UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA - DOUTORADO
PESQUISA EM HISTÓRIA E FILOSOFIA DA PSICOLOGIA

Hugo Tannous

The epistemological problems of the clinical research in psychoanalysis after Grünbaum: some directions and resolutions

Juiz de Fora
2021

Hugo Tannous

The epistemological problems of the clinical research in psychoanalysis after Grünbaum: some directions and resolutions

Tese apresentada ao Programa de Pós-Graduação em Psicologia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Psicologia. Área de Concentração: História e Filosofia da Psicologia.

Orientador: Richard Theisen Simanke

Coorientador: Stephen Frosh

Juiz de Fora

2021

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Tannous, Hugo.
The epistemological problems of the clinical research in psychoanalysis after Grünbaum : Some directions and resolutions / Hugo Tannous. -- 2021.
342 p.

Orientador: Richard Theisen Simanke
Coorientador: Stephen Frosh
Tese (doutorado) - Universidade Federal de Juiz de Fora, Instituto de Ciências Humanas. Programa de Pós-Graduação em Psicologia, 2021.

1. Clinical psychoanalysis. 2. Clinical method. 3. Clinical research. 4. Epistemology. 5. Adolf Grünbaum. I. Theisen Simanke, Richard, orient. II. Frosh, Stephen, coorient. III. Título.

FEDERAL UNIVERSITY OF JUIZ DE FORA
RESEARCH AND GRADUATE PROGRAMS OFFICE

HUGO TANNOUS JORGE

THE EPISTEMOLOGICAL PROBLEMS OF THE CLINICAL RESEARCH IN PSYCHOANALYSIS AFTER GRÜNBAUM: SOME DIRECTIONS AND RESOLUTIONS

Thesis submitted to the Postgratuate Program in Psychology of the Federal University of Juiz de Fora as a partial requirement for obtaining a Doctoral degree in Psychology. Concentration area: Psychology

Approved on the 15 of December of 2021.

EXAMINING BOARD

Doctor / Professor Richard Theisen Simanke – Academic Advisor
Federal University of Juiz de Fora

Doctor / Professor Caio Padovan
Université Paul Valéry - Montpellier 3 - Pontifícia Universidade Católica do Paraná

Doctor / Professor Michael Lacewing
University College London

Doctor / Professor Michael T. Michael
Yonsei University - Underwood International College

Doctor / Professor Paulo Antonio de Campos Beer
Universidade de São Paulo

FEDERAL UNIVERSITY OF JUIZ DE FORA
RESEARCH AND GRADUATE PROGRAMS OFFICE

HUGO TANNOUS JORGE

THE EPISTEMOLOGICAL PROBLEMS OF THE CLINICAL RESEARCH IN PSYCHOANALYSIS AFTER GRÜNBAUM: SOME DIRECTIONS AND RESOLUTIONS

Thesis submitted to the Postgratuate Program in Psychology of the Federal University of Juiz de Fora as a partial requirement for obtaining a Doctoral degree in Psychology. Concentration area: Psychology

Approved on the 15 of December of 2021.

EXAMINING BOARD

Doctor / Professor Richard Theisen Simanke – Academic Advisor
Federal University of Juiz de Fora

Doctor / Professor Caio Padovan
Université Paul Valéry - Montpellier 3 - Pontifícia Universidade Católica do Paraná

Doctor / Professor Michael Lacewing
University College London

Michael

Doctor / Professor Michael T. Michael
Yonsei University - Underwood International College

Doctor / Professor Paulo Antonio de Campos Beer
Universidade de São Paulo

AGRADECIMENTOS

Escrever uma tese de doutorado durante o fim do mundo não é fácil! Mas é possível, se as pessoas certas estiverem ao seu lado.

Sou grato ao Giovane, que esteve ao meu lado em absolutamente todos os momentos do doutorado, desde os mais auspiciosos até os mais enlouquecedores; que transformou minhas neuroses em piadas. Sem sua ternura, entrega, confiança, suas curadorias de filmes e livros, suas logísticas, seus trocadilhos, a tese seria uma pobreza só e a jornada do doutorado, uma "brutíssima confusão". Serei eternamente grato por ele ter me ajudado a lapidar as "brutíssimas", e pelas tantas aventuras que vivemos juntos. (Quem se cansa de Londres se cansa da vida, e Deus sabe que a gente não se cansou daquela cidade).

À minha mamãe, que ficou mais eufórica do que eu quando a Defesa terminou bem. Ela nunca estremeceu diante do meu desejo de ser psicólogo e um pouco filósofo; sempre acolheu as grandes escolhas da minha vida com seu amor infinito e sua coragem pra amar infinito. Não são hipérboles: "nunca" e "sempre" são as palavras mais apropriadas aqui. Deve-se à minha mamãe a liberdade de que hoje usufruo pra pensar, sentir e criar, pra buscar um sentido e um propósito no mundo. "Só" isso... E mais um pouco: vem dela, professora de Matemática, o meu desejo de ser professor, bem como o meu interesse pela Lógica.

Ao meu papai, que está sempre aqui quando escrevo. Interpretou textos comigo desde que eu me entendi por gente, refletiu comigo sobre os mais variados temas madrugada afora, foi o primeiro a ler e editar meus *drafts* (Hugo criança queria ser escritor de fantasia e terror); *last but not least*, ele que me fez ter coragem frente ao mar (que me ensinou a curtir suas bravas e fugazes ondas). Sua consciência se afastou de nós há bastante tempo, e tive que reconstruí-la dentro da minha própria; tarefa dolorosa e, convenhamos, impossível, mas... tento, e pelo caminho, largo uns parágrafos.

Ao meu irmão. Ao longo de todas as minhas andanças em busca de uma formação de qualidade, ele sempre se contorceu com seu carro e seu tempo pra que eu não ficasse esbodegado em minhas visitas de férias e feriados; as conversas que tivemos nessas "baldeações" me faziam esquecer do sono que se atrasava. Sou grato por todos os *memes* e vídeos do TikTok que ele compartilhou comigo; alguns me faziam esquecer que, com doutorado e pandemia, a vida também, desesperadamente, se atrasava.

À minha irmã, que sofreu mais que eu durante a defesa. Que me trata como bebê, mas também como adulto. Que foi a primeira a dizer "acho que você tem bastante de professor universitário" ainda em meus anos de faculdade; o primeiro espelho em que me vi mestre e doutor. Sou grato a ela também por ter dado à luz com tanto amor Ayla, a Anciã, Aurora, a Sorridente, e Antonela, a Serena, e por estar me dando a honra de carregar essas gordinhas, de rir das ternurinhas delas e de ajudar a educá-las. A essas bebezinhas sou grato. À Ayla, pelas sonecas e canções que curtimos juntos; à Aurora, pelas estórias e abraços que trocamos; e à Antonela, que veio à luz mais suave que minha tese. Ao meu cunhado Marcelo, pelo incentivo e ajuda; por ter dirigido mais de 1300km pra que eu pudesse estar com minha família sem o risco de repassar o infame vírus.

À Vovó Bassima, por ser lembrança de conforto e carinho; por ter me comparado ao seu pai, professor, contador de histórias e escritor. À Tia Samira, por ser exemplo de bom humor e por ter cuidado de mim. Às primas e aos primos, pela torcida.

Sou grato ao meu orientador, Professor Richard Simanke, pelo exemplo que me deu ao filosofar de forma escrupulosa e iconoclasta. E iconoclasta ele é, não só com os autores que estuda, mas também consigo mesmo; é o professor universitário menos vaidoso que eu conheço, recebe gabações com uma involuntária careta. Ele realmente acha que o que produz não é nada demais, e isso me intriga bastante, já que foi o brilhantismo de seus artigos (os quais comecei a ler ainda na Graduação) que me trouxe à distante e solitária Juiz de Fora pra fazer mestrado. Sou grato a ele também por ter confiado tanto em mim, principalmente quando sugeriu que eu escrevesse minha tese em inglês.

Sou grato ao Professor Jubel Barreto por ter me dado a oportunidade de fazer parte do meu Estágio em Docência nas disciplinas de Psicologia Médica I e II; sua orientação foi atenta e rica, e ele também depositou uma lisonjeante confiança em mim. À Professora Fátima Caropreso, por ter me dado a oportunidade de estagiar na disciplina de Seminários de Psicanálise Freudiana; sua risada contagiante, seu afeto e seu rigor são inesquecíveis. Sou também grato aos meus alunos de estágio e aos monitores com quem trabalhei: pelo interesse, pelas perguntas, argumentos e assistências. Aprendi e me diverti com eles.

Sou profundamente grato à CAPES pela bolsa do Sanduíche no Exterior. As fontes e os argumentos da minha tese não teriam sido empolgantes não fossem os debates que assimilei e o material que coletei no continente que deu origem à psicanálise (sim, eu acho a minha tese empolgante). No período-sanduíche, tornei-me um pesquisador melhor; desenvolvi habilidades que vou levar pra vida toda. Que essa instituição e o regime democrático que a sustenta se recuperem em breve, que tenham vida longa, e que os inimigos da Educação e da Cultura morram aos poucos nas mentes do Brasil e do mundo.

Também sou profundamente grato ao meu colega TAE Jaime e à Professora Virgínia Pereira. Jaime se dispôs a declarar em papel timbrado que cobriria, durante meu afastamento pro período-sanduíche, as funções do meu cargo, e defendeu ferozmente meu pedido no Conselho de Unidade, enquanto sua chefe Virgínia favoreceu seu gesto de todas as formas que podia. Sem a empatia e a coragem dessas pessoas, eu teria sucumbido frente à zelotipia-em-pele-de-legalismo de que fui vítima àquela época e não teria usufruído da minha bolsa da CAPES; teria me tornado um servidor infeliz e um doutorando infeliz. Também sou grato ao meu colega TAE Flávio Sereno, coordenador do SINTUFEJUF, por ter se sensibilizado com os problemas relativos ao meu afastamento e por ter agido pra tentar resolvê-los.

Sou grato aos meus colegas TAEs da UFJF, principalmente a Rosi, Lílian, André e Gisele, por terem demonstrado apoio ou colaborado pra que eu completasse minha jornada. E também aos melhores chefes que tive na vida: Professora Arise Galil e Professor Diego Sarkis.

Sou grato aos Professores Francis Justi e Lélio Lourenço, por sempre terem se mostrado dispostos e atuantes enquanto coordenadores do PPG. À época dos conflitos envolvendo o pedido de afastamento, o Professor Francis também demonstrou empatia e coragem ao defender em documento oficial pro Conselho de Unidade da FAMED que abrir mão de uma bolsa de Doutorado-Sanduíche no Exterior seria danoso pra Universidade como um todo. O mesmo fez a Professora Mônica de Oliveira, Pró-Reitora de Pós-Graduação e Pesquisa de então. Sou grato a ela e às TAEs Gabriella e Eduarda, que com toda a sua boa vontade fizeram meu pedido de apoio chegar até a Pró-Reitora.

Sou grato ao Professor Jim Hopkins, por ter sido caloroso e incansável frente ao meu interesse em passar um período-sanduíche na UCL, por ter negociado o meu cargo de *Honorary Research Assistant* nessa prestigiada universidade, pelas suas inspiradoras aulas e pelas nossas conversas na *ASCR*. Minha ideia de relacionar psicanálise clínica e noções de probabilidade se assentou com ele; as primeiras lições que aprendi em Neuropsicanálise vieram dele e através dele. (Minha grande estima por ele está, aliás, refletida nos argumentos desta tese.)

Sou grato ao Professor Stephen Frosh. Ele me acolheu na eminente Birkbeck e nunca hesitou em colaborar pra que a papelada da CAPES estivesse em dia; ele me direcionou a eventos excepcionais da universidade (assisti a palestras e seminários de Slavoj Žižek, Judith Butler, Derek Hook e Jacqueline Rose, por exemplo); mas mais importante que tudo isso, ele leu atentamente as primeiras loucas versões dos meus argumentos, e me recebeu mensalmente em sua sala pra me dar *feedbacks*, debater temas e me indicar novas direções (sempre com uma polidez única e muito *British-Jewish humour*). Que sua escrita elegante continue me inspirando; que suas ideias continuem me desafiando.

Aos amigos que fiz no Sanduíche: Alexandros, Aline, Ana, Anna, Cameron&Savannah, Guilherme, Irene, Isla&Carlos, James, Patrícia, Philip, Reg&Kev, Sharon, Timmy, Tobias, Tom, William e muitos outros. Meus amigos Alexandros, Philip, Timmy, Tom e William debateram comigo sobre psicanálise, o mal-estar pós-moderno e diferenças culturais nos pubs ao redor de Birkbeck; Alexandros, com sua simpatia mediterrânea, até leu meu texto de qualificação e escreveu extensos comentários a ele (!). Aline, Ana, Anna, Guilherme, Isla&Carlos e Tobias foram confortos de brasilidade em Londres. Irene e Patrícia, surpresas bárbaras em congressos fora da Inglaterra, ambas também me trazendo o conforto das origens. Guilherme e Irene, em particular, pra sempre serão meus interlocutores teóricos. Cameron&Savannah nos divertiram bastante. James e Sharon foram anjos britânicos quando Giovane e eu estivemos no *International Students House*. E Reg&Kev foram anjos australianos em nosso primeiro mês em Londres, fazendo muito mais do que qualquer *host* do Airbnb faria (pra começar, nos doaram uma TV e levaram nossa mudança de carro de Bromley até NW1); são hoje nossos amigos.

Sou grato aos Professores Caio Padovan e Paulo Beer, que além de serem grandes amigos foram parte da minha banca. Caio é uma enciclopédia ambulante sobre história e epistemologia da psicanálise e, eloquente e calmo, parece a reencarnação de um intelectual do século XIX; seu lisonjeante interesse pelo meu trabalho, já em uma época atrapalhada do meu mestrado (UFES, 2015), e mais ainda ao longo do meu doutorado, me fez entender pela primeira vez que eu poderia, sim, dar uma relevante contribuição pra nossa área. Isso e as suas ajudas com fontes bibliográficas fizeram toda a diferença... Com seus desbravadores trabalhos de mestrado e doutorado e seus gestos calorosos, Paulo também trouxe pra minha jornada muitas faíscas filosóficas e um sentimento de comunidade. Ainda me lembro da minha animação quando eu soube que um cara da USP estava estudando o tema que eu queria estudar no doutorado... Dei um *add* no Face, fui tietá-lo no lançamento do seu livro e nos tornamos amigos. Outro sem o qual eu não teria feito o Sanduíche, e que trouxe uma graça extra pra esse período, juntamente com Lu e Tereza. Sou grato também por ele ter colocado uma piadinha ou outra em seus trabalhos acadêmicos; me deixaram seguro pra que eu colocasse um monte de piadinhas nos meus. Paulo e Caio são acadêmicos brilhantes, e é uma honra pra mim ser amigo deles.

Sou grato aos Professores Michael Lacewing e a Michael T. Michael. Nas vezes em que lia algum dos artigos ou capítulos de livros de um deles, sempre surgia em mim o pensamento: “Isso daqui ‘tá muito bom. Argumento e estilo. É esse tipo de coisa que quero escrever um dia”. Epítomes de rigor e lucidez dentro da epistemologia da psicanálise de tradição anglo-saxã, e foram parte da minha banca, e debateram comigo, e me criticaram e elogiaram, desde a

qualificação. Coisa fina... Sou grato à disposição e à diligência deles, e certamente carregarei suas intervenções comigo pros meus próximos anos de trabalho.

À Teacher Paola Temponi, por ter libertado minha língua e minha mente pra a Defesa. Grato pelo cuidado e preocupação, pelo tempo despendido, pelo bom humor.

Sou grato ao Professor Weiny Pinto, por sempre ter sido tão entusiasmado com o tema da minha tese e por exercer tão efetiva liderança em eventos e publicações sobre história e filosofia da psicanálise. Como se não bastasse, é um amigo que não economiza em gentilezas. Que bom que fiz aquela pergunta no Congresso em Curitiba!

Sou grato aos amigos com quem convivi em Juiz de Fora durante o doutorado: Diego, Eduardo, Liliam, Lucas, Luciana, Rayssa, Rondineli. Fizeram-me rir e pensar e sentir menos solidão. Diego, Eduardo, Lucas, Rayssa, Rondineli e eu: não se engane com os *nerds* da História e Filosofia da Psicologia, somos Vida Loka, adoramos fofocar sobre os professores e filósofos enquanto tomamos uma cerva (ou um pingado na padaria, acompanhado, claro, de um pãozinho com manteiga; como eu disse, Vida Loka). Liliam cuidou das minhas dores corporais com Pilates e conversas. Luciana foi minha companheira de piração e de receitas *low carb* no exótico ano de 2020, e quase me convenceu que meus conhecimentos sobre método científico eram vastos. (Eu deveria ter feito mais receitas *low carb*, porque acabei por desenvolver uma bolota barrigal, a mais universal de todas as heranças doutorais).

Aos amigos de Floripa Allan, Caio, Juliane e Pedro, que durante um momento crítico do doutorado me entreteram com tirinhas, debates acadêmicos, e dicas de filmes e livros através do grupo de *Whatsapp* "Neuróticos Comun(a)s". A jocosidade das tirinhas variou, a seriedade dos debates variou; a única coisa que não variou foi a menção a genitais humanos. Tenho que ir a Floripa em breve; restaurarei assim meus níveis antigos de alegria. Com esses amigos, e com muitos outros (Salsi, Adri, Brunfan,...).

Às amigas de infância, Daniela, Lídia, Luma, e Mayra. Pelo carinho, pelas crises de riso, trocas e memórias. Que a Luma renove seus votos; que possamos fazer festonas, não importa o motivo, que é pra gente se reunir e ser bagaceiro em roupas de gala.

Ah, sou grato também ao meu superego, que me obedeceu algumas das vezes em que eu gritei "*silenzio, Bruno!*" pra ele.

E, por fim...

(Pega bem ser grato ao analista?)

Once upon a time, there was a young sculptor named Cintolo. He served a king in a realm so far underground that neither the sun's beams nor the moonlight could find it. He filled the royal gardens with flowers sculpted from rubies and fountains sculpted from malachite. He carved busts of the king and queen that were so lifelike, everyone believed they could hear them breathe.
Their only daughter, the princess Moanna, loved to watch the sculptor work, but Cintolo never managed to sculpt her form. "I can't sit still for that long, Cintolo," she said. "There's too much to do and too much to see."
Then one day Moanna was gone. And Cintolo remembered how often she'd asked him about the sun and the moon and whether he knew what the trees, whose roots laced the ceiling of her bedroom, looked like above the ground.
[…]
Cintolo didn't use stone for this sculpture. He carved the Faun's likeness from wood, for wood always remembers it was once a living tree, alive and breathing in both kingdoms, the one above and the one below.
It took Cintolo three days and three nights to finish the sculpture and, when he told the Faun to rise from his chair, so did his wooden image.
"Tell it to find her, Your Horned Highness," the sculptor said. "I promise it will neither rest nor die before he does." (Del Toro & Funke, 2019, pp. 39-42).

1. You can only really wish what you believe to be possible.
2. You can only believe to be possible what is part of your history.
3. It can only be a part of your history that which you truly wish. (Ende, 2008)[1].

[…] Mr Segundus wished to know, he said, why modern magicians were unable to work the magic they wrote about. In short, he wished to know why there was no more magic done in England.
[…]
The President of the York society (whose name was Dr Foxcastle) turned to John Segundus and explained that the question was a wrong one. "It presupposes that magicians have some sort of duty to do magic – which is clearly nonsense. You would not, I imagine, suggest that it is the task of botanists to devise more flowers? Or that astronomers should labour to rearrange the stars? Magicians, Mr Segundus, study magic which was done long ago. Why should any one expect more?" (Clarke, 2015, p. 4).

[1]My translation from

1. Du kannst nur wirklich wünschen, was du für möglich hältst.
2. Du kannst nur das für möglich halten, was zu deiner Geschichte gehört.
3. Nur das gehört zu deiner Geschichte, was du in Wahrheit wünschst.

ABSTRACT

Since the advent of the clinical-psychoanalytic method, philosophers and scientists have identified problems in its epistemological pretensions or at least in the degree of disposition of psychoanalytic communities to try to overcome its epistemological shortcomings. The aim of this doctoral dissertation is to investigate to which degree and by which rationale the data ensued by the clinical-psychoanalytic method would be able to cogently support hypotheses about the behaviour of individuals and classes of individuals as well as about some operations of the human mind in general. It is thus an elaboration on both the actual and the potential logic of both the generation and the justification of clinical-psychoanalytic hypotheses. For its enduring influence, I choose Adolf Grünbaum's critique on the foundations of that method to frame such an elaboration. With his Millian orientation, Grünbaum concludes that Freud and his followers are unable to defend the probative character of their clinical method. He argues that Freud's attempt to link therapeutic outcome with truthfulness, his "Tally Argument", is unsound; that the clinical data from which clinical-psychoanalytic inferences are made are probably contaminated by the analysts' suggestions; and that, in any case, many of such inferences instantiate causal fallacies and biases. I take these three classes of arguments in Grünbaum as three classes of epistemological problems to be redirected and solved rather than just evaluated in their fairness with the psychoanalytic corpus and practice. I counterargue that clinical-psychoanalytic inferences are or can be essentially Millian as long as they are or can be almost identical to the Explanationist-Bayesian inferences in the historical sciences. I discuss the concept of suggestion and the theories about the conditions to its occurrence, thus deflating Grünbaum's contamination argument but also revealing that analysts should be able to include heterosuggestions in their database; I formalize a technical device to help them in this task. Faced with the problem of how truthful enunciations during treatment are related to therapeutic effects in the patient, I argue that the former can be considered INUS causes of the latter, which implies that therapeutic effects cannot be considered unequivocal evidence, but only complementary evidence, for antecedent truthfulness; that, therefore, Grünbaum is right, but shallow, about Freud's Tally Argument. I conclude, in contrast with Grünbaum, that research from clinical-psychoanalytic data is feasible, as the method involved in this research is already cogent in many respects and can become more and more cogent through the stabilization or implementation of certain principles and guidelines; and that, moreover, it is an irreplaceable and extremely relevant kind of research. I also conclude, however, that a question of how consistent the results of this method can be across its applications is still barely raised by analysts and by those committed to advancing this kind of research. To make clinical-psychoanalytic research widely recognized as scientific, a combination of philosophical, technological, and institutional efforts is needed.

Keywords: clinical psychoanalysis; clinical method; clinical research; epistemology; Adolf Grünbaum.

RESUMO

Desde o surgimento da psicanálise, filósofos e cientistas identificaram problemas nas pretensões epistemológicas de seu método clínico ou ao menos na disposição das comunidades psicanalíticas para buscar superar as fragilidades epistemológicas de tal método. O objetivo desta dissertação doutoral é investigar até que ponto e por que razões os dados produzidos pelo método clínico-psicanalítico seriam capazes de sustentar de forma cogente hipóteses sobre o comportamento de indivíduos e classes de indivíduos, bem como sobre algumas operações da mente humana em geral. Ela consiste, portanto, em uma elaboração da lógica, vigente ou potencial, presente tanto na geração quanto na justificação de hipóteses clínico-psicanalíticas. Por sua duradoura influência, a crítica de Adolf Grünbaum sobre os fundamentos de tal método serviu de pivô para tal elaboração. Com sua orientação milliana, Grünbaum conclui que Freud e os freudianos são incapazes de defender o caráter probatório de seu método clínico. Ele argumenta que a tentativa de Freud de ligar resultados terapêuticos à veracidade, seu "Argumento da Correspondência", depende de uma premissa falsa; que os dados clínicos a partir dos quais se fazem inferências clínico-psicanalíticas estão provavelmente contaminados pelas sugestões dos analistas; e que, de qualquer modo, muitas dessas inferências são instâncias de falácias e vieses causais. Essas três classes de argumentos em Grünbaum são tomadas nesta dissertação como três classes de problemas epistemológicos a serem encaminhados e resolvidos em vez de apenas avaliados por sua justeza em relação à obra e à prática psicanalíticas. Contra-argumenta-se que inferências clínico-psicanalíticas são ou podem ser essencialmente millianas na medida em que são ou podem ser quase idênticas às inferências bayesiano-explicacionistas das ciências históricas. Discute-se o conceito de sugestão e as teorias sobre as condições para sua ocorrência, o que deflaciona o argumento grünbaumiano de contaminação, mas também revela que analistas devem ser capazes de incluir hétero-sugestões em sua base de dados; é formalizado um dispositivo técnico para ajudá-los nessa tarefa. Face ao problema do quanto enunciações verazes durante o tratamento estão relacionadas a efeitos terapêuticos no paciente, argumenta-se que aquelas podem ser consideradas causas INUS destes, o que implica que efeitos terapêuticos não podem ser considerados evidências inequívocas, mas apenas complementares, de uma veracidade antecedente; que, portanto, Grünbaum é preciso, mas superficial, em relação ao Argumento da Correspondência de Freud. Conclui-se, contrariamente a Grünbaum, que a pesquisa a partir de dados clínico-psicanalíticos é viável, já que o método envolvido em tal pesquisa já é cogente em muitos aspectos e pode se tornar cada vez mais cogente através da estabilização ou implementação de certos princípios e diretrizes; e que, ademais, trata-se de um tipo de pesquisa insubstituível e extremamente relevante. Também se conclui, no entanto, que a questão do quão consistentes os resultados desse método podem ser através de suas aplicações, ou seja, a questão de sua confiabilidade, é ainda pouco levantada por psicanalistas e por aqueles comprometidos com o desenvolvimento desse tipo de pesquisa. Para que a pesquisa clínico-psicanalítica se torne amplamente reconhecida como científica, uma combinação de esforços filosóficos, tecnológicos e institucionais é necessária.

Palavras-chave: psicanálise clínica; método clínico; pesquisa clínica; epistemologia; Adolf Grünbaum.

CONTENT

1 INTRODUCTION

According to Freud's encyclopaedia article published in 1923, psychoanalysis can be defined as a procedure for the investigation of mental processes, as a method to treat neurotic disorders, or as a collection of psychological information (Freud, 1955e). That investigation is a condition for this treatment and psychological theory; its foundations, therefore, are the foundations of everything Freud called psychoanalysis. Anyone interested in psychoanalysis, thus, should ask oneself: what are such foundations made of?

In "Notes upon a case of obsessional neurosis", Freud (1955d) presents the context of one of his interpretations in the clinical setting. After saying to the Rat Man that his obsessive thoughts were linked to a hateful affect toward his father, born out of certain events in his childhood, the patient gets sceptical: how can it be explained that these thoughts, thus this supposed hate, had disappeared for many years? The psychoanalyst, then, affirms that the patient already knew the answer to that question, after which the Rat Man goes on speaking "somewhat disconnectedly" that

> he had been his father's best friend, and that his father had been his. Except on a few subjects, upon which fathers and sons usually hold aloof from one another (what could he mean by that?), there had been a greater intimacy between them than there now was between him and his best friend. As regards the lady for whose sake he had sacrificed his father in that idea of his, it was true that he had loved her very much, but he had never felt really sensual wishes toward her, such as he had constantly had in his childhood. Altogether, in his childhood his sensual impulses had been much stronger than during his puberty (p. 182).

After this, along with elements of the patient's former statements, the psychoanalyst concludes that his hateful affect, thus also his "parricidal" thoughts, were caused by his childish perception of the father as deterrent of his sexual appetites; the gaps between the obsessive outbreaks, therefore, were correlated to the gaps between the irruptions of these appetites. Basic epistemological questions can be posed here. Why would the words of the Rat Man be evidence for the enunciated interpretation and also for its extended version? Which elements of the sessions until then, and of what happened after the extended interpretation was enunciated, supported the inference on the cause of the patient's obsessive thoughts? How can we unequivocally assure that the patient's cognitive processes, such as his memory, were not unconsciously tuned to what the medical authority, unconsciously or not, expected of him? And, finally, how can the Rat Man help us know something of the obsessional neurosis *as a category*?

In this doctoral thesis, we shall discuss the logic of discovery and justification of psychoanalytic hypotheses on the ill and the sane human mind insofar as this logic can unfold in the clinical setting alone. That is, we shall discuss the possibility of producing knowledge from the actions and circumstances of psychoanalytic patients around what makes them love, suffer and tolerate.

Since the outset of psychoanalysis, its clinical method has suffered from various charges. Philosophers and scientists have identified problems both in the very nature of the method and in the degree of disposition of psychoanalytic communities to improve it. Glover (1952) notes that, if the problems are owed

> partly to the conditions under which clinical analysis is conducted, and partly to the use of interpretive techniques of investigation, [the conclusion is that] it is not possible either to employ fully or fully to depend on the forms of scientific control customary in most other sciences (p. 403).

But the problems can also be due to a "tendency not to apply to the data of observation or to the methods of interpretation such scientific controls as are available" (p. 403).

One task of the epistemology of psychoanalysis is to embrace the first kind of problems and, resisting the cry of the psychoanalytic communities, to suggest solutions for the rest of them. But how to get to know such unavoidable limits and unapplied possibilities?

Listing some persistent expressions of discomfort with clinical psychoanalysis may be the first step toward answering this question. Many scholars have held and still hold:

a) that psychoanalysis' principles for interpreting a patient's mental products were never formalized;
b) that generalizations in clinical psychoanalysis are ill-supported since they come from a small set of cases (or from just one case);
c) that its concepts are not operationally defined;
d) and that its theoretical results are tinted with the subjectivity of analyst and patient.

Such facts would allegedly be the causes of the deep theoretical divergences between psychoanalysts. To make things worse, between analysts we can also see a longevous metatheoretical war, a war on the possibility, origin and essence of psychoanalytic knowledge.

The shortage of precepts for interpreting mental products is the "Achilles heel" of psychoanalytic research, according to Glover (1952, p. 405). Also a long time ago, Seitz (1966) identifies a "consensus problem" in psychoanalytic and psychotherapeutic research: the typical difficulty of clinicians to agree upon the interpretation of the same set of data. Only 15 years ago, citing Glover, Seitz and also Rapaport (1960), Wallerstein (2006) claimed that part of what had been causing the anti-science tendency in clinical psychoanalysis was the fact that "skilled psychoanalytic clinicians can construct differing, but often equally compelling, formulations of psychoanalytic case material and that no ready method to establish the truth claims of alternative formulations has yet come to win wide acceptance" (p. 304). In confounding disagreements over operational definitions with inferential disagreements, though, Wallerstein (2006) also felt entitled to affirm that the problem is not urgent for a scientific psychoanalysis, as solutions for operational problems were already on their way. Very recently, Eagle (2019) argued otherwise. He held that the psychoanalytic community still needs to overcome their indifference or hostility toward studies that try to tell the difference between "arbitrary theoretical impositions on to the clinical material" and "interpretations warranted by the clinical evidence" (p. 370).

Even if the interpretation of a case were warranted, using it to support general statements about the psyche of a category of humans or of all humans would still be controversial. According to Hinshelwood (2013), one of the 4 key challenges of psychoanalysis is to get to know how to generalize from unique case-studies; the challenge "appears insurmountable, and leaves psychoanalysis way behind most experimental psychology [...]" (p. 9). Citing other authors, Willemsen, Della Rosa and Kegerreis (2017) write that the generalizability problem can still be summoned to dispute the case-study method, and that reading, writing and presenting case studies has been treated as a "ritual to affirm analysts in their professional identity, rather than [as] a research method" (p. 2). Meganck, Inslegers, Krivzov and Notaerts (2017) remind us that "the potential for knowledge aggregation across cases [...] remains unexplored despite the existence of methodological tools to do so, for example, meta-synthesis methods (Finfgeld, 2003; Willemsen et al, 2015) and case comparison methods (Iwakabe and Gazzola, 2009)", being in this matter in unison with Willemsen et al (2017).

The charge that what primarily hinders the scientific accreditation of psychoanalysis is that its concepts are too vague, its community not trying to define them operationally, dates back to Glover's discontents and to the critiques of Nagel (1959) and Popper (1963). Glover (1952) held that "before we can even begin to conduct systematic research on psychoanalysis, a preparatory phase is essential during which a process of standardization and definition of terms and concepts can be effected" but that, "it must be confessed", "no such agreement exists" (pp. 405-406). "To take a simple instance", says Glover, "no one to my knowledge has yet defined the term 'castration complex' in a form which would permit statistical correlation […]" (p. 405). Nagel (1959) argued likewise that Freudian theory presents flaws related to the linking of theoretical notions to "fairly definite and unambiguously specified observable materials" (p.

40). Somehow mimicking Nagel, Popper (1963), argued that Freudian theory is similar to "astrology, rather than [to] astronomy", since "every conceivable case could be interpreted in the light of […] [the] theory" (p. 35). This charge persists until our days, according to Eagle's testimony:

> Unfortunately the psychoanalytic literature is replete with formulations that are so vague and/or so obscure that it is difficult to conceive of any evidence that could play a role in their evaluation. Especially unfortunate is the fact that the authors of these formulations are not obscure figures, but include the most prominent and influential psychoanalytic theorists in the world today (Eagle, 2019, p. 363).

Finally, the psychological constraints of analyst and patient due to their engagement in a therapeutic venture are also said to burden the clinical research in psychoanalysis. Lacewing (2018) contends that Popper's diagnosis indicates not merely loose definitions, but, more widely, "confirmation bias, which involves the selective gathering, weighing, or interpretation of evidence that supports one's existing beliefs or favoured hypothesis while neglecting or discounting evidence that tells against one's view" and which "many psychoanalysts have undoubtedly manifested" (p. 100). In addition, there is the problem of the patient's suggestibility:

> […] it is possible that the theoretically based expectations of the analyst influence the behaviour and free associations of the analysand so as to produce erroneous data that support those very expectations. Unless this possible bias can be corrected for, clinical data do not supply independent support for psychoanalytic theoretical claims (Lacewing, 2018, p. 98).

These are recent paths to discuss psychological constraints in the clinic, but they are acknowledged since Freud; in his jargon, they are the resistant dimension of transference and countertransference.

Such problems were, since Freud, largely neglected. As claimed by Rubovits-Seitz (1998), Freud minimized the role interpretations had in his method, never outlined his interpretive method acutely or as a system, used to become hypersensitive if his empiricism and objectivity were questioned and to diminish methodologists and their high standards of objectivity, among other inauspicious conducts. Freud averted the delivery of methodologic elucidations even in "The interpretation of dreams", in his papers on technique, and in a case study such as Dora's (Rubovits-Seitz, 1998). Thereby, he "set a problematic precedent for psychoanalysis and dynamic psychotherapy which continues to the present day" (Rubovits-Seitz, 1998, p. 9).

More than half a century ago, Glover (1952) depicted the post-Freudian power dynamics through which this neglect with the clinical method is buttressed:

> An analyst, let us say, of established prestige and seniority, produces a paper advancing some new point of view or alleged discovery in the theoretical or clinical field. Given sufficient enthusiasm and persuasiveness, or even just plain dogmatism on the part of the author, the chances are that without any check, this view or alleged discovery will gain currency, will be quoted and requoted until it attains the status of an accepted conclusion. Some few observers who have been stimulated by the new idea may test it in their clinical practice. If they can corroborate it they will no doubt report the fact: but if they do not, or if they feel disposed to reject it, this scientific "negative" is much less likely to be expressed, at any rate in public; and so, failing effective examination, the view is ultimately canonized with the sanctioning phrase, "as so-and-so has shown". In other words an *ipse dixit* acquires the validity of an attested conclusion on hearsay evidence only (p. 403).

Yet, sadly, a similar state of affairs was debunked as recently as 2019 by Morris Eagle. He claimed that psychoanalytic schools are not born out of "the presentation of evidence that refutes earlier theories or compelling new evidence that supports […] [a] new theory", but from "the emergence of a charismatic figure who attracts loyal adherents, often culminating in the establishment of psychoanalytic institutes that provide training from a particular theoretical perspective […]" (Eagle, 2019, p. 365). In other words: "if one looks at the historical trajectory of psychoanalytic theorizing from Freudian to dominant contemporary theories, it appears that at least as much influential psychoanalytic theorizing is increasingly characterized by obtuse formulation far removed from observation and evidence" (p. 371).

The most remarkable of the discontented with the clinical method of psychoanalysis was the philosopher Adolf Grünbaum. Mitchell (1998) playfully admits the existence of a syndrome related to the contact with the philosopher's attack on that method:

> What follows for an analyst afflicted with the Grünbaum Syndrome is several days of guilty anguish for not having involved oneself in analytic research. This may include […] pulling a twenty-year-old statistics text off the shelf and quickly putting it back. There may be a sleep disturbance and distractions from work. However, it invariably passes in a day or so, and the patient is able to return to a fully productive life. (p. 5)

According to Edelson (1984), though, such symptoms should not go away that easily. In considering the three major epistemological challenges to psychoanalysis, Nagel's (1959), Popper's (1963) and Grünbaum's (1984), he comments: "The first two challenges may now seem to some in retrospect to have been hardly worth the bother of a response, but I cannot think that will ever be said about the third" (p. 1). Already in our century, one of the first philosophers of psychoanalysis in Brazil, Renato Mezan, keeps Edelson's statement alive:

> Grünbaum's critique of psychoanalysis is not only more caustic and forceful than the ones emerging from the Logical Positivism and Popper. It is also graver, since it does not aim, like those do, to frame psychoanalysis in an abstract definition of science […], but *to annihilate the belief in the validity of the clinical method as knowledge-producer*. Unfortunately for us analysts, Grünbaum is no negligible opponent: his knowledge of Freud and of the

> psychoanalytic literature is vast and precise, he erects his reasoning in an astute manner, he writes clearly and with a hint of irony. In sum, it is not easy to refute his arguments. (Mezan, 2006, p. 229)

Grünbaum's first foray into the philosophy of psychoanalysis occurred in 1958, in a Meeting of the New York University's Institute of Philosophy to discuss the issue "Psychoanalysis, scientific method and philosophy" (Hook, 1959). In the Meeting, he condemns Lawrence Kubie's view that psychoanalysts do not need to state what would be the evidence for a case of Oedipus Complex's absence, "for in rendering that hypothesis *proof against disproof* [...], Dr. Kubie's conception makes it *irrelevant* to the explanation of observable human behavior" (Grünbaum, 1959, p. 225). As of 1975, he began to scrutinize Popper's philosophy of science and his argument that psychoanalytic theory does not have potential falsifiers; he held the Popperian indictment to be wrong given some instances of that theory. In turning to philosophical works on psychoanalysis, he became dissatisfied that most of them "did not come to grips at all with the key issues of its acceptability as an explanatory theory and was also […] unilluminating concerning the logical structure and rationale of Freud's edifice" (Grünbaum, 2008, p. 576).

Eventually, throughout many papers he came to develop an original epistemological critique of psychoanalysis; the eminent book "The Foundations of Psychoanalysis", in 1984, was to some extent a synthesis of those papers. He expanded the arguments of the book along the decades (his last publication of the kind was in 2015) but their common core endured: the defence that Freud and his followers never presented a good logical and empirical basis for eliminating some hypotheses that could alternatively account for the data of clinical psychoanalysis, such as the hypothesis of the action of cognitive and therapeutic suggestion and the hypothesis of incidental (unmotivated) affinities between mental products. He concluded that clinical psychoanalysis can justify itself as a heuristic device, but not as a probative one; in other words, that it is competent at generating hypotheses, but not at checking whether they are true.

The critique was first viewed as a fatal blow to psychoanalysis; supposedly, the "Freud Wars" had been over, and the invaders had won. However, the Grünbaumian critique has proved to be a powerful cue for psychoanalysts to develop their epistemological assumptions. With philosophers like Grünbaum, psychoanalysis ends up living always a little longer. As Edelson (1988) pointed out decades ago:

> [...] because of the explicitness and utter lucidity of his argument, and the thorough scholarship with which he has documented his depiction of psychoanalysis, the critique can function as a

> powerful stimulus to hard thinking about the issues he has raised. I do not know for what more one could ask from a philosopher of science (p. 13).

The impact of "The Foundations of Psychoanalysis" in philosophers and psychologists was so great in 1984 that, in 1986, 40 authors published responses to the book in the "Open Peer Commentary" of the journal "Behavioural and Brain Sciences". But this atmosphere was not confined to the 1980s. The disposition in time of the responses to Grünbaum demonstrates his influence and actuality: some instances are in Edelson (1984), Hopkins (1988), Bucci (1989), Sachs (1989), Wallace (1989), Erwin (1996), Frosh (2006), Wallerstein (2006), Lacewing (2012a, 2013b, 2018), Azcona (2017) and Michael (2008, 2015, 2019). The philosopher has also given back many responses (Grünbaum, 1986a, 1993, 2007), causing the debate to be developed even further.

Many degrees of consent to the Pittsburgh's philosopher have been shown in those responses and, among psychoanalysis' advocates, there have been highly diverse strategies. Lacewing (2012a) divided these strategies into 4 groups: (1) to reject his assertion that psychoanalysis makes causal inferences (the hermeneutic line); (2) to reject his portrayal of the clinical method, for instance by saying that psychoanalysts have implicitly employed the epistemological cannons he endorses; (3) to reject that one should commit to these cannons, and to argue that psychoanalytic inferences are guided by alternative scientific codes; (4) to reject his assertion that these cannons should be *the only ones* to justify causal inferences.

Luyten, Blatt and Corveleyn (2006) indicated that Grünbaum's critique "implicitly or explicitly influenced" the division of the psychoanalytic community into "two radically different cultures regarding the nature and role of empirical research in psychoanalysis" (p. 578): a culture upholding that psychoanalytic research should be confined to the old case study method and another upholding that the clinical method does not satisfy the canons of science and so that psychoanalytic research should focus on experiments or quasi-experiments. But since this division existed for some time before Grünbaum, it would be more accurate to say that the philosopher influenced its *recent figure* or that he *reinforced* it in the turn to the 21st century; in any case, Grünbaum was a milestone in the history of psychoanalysis.

The depiction above may suggest that the debate on the Grünbaumian framing of the epistemological problems of clinical psychoanalysis is by now saturated. With surprise we see, though, that it is still very much alive in the English scholarship. The problems are still whispering new directions, the more so since the recent arguments related to it are especially thought-provoking.

The verdict that such problems must be further examined is even more pertinent to Brazilian philosophers and psychoanalysts. To them, Grünbaum is still a perfect manifestation of the oxymoron "illustrious unknown". 0,63% of the academic works in English with the words "philosophy", "science" and "psychoanalysis" in it include the word "Grünbaum", while only 0,14% of the works in Portuguese on "*filosofia*", "*ciência*" and "*psicanálise*" include it. In a sense, then, within the scholarship on the epistemology of psychoanalysis, the Professor is 4,5 times less important in Portuguese than he is in English. Could it be because he wrote mostly in English? Probably not: if we make the same comparison not to English but to Spanish scholarship, we may say that Grünbaum is more than two times less important in Portuguese than he is in Spanish2. Moreover, in the Brazilian scholarship of the last 16 years, thorough and primary analyses of Grünbaum's arguments can be counted in less than two hands; they are: Marinho (2006), Mezan (2006), Martins (2012), Davidovich (2014), Beer (2015), Kaszubowski (2016) and Pinto (2016). Of these, only Marinho (2006) and Kaszubowski (2016) make use of the arguments to come up with solutions for the epistemology of clinical research in psychoanalysis.

Although still today experts barely discuss what are the principles and models involved in generating, accepting, confirming, refuting and changing hypotheses in clinical psychoanalysis, the ones who do it depart, as a rule, from Grünbaum's critique. For our discussion about the possibility of producing knowledge from clinical-psychoanalytic evidence, the critique shall be, thus, a privileged pretext – a compass to start the exploration.

In other words, I shall address the problems within the epistemology of clinical research in psychoanalysis as they were circumscribed by Grünbaum. For this, I revisit Grünbaum's discrediting arguments, present some of the responses given to them and, inspired by these responses, develop at length some arguments for redirecting or resolving the contingent parcel of those problems. The overarching aim is to question the grünbaumian conclusion that it is impossible to do cogent psychological research with the data obtained in the clinical context.

Chapter 1 presents a panorama of the critique. Chapter 2 displays a sample of responses to it. Chapter 3 is a preparation for the rest of the thesis: it presents the presuppositions of my own responses to Grünbaum to be seen in Chapters 4, 5 and 6. Chapter 4 discusses the grünbaumian problems around the retrospective or *ex post facto* nature of causal inferences in clinical psychoanalysis. Chapter 5 discusses Grünbaum's argument that there is no way to free clinical-psychoanalytic data of the suspicion that it is probably contaminated by suggestion.

2For this calculation, I have consulted Google Scholar on the 30th October 2021.

Chapter 6, at last, faces the logical and empirical complications of the link that should exist between the therapeutic and the epistemic dimensions of clinical psychoanalysis.

Apart from the standard third person, I shall use both first persons ("I" and "we") and, more seldom, a second person ("note that…"). I do this to serve a style which I think is the most adequate to articulate the philosophical problems at stake – the so-called "classic style", where one writes as if one were giving the readers a tour, pointing at things that they have not yet noticed on the way and exchanging impressions with them about such things (Thomas & Turner, 1994; Pinker, 2015). "A writer of classic prose must simulate two experiences: showing the reader something in the world, and engaging her in conversation" (Pinker, 2015, p. 29). Let us start, then.

2 GRÜNBAUM’S CRITIQUE OF PSYCHOANALYSIS

2.1 “TRUTH IS NOT NECESSARY FOR CURE”: FREUD’S UNSOUND TALLY ARGUMENT

Grünbaum’s eminent critique on psychoanalysis (1980, 1984, 1993, 2015) can be schematically understood within the frame of the basic principle of eliminative inductivism: in upholding a hypothesis from data, a theorist must have reasons to believe that all the other hypotheses that could come from the same data are unlikely. Along these lines, Freudians could only present evidence that some of the patients’ most archaic and dramatic memories and fantasies are the causes of their actions, and that repression of these memories and fantasies is the cause of their symptoms, by eliminating the alternative explanations for the concurrence of these phenomena. One of the rival hypotheses psychoanalysis should demonstrate unlikely is that symptoms, cures, actions, memories and fantasies are caused by suggestion. Grünbaum defends, though, that no argument over the Freudian method of clinical investigation provided the tools for this task.

The philosopher of science claims to have derived the most fundamental instance of this flaw from the treatment of the problem of suggestion undertaken by Freud in one of his “Introductory Lectures” of 1917. He names it “the Tally Argument” (Grünbaum, 1980). To get to his point, Freud (1963a) stresses the empirical premise that “[suggestion in] hypnotic treatment leaves the patient […] unable to resist any fresh occasion for falling ill”, while suggestion in the analytic treatment is used to an overcoming of resistances through which “the patient’s mental life is permanently changed […]” (p. 451). Freud (1963a) goes forward and claims that, even though the psychoanalytic authority can easily make the patient support some particular theory, this would only affect

> his intelligence, not his illness. After all, his conflicts will only be successfully solved and his resistances overcome if the anticipatory ideas he is given *tally* [emphasis added] with what is real in him. Whatever in the doctor’s conjectures is inaccurate drops out in the course of the analysis [footnote omitted]; it has to be withdrawn and replaced by something more correct. (p. 452)

In this context, Freud (1963a) argues that the analyst’s conjectures should be dropped out if he or she is led to witness a false or non-psychoanalytic cure. Symptoms may be overcome before any analytical work is done, or they may return. In these cases, the interpretation given to the patient right before the bogus cure is not yet accurate or correct, have been part of a suggestive performance, and should be replaced by another interpretation.

The "Tally" in Grünbaum's dubbing refers to the term Freud uses to convey the accordance between the psychoanalytic interpretation and the mind of the patient. The word "tally" comes from the Latin "*talia*", a cutting from a plant, a twig. In age-old times,

> a tally was a simple wooden rod with notches cut across one of its faces to represent the amount of money owed or received. The rod was then split in two lengthways, both parties in the transaction receiving half. Brought together the halves corresponded exactly and were legal proof of the debt incurred or payment made. (Flavell & Flavell, 1995, p. 238)

The English words "tailor" and "retaliation" (as in the Babylonian law) also have the same origin[3]. An eye for an eye and a word for a mind. The analyst cuts a garment that must fit the patient. The rod of interpretation must have the same number of notches as the rod of the unconscious…

But Freud did not write in English; the word he used was "*übereinstimmen*", to accord with, to coincide with, to be conformed to, to correspond to. "*Erwartungsvorstellungen […] die mit der Wirklichkeit in ihm übereinstimmen*": "anticipatory ideas […] that accord with the reality in him [the patient]". "*Stimmen*" means to tune, to accord, to be correct as to something (and "*Stimme*" means "voice"…). The prefix "*über*" intensifies the prefix "*ein*", which means "into", suggesting an image, along with "*stimmen*", of one thing getting inside another and being fused with it.

Anyhow, by formalizing the Tally Argument, Grünbaum intended to spotlight Freud's thesis that a therapeutic effect on the analysed patient is evidence confirming the hypotheses the analyst enunciated to the same patient before – his thesis of the causal relation between truthfulness and therapeutic effect. To this, Grünbaum (1984) gave the name of Necessary Condition Thesis (NCT). It can, according to him, be split in two: (1) the thesis that the patient's correct insight on the dynamics of his or her character and neurotic afflictions is causally necessary for his or her healing; (2) the thesis that only the clinical method of psychoanalysis can draw these dynamics out and promote this insight to the patient.

The NCT would be one of the premises of the Tally Argument; the other is the evidence that the patient got better. The conclusion of the Argument could also be split in two: (1) if these premises are true, the interpretations given by the analyst are correct, that is, "tally with what is real" in the patient; (2) if these premises are true, only the analytic method could have healed the patient's neurosis. According to Grünbaum (1984), Freud would be saying in 1917 that,

---

[3] The Portuguese words "*talo*", "*talhar [a madeira]*", "*tal*" and "*[lei de] talião*" also probably stem from this Latin root.

> […] collectively, the successful outcomes of analyses do constitute *cogent* evidence for all that general psychoanalytic theory tells us about the influences of the unconscious dynamics of the mind on our lives. In short, psychoanalytic treatment successes as a whole vouch for the truth of the Freudian theory of personality, including its specific aetiologies of the psychoneuroses and even its general theory of psychosexual development (pp. 140-141).

As for any argument, for this one to be valid (in the sense of classical logic) the conclusions cannot be false if the premises are true. It is worthwhile to note that Grünbaum does not work with the idea that the Tally Argument is invalid, and anyone who applies this last definition to the Argument could not do otherwise. The scholar argues, actually, for the falsity of one of its premises, the NCT. This makes Freud's argument not invalid, but *unsound*, and an unsound argument "fails to establish the truth of its conclusion even if in fact the conclusion is true" (Copi, Cohen & McMahon, 2014, p. 31).

Table 1 - The Tally Argument

Premise 1: Necessary Condition Thesis. Premise 1.1.: The patient's correct insight on the dynamics of his or her character and neurotic afflictions is causally necessary for his or her healing. Premise 1.2.: Only the clinical method of psychoanalysis can draw these dynamics out and promote this insight to the patient.
Premise 2: The patient got better.
Conclusion 1: The interpretations given by the analyst are correct.
Conclusion 2: Only the analytic method could have healed the patient's neurosis.

Source: Elaborated by the author (2021) based on Grünbaum (1984).

It is worthwhile to note this because the NCT is an empirical premise – the critique, then, attacks the Tally Argument not on its logical or conceptual dimensions, but on its empirical dimensions. And he notably provides a sample of empirical researches which would show that components other than veridical psychoanalytic interpretations can be causes of qualified therapeutic effects. He concludes that "it becomes quite reasonable – though *not* compelling – to interpret […] [psychoanalysis'] therapeutic achievements as placebo effects" (p. 161).

Researches that compare the outcomes of different kinds of psychotherapies (Smith, Glass & Miller, 1980; Rachman & Wilson, 1980; Strupp, Hadley & Gomes-Schwartz, 1977) show that psychoanalysis does not supply greater therapeutic effects than its rivals (Grünbaum 1984). This success of other psychotherapies would clearly make NCT inconsistent: if one considers the hypothesis that the other psychotherapies do not supply veridical insights, then NCT's first part is false, since in this state of affairs a false insight would be causing a therapeutic effect; and, if one considers that the other psychotherapies do supply veridical insights, then NCT's second part is false (Grünbaum, 1980).

The classic psychoanalytic rebuttal to this is to question if the quality of the therapeutic effects was properly conceptualized, operationalized and measured in these researches. Psychoanalytic theory states that if a patient presents symptom remission but was not subject to the lifting and working-through of repressed conflicts, these will certainly cause another symptom. But Grünbaum (1984) also cites proofs of symptom remission without symptom substitution in behavioural therapies (Fisher & Greenberg, 1977; Munby & Johnston, 1980).

Of course, the research that could ultimately show that psychoanalysis is a placebo would compare it to placebo-therapies, and that is why Grünbaum pushes the epistemologist to the belief of a placebogenic psychoanalysis only in a reticent way. Anyhow, even if all the researches above were to show contrary results, the therapeutic superiority of psychoanalysis being thus demonstrated, it would still be necessary to demonstrate the causal link between therapeutic superiority and veridicality. Is it always therapeutic in the long run to know the truth about ourselves? Could self-deceit provide everlasting well-being? What would be the independent criteria to claim that an interpretation is true, so that psychoanalysis could show, without being circular, that this true interpretation is related to qualified therapeutic results? In sum, even though those results show that NCT is inconsistent or false, contrary results would not show it to be true or tenable.

Having the risk of placebo cures in psychoanalytic therapy and the resultant falsity of the NCT been argued for, two bad omens for the epistemology of clinical psychoanalysis are cast: 1) a therapeutic effect can no longer, as Freud intended, work as an indicator of the truth of the interpretation; 2) the evidence for the Repression Aetiology becomes poor.

Let us start with the second. The Repression Aetiology is the thesis that the repression of specific events is the necessary cause of the psychoneuroses. The relatedness between it and the NCT is direct. If repression of specific events is the necessary cause of any neurosis (the Repression Aetiology), then to undo repression by informing the patients of this event and of why it has been repressed – in other words, *to deliver a true interpretation on the cause of their neuroses* – will undo their neuroses (the NCT). The NCT is just the therapeutic consequence of the Repression Aetiology. However, if the NCT is false, therapeutic effects can no longer be regarded as evidence for this Aetiology; if the factor of psychoanalytic therapy may be a placebo, then the aetiology of the psychoneuroses may not be linked to repression[4]. The problem is that the Repression Aetiology is maybe the most fundamental thesis of psychoanalytic theory.

The psychopathological thesis is discussed by Grünbaum (1984) in the context of Breuer and Freud's developing of the pre-psychoanalytic cathartic method. He shows that the parents of psychoanalysis did write an argument for *ruling out* the possibility that their cathartic method was a medium for suggestive cures but that this argument was notwithstanding insufficient. Let us just clarify that, by the time of the cathartic method, the Aetiology was slightly different from the more general one stated in the last paragraph; by then, the thesis was that the repression of a traumatic idea hinders the discharge of its related affect, and that the cause of *both the genesis and the conservation* of a neurosis is this *affect strangulation fostered by repression*. Freud and Breuer's scientific defence of the thesis was this: after they enabled the patient to relive intensely the affect of a specific memory by hypnotically lifting a repression, only one specific symptom among the patient's symptoms, the one thematically related to the memory repressed, was to fade away; since symptoms were removed in isolation and according to the intervention made, one could not say that the cure was caused by the suggestive authority of the doctor rather than by the patients' remembering-and-discharging.

Grünbaum (1984) argues that this idea of a specific action advocates nothing against the possibility of suggestion. Along the cathartic therapy, by guiding the patient to recall when a distinct symptom first appeared, Breuer and Freud may have implicitly communicated that they attached potential therapeutic significance to the recall of an event thematically related to this symptom – and that this therapeutic significance would apply only to this symptom. If they wished to reject the hypothesis of placebo effect, Breuer and Freud should have compared their patients with subjects getting the same hypnotic nurture as their patients but without an abreactive lifting of repressions. Their "attribution of remedial efficacy" was "devoid of adequate evidential warrant" (Grünbaum, 1984, p. 180), which would push us to wondering if something else, still hazy for us, would be the necessary cause of a neurotic symptom.

Grünbaum (1984) argues, furthermore, that even if they had proven that a cathartic undoing of repressions is the cause of a cure, it would have been evidence for a thesis other than their Repression Aetiology: the thesis that the repression of a specific event and the

[4] Note that the Repression Aetiology is not demonstrated false if the NCT is demonstrated false. In that Aetiology, repression is a necessary, *but not a sufficient*, cause; it can be present without causing a neurosis. If something other than the lifting of repressions is causing the cure (i.e, the NCT is false), repression could still be a necessary cause of the neurosis in question.

strangulation of its related affect is the necessary cause of the symptom *conservation*, but not of its *origination*. The philosopher cites Morris Eagle, who remarks that Breuer and Freud's therapeutic conclusion is consonant with the thesis that the origination of the symptom is due to the conscious traumatic experience itself rather than to its repression, while repression (due to anxiety) merely conserves this symptom. Repression, in this possibility, is irrelevant to the genesis of the symptoms.

In sum, just as long as the NCT is false, Freud was not able to show the rival hypothesis of therapeutic suggestion to be unlikely. And, if there is the risk of therapeutic suggestion, then the vital Repression Aetiology is far from confirmed.

Now let us consider the following epistemological consequence of the unsoundness of the Tally Argument: a therapeutic progress in the patient cannot be an indicator of the truth of the interpretation given before by the analyst. Moreover, according to Grünbaum (1984), clinical psychoanalysis would not have a truth-ensuring device as fundamental as the debunked Tally Argument and its "bold lawlike premise" (p. 127).

For him, it would have such relevance among Freud's arguments for the conditions under which the new science could produce cogent theses that it would cope alone with the entirety of the epistemological challenges of clinical psychoanalysis. Freud's NCT would be a kind of epistemological insurance. It would supplement the holes in causal inference and certify that the clinical data are not contaminated by suggestive moves of the analyst, warranting savagely for the truth of the analyst's interpretation:

> As long as Freud saw himself entitled to adduce his NCT, he felt able – with a *single* stroke – to rebuff the twin suggestibility attacks on the dynamics of his therapy as well as on the cognitive reliability of the clinical data gathered by psychoanalytic investigation. Precisely because the crucial NCT premise of his Tally Argument declared correct *etiologic* insight to be *therapeutically* indispensable, this argument legitimated Freud's confidence in the following proposition: his *retrospective* clinical ascertainment of the aetiologies of the psychoneuroses and of the causes of normal personality development *by the psychoanalytic method* was not vitiated by pitfalls of causal inference such as *post hoc ergo propter hoc*, but rather was methodologically sound. […] the attribution of therapeutic efficacy to the lifting of repressions was indeed the epistemic basis for endowing Freud's method of free associations with the ability to certify causes (e.g., pathogens). (p. 146)

Moreover, such testing power of the healings' setup would have been accepted throughout the decades of the 20th century, to contexts presumably more methodologically developed than Freud's, as may be demonstrated by quotes from Waelder (1962) and Basch (1980) (Grünbaum, 1984, pp. 146-148). As the philosopher concludes in the most forceful way: "though it is widely overlooked, *the attribution of therapeutic success to the removal of* *repressions not only was but remains to this day, the sole epistemic underwriter of the purported ability of the patient's free associations to certify causes*" (Grünbaum, 1984, p. 185).

We have seen the fall of the Tally Argument. That is why, he asserts, we must consider the possibilities that interpretations made in clinical psychoanalysis: 1) are based on contaminated (suggested) data and/or; 2) are enacting logical fallacies. These two possibilities set by the philosopher will be discussed in the next two sections.

2.2 "THE CLINICAL EVIDENCE MAY BE CONTAMINATED": THE CHARGE OF COGNITIVE SUGGESTION

2.2.1 Per Via di Porre: "Analysis Adulterates the Patient's Flow"

Grünbaum's charge that clinical evidence may be contaminated has more than one facet. Besides the expectancy effects that may be caused by the analyst, there are those that may be caused by events outside the session: since Freudian hypotheses are well adhered to Western culture, it is probable that patients try to guess from books, papers, Wikipedia, TV shows, hearsay, etc., how these hypotheses would explain their suffering. In Grünbaum's words:

> […] does the patient's adherence to the fundamental rule of free association indeed safeguard the *causally uncontaminated* emergence of actually existing repressed wishes, anger, guilt, fear, etc.? Or is the process of association contaminated by the analyst's injection of influence of one sort or another? Clearly, the answer will depend, at least partly, on just what the analyst does while the patient is busy fulfilling his share of the analytic compact. This answer is also likely to depend on the antecedent beliefs that patients going into analysis bring into the analytic situation, for many an intelligent analysand is consciously aware of the sort of material that his Freudian therapist does expect from his free associations. For example, male patients are expected to have repressed castration anxiety, and females are to have unconscious penis envy (Grünbaum, 1984, pp. 208-209).

We must also consider the strength with which the analyst may contaminate the patients' mental products: the analyst may do it mildly by stressing only the patients' products the theory indicates beforehand as relevant, thus inducing their reports of some kinds of products rather than of others; or strongly, by making inapposite and baseless interventions, thus adulterating the patient's cognitive flow. Accordingly, our last citation proceeds thus:

> While dealing with our question, we shall need to be mindful of another, since it likewise pertains to the epistemic effects of the analyst's intervention: if a plethora of unconscious thoughts surface, by what criteria does the analyst decide when to call a halt to the surfeit of associations, while investigating parapraxes and dreams? Hence, let us canvass in what respects overt and subtle interventions by the analyst affect the data yielded by the patient's associations (Grünbaum, 1984, p. 209).

Conceptually, one partition – cognitive suggestion caused by the analyst *versus* caused by extraclinical cues – may be crossed with the other – reinforcement *versus* adulteration; that is, the latter two processes may be present inside or outside the clinic.

Freud once convoked Leonardo Da Vinci's discussion on painting and sculpture to contrast his therapy with suggestive ones (Freud, 1953a). Suggestive therapies would be like painting – would work *per via di porre*, "by way of laying" –, for they would, as it were, *add matters on top of* the patient's psyche, here metaphorized as the canvas, so that the pathogenic idea would not come to the fore. Psychoanalysis, on the other hand, is analogue to sculpture, which works "*per via di levare*", to wit, "by way of extracting" the useless matters hiding the handsome figure that lies underneath. The two ways indicated by Grünbaum along which the analyst may contaminate clinical evidence, I suggest, are on the lines of this metaphor of Freud: the analyst's selection bias, and the discriminating reinforcement thereby resulted, work *per via di levare*, by taking away the excessive, while the analyst's meddlesome, out-of-nowhere, hasty interventions work *per via di porre*, by "dyeing" the patient's associations[5]. We shall start by presenting Grünbaum's arguments on the latter, that is, on cognitive suggestion as sheer adulteration.

In Chapter 4 of *Foundations*, Grünbaum reminds us of a famous instance of motivated forgetting in Freud's work, the forgetting of the word *aliquis* by a man identified as AJ. Resenting the fact that he was discriminated for his religion, the Jewish man made a foiled effort to quote Virgil's "*Exoriare aliquis nostris ex ossibus ultor*", a line that translates "Would that someone arise from our bones as an avenger!". AJ asks for Freud's help, at which Freud restores "*aliquis*" to the line. Then, the young man, "who was familiar with […] psychoanalytic writings" (Grünbaum, 1984, p. 190), asked Freud to explain the memory lapse; standing before the urge, the doctor instructed him to free-associate to the restored word. AJ's associations led them to the fear that a lady with whom AJ had had sexual relations had been impregnated. Freud concluded that this was the cause of AJ's lapse: an unconscious fear that the affair had yielded a progeny had *conflicted* with the overt will to have a progeny to avenge the hardships suffered by the Jews.

Grünbaum notes that, in this case, Freud did not let AJ's associations flow freely: "[…] it was Freud's correct statement of the quotation from Virgil, but *not* the undoing of AJ's

[5] Considering that I am Grünbaum's spokesperson in this chapter, I should say why it is interesting to view the charge of cognitive suggestion through this Freudian metaphor even though Grünbaum never proposed such a thing: it calls one's attention to suggestion's unsuspected pervasiveness and to Freud's attempt to nullify this pervasiveness through a questionable contrast.

repression of his pregnancy fear, which served to *remove* the mnemonic lacuna of *aliquis* in the young man's awareness" (p. 193). Well, one can get to any content from a given stimulus, if only no limit is posited to the allowed number of associative chains[6]; from the word Freud restores, AJ arrived at his most recent worry, which is only trivial. If the repressed fear were the necessary cause of AJ's forgetting, then his awareness of this fear after free-association would have, naturally and immediately, opened the mental path to *aliquis*; so Freud should have performed this test, letting him associate to any other element related to the forgetting and checking if the word became manifest anew[7]. On this line of argument, Grünbaum cites Timpanaro (1976), to whom "Freud did not hesitate to intervene occasionally in the flow of AJ's associations" and "thereby subtly steered the associations in a manner akin to the Socratic method of eliciting answers to leading questions" (p. 192).

The studies by Shapiro and Morris (1978) are also brought to benefit the argument. According to the authors, therapists may unintentionally convey information to patients, such as expectations and attitudes. This confines therapists to find just bogus confirmations for their psychological hypotheses; not knowing they are bogus, the therapists strengthen their assumptions. There comes, henceforth, a circular and long chain of suggestibility.

> This epistemological difficulty is […] compounded by the operation of those phenomena that Freud termed "countertransference" phenomena after shrewdly discerning them: the distorting effects of the therapist's feelings toward the patient on the accuracy of the former's perception of the latter's behaviour (S.E. 1910,11:144-145) (Grünbaum 1984, p. 212).

In Chapter 4 of *Foundations*, Grünbaum presents a psychoanalytic ideal-instance of probable adulteration of data through suggestion. The instance came from a publication by "two respected teaching analysts", Gertrude and Rubin Blanck (Blanck & Blanck, 1974), and concerned "penis envy", the Freudian hypothesis that female subjects typically envy male sexual functioning and genital architecture. It was a recent publication for when he was writing *Foundations* and it markedly exhibited a process of interpretation in clinical psychoanalysis in which no knowledge is really questioned or produced, only (forcefully) applied. It was ideal, hence, for demonstrating that

> even nowadays *some* analysts do intervene unabashedly in the associations of their patients. […] it is then pretended that the products of the ensuing associations are the subject's previously unconscious ideas, which have surfaced in unbiased fashion. Thus, it would be wrong to suppose that Freud's own "activist" handling of a patient's associations when giving interpretations is a thing of the past in psychoanalytic research and practice (Grünbaum, 1984, p. 212).

---

[6] This may be called thematic liability or elasticity. In this thesis, we will further explore critical arguments related to associative liability below and in chapter 4.

[7] Grünbaum (1984) says that even this would not qualify for good evidence, but at this point develops no further on the alternative hypotheses that would still have to be eliminated to make Freud's hypothesis likely (see p. 193).

Although italicizing the word "some" above, thus acknowledging a contingency, the philosopher argues that the latter does no harm to his charge. He tells us that the analyst Benjamin Rubinstein corrected him for taking Blanck and Blanck's or even Freud's conduct as the model in contemporary psychoanalysis; according to Rubinstein, there are contemporary analysts who are vigilant about "overt, covert, and even unconscious suggestions on the patient" and struggle to be far less "activist" than those other analysts. But we would have to be sure, he contends, that this struggle does have the effect it betokens; the mere effort to avert "unintended yet potent suggestive influences" should be shown successful or it would be just hollow epistemic sophistication (Grünbaum 1984, p. 212).

A contingency could appear here as well: some analysts would show this success, others would not. Thus, if Grünbaum want us to consider this attack as a "check-mate", we must read between his lines: "The effort is nice, but the clinical method does not have the logical elements that would be necessary to demonstrate there was no suggestion". Regrettably, he does not invest in this supposed point here. Instead, he goes back to a good-enough contingency: "In any case, there is telling evidence in the current literature that the effort is not even properly made among some influential practitioners" (1984, pp. 211-212).

Let us then finally present his debate of Blanck and Blanck's case report. A female patient had anxiety related to her skin and general image and started therapy with a woman psychoanalyst. Grünbaum spotlights three strings of interpretation the analyst gives to the patient:

Table 2 - Strings of Interpretation in Blanck and Black's (1974) Case Report

String	Patient	Analyst
1	"Today I feel that I should see a dermatologist about my skin"	"You think constantly about your appearance because you are not sure that your body is always as it should be" (p. 320).
2	"Sometimes I think I look better than at other times"	"You are not always certain that your body is the same"
	"I always feel there is something wrong"	"This is a classical phallic statement"
	Dreams about one of these interpretations	Thinks the interpretation was thereby confirmed (p. 321)

3	(In response to a question of the analyst) "Yes, men are always more admired"	"They have something more to be looked at"
	"Oh, you mean a penis"	Thinks: "When the patient says, 'You mean' to her own association, it is a projection which represents the last defence against allowing the thought into consciousness" (pp. 321-322).

Source: Elaborated by the author (2021) based on Grünbaum (1984).

On string 1, Grünbaum (1984) makes us note the fact that there was nothing in the patient's line that avowed or implied that lack of sureness: "this assertion is not based on data furnished by the patient but originates in the envy hypothesis" (p. 213). Moreover, since not to be sure that our body is as it should be is a very common state of mind, the interpretation is likely to be accepted. By stating trivial things such as these, the analyst paves the way to the non-trivial hypothesis of penis envy.

The philosopher comments nothing on string 2, maybe because he thinks its suggestive character is self-evident. But he comments on the spurious confirmations promoted in string 3. Once the patient has been "initiated into engaging in penis talk about herself", she may take the discovery of her penis envy as authentic and go on "furnishing her analyst with ever more spurious confirmation" (pp. 213-214).

After the case report, he cites Emanuel Peterfreund (1983) to concede once more that there are some rational psychoanalysts who do not find an investigative spirit in the procedure of the Blancks. Resonating Rubinstein, Peterfreund says that the Blancks did not take "penis envy" as a hypothesis for which evidence may or may not show up, but rather indoctrinated the patient into its truth; this would be a "stereotypic approach" in which the results of the investigation are known beforehand and the task of the analyst is to make the patients come to these results, so to say, by themselves. Grünbaum (1984) insists: "Though Peterfreund's own 'heuristic' approach to clinical investigation and therapy does avoid *some* of the pitfalls of the stereotypic one, he does not, alas, come to grips with the fundamental difficulties besetting intraclinical validation, as set forth in this book" (p. 215). His conclusion would have been more telling, though, if he had stated which pitfalls Peterfreund's approach is able to avoid, and which difficulties set forth in *Foundations* outweighs the efforts of that analyst.

2.2.2 Per Via di Levare: "Analysis Reinforces Parts of the Patient's Flow"

The patients under psychoanalytic work must associate to their mental states. But for how long? How many links are necessary until the memories and fantasies that caused symptoms, dreams and parapraxes can get a lane to the surface? How does the analyst identify what is relevant and what is not in the patient's range-varying string of spoken thought?

An orthodox psychoanalyst may rebut the significance of such problems by pointing to the principle that *every* string of associations triggered by a mental state *reproduces in reverse* its former causal unfolding. But even this hypothetical psychoanalyst would concede that there may be censure operations on many moments of a string of associations. Well, that would not be a problem for this analyst, for theoretical principles for which Freud presented support – for instance, the principle that sexual conflict plays a part in every mental suffering – could guide a selection of relevant elements in the clinical material in spite of resistance. On this stance, *everything* the patient says under the fundamental rule is relevant, even though some of these things are masks and some are grimaces.

But that definitely is a risky bet. The patients' psyche would *not* become strictly conservative, confined to neural circuits that are highly facilitated (to borrow the terms of Freud's *Project for a Scientific Psychology*), with a diversion here and there to evade disgust, *only because* it obeys the demand to depict everything that passes by. In the human psyche, the tendency to be conservative coexists with the tendency to make new connections, to try new circuits. Thus, the problem of identifying psychoanalytically relevant, cause-related, associations and random, groping, associations persists. Let us name it "the selection problem".

It is worsened by what may be named "thematic elasticity" or "thematic liability": the fact that it is always possible to associate one theme to another, especially when the number of medium links between those themes is not predetermined. Standing before this Cartesian demon, could psychoanalysts identify what is relevant in the patients' associations through theoretical principles? If so, analysts would basically perform a *petitio principii*, for they would infer the causes of the patients' mental states way before the latter have presented any data. (Let us note that this conclusion also applies to the orthodox analyst we imagined above). Standing before the selection problem, we only find a "selection bias".

The pervasion of selection bias in clinical psychoanalysis is part of Grünbaum's charge of cognitive suggestion as well, since this bias is also a promoter of suggestion effects – an understated yet powerful promoter. It also may inform the patient, now in the most implicit manner, what the psychoanalyst expects to hear. Even though it does not impose alien stuff to the patient's *Lebenswelt*, it reinforces some of its elements rather than others, contaminating it just the same.

The philosopher claims that an influencing selection may operate in two complementary ways. The "go on" or the "tell me more" of the analysts implies that something important has not yet been said, while a curious "hum" or a "stop right there" followed by a request for elucidation implies that what has just been said is perhaps a nugget of gold or an Arkenstone. If inventive patients could associate long enough, they could yield almost any kind of content: "thoughts about death, God, and indeed cabbages and kings" (Grünbaum, 1984, p. 209). This thematic elasticity forces the analysts to enact a question-begging selection bias. They may interrupt the patients' spontaneous causal dynamics at a definite point or, if the patients falter, demand that they go further – all in function of what they consider relevant or irrelevant material. How would they manage "not to load the dice by ever so subtly hinting to the patient what kind of material [...] [they expect] to emerge?" (Grünbaum, 1984, p. 209).

Grünbaum (1984) imagines an alternative version of AJ's vignette that would demonstrate that thematic elasticity and a manipulation of the range of the association string result in a selection bias:

> Suppose that Freud had allowed AJ to continue well past the disclosure of the pregnancy fear. Perhaps it would then have emerged that AJ's parents had taught him early that the Romans had crucified Jesus, but that Christians had then unfairly blamed the Jews for deicide. It might furthermore have emerged that AJ had repressed his ensuing hatred of the Romans when Virgil, Horace, and other Roman poets were shown great respect in his Austrian educational environment. […] Would AJ's hypothesized repression of his hatred for the Romans not have had greater thematic "suitability as a determinant" of his *aliquis* slip than his anxiety about the pregnancy, even though the former assumedly emerged only later in his associative chain? After all, Virgil was a Roman, and AJ was citing the line from the Aeneid to express his conscious resentment of Christian anti-Semitism. What a golden opportunity to punish the unconsciously resented Romans simultaneously by spoiling Virgil's line! (pp. 209-210)

In the context of this argument on suggestion *per via di levare*, Grünbaum (1984) cites the paper "Limitations of Free Association", by the analyst Judd Marmor (1970). After an account of the experimental studies by L. Krasner, W. K. Kaplan and K. Salzinger, Marmor (1970) would conclude that the most delicate gestures of an analyst are able to convey expectations and alter the otherwise free-spirited flow of the patient, gestures such as a glance, a lift of the eyebrow, a shake of the head, a shrug of the shoulder, "uh-huhs", silences, tones of voice, etc.. They would be like radio signals to the patient, reinforcing some of its mental products and inhibiting others. Free-association would not be free, as Freud would have it.

Therefore, according to Grünbaum (1984), the analysts' selective questions, punctuations and reactions as well as their rough inputs to the patients' flow, discussed in the last section, risk to contaminate clinical data. As a consequence, any clinical investigation in psychoanalysis is invalidated from the start: before any flaw with data analysis, there is a flaw with the data itself, with its gathering or its integrity. To extend Freud's metaphor, the analyst is like an archaeologist who, when digging up an object, spills some reactive substance onto it or does not care to check if there are other objects in the encircling lands that would shift an immediate interpretation of it. Moreover, this risk of cognitive suggestion makes us ignorant of whether patients are producing just the kind of phenomenological data that confirms the hypotheses of their analysts, a fact that makes the psychoanalytic corpus of evidence awfully questionable.

2.3 "CLINICAL INFERENCES ARE WEAK": THE CHARGE OF BAD INFERENCE

2.3.1 The Post Hoc Ergo Propter Hoc Fallacy and the Sampling Bias: "Retrospective Clinical Inferences Are More Limited than Freud had Thought"

As hinted before, Grünbaum's arguing moves are comparable to check-mate moves in chess. His logical inflexions regularly look like this: "Let us suppose, for argument's sake, that charge x is demonstrated false or improbable (or, fantasy-wise, correctable); even then, we would still have another charge y...". His readers are urged to make some moves of challenge, but they end up facing a resignation-inspiring endgame. This is his inflexion betwixt his charge of data contamination and his charge of weak inference. Even if the clinical data were uncontaminated and reliable, he says, inferences in clinical psychoanalysis would be weak: they would be inductive arguments in which the truth of the premises does not imply a high probability of the conclusion. (It also works backwards: "even if inferences in clinical psychoanalysis were strong, its data would be polluted").

Grünbaum (1983) considers the way Freud infers the necessary cause of obsessional neurosis from the Rat Man case as the standard way to infer the cause of a mental pathology in psychoanalysis. He points out that Freud's evidence that a premature sexual incident is the necessary cause of obsessional neurosis comes down to the Rat Man being obsessive *and* having repressed a premature sexual incident. One or some cases, one co-incidence, and Freudians would find a necessary cause to a psychopathology. But this inductive argument, says Grünbaum (1983), is weak.

From a neo-Baconian or Millian standpoint, he develops two arguments to undermine this method of inference:

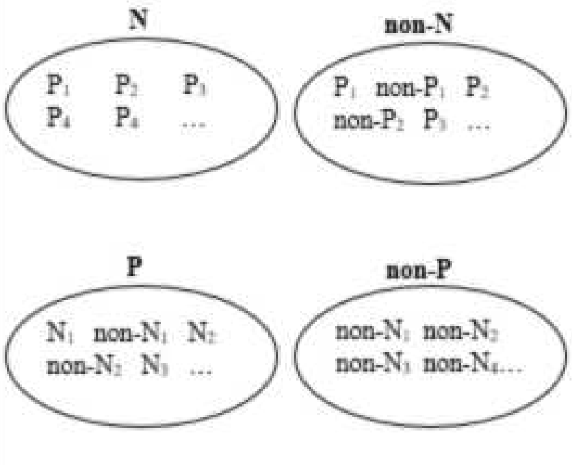
1. The mere co-incidence of events is no evidence for a necessary cause. We have evidence for a necessary cause only if we show, apart from this co-incidence, a co-non-incidence, namely: that in every case in which the hypothesized cause is absent, the hypothesized effect is also absent.

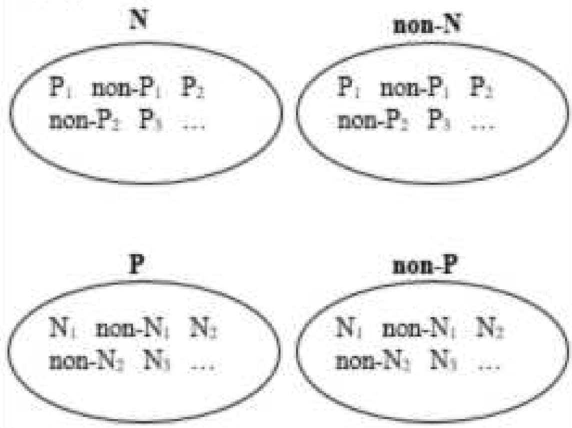
2. In a clinical research, the research subjects are not randomly chosen; they can only be persons who are ill and/or who went after a therapy. In clinical psychoanalysis we are never able to compare the neurotics with the non-neurotics, but only one category of neurotics that went after therapy with all the other categories of them that went after therapy. Hence, even if the condition of 1 were present, clinical psychoanalysis could infer only neurosis "choice" but could not infer anything related to its genesis.

To clarify argument 1, let us state Freud's argument once more: every case of the psychopathology in question (N, for neurosis) presents (a report of) a specific kind of event (P, for pathogenic event), so P is the necessary cause of N. Let us now inspect the logic of necessary – that is, necessary *but not sufficient* – causes. If P is necessary for the advent of N, then: N cannot advent without P, so a non-P will always cause a non-N; *and* it is possible for a P to cause a non-N, if only the sufficient set of causes for N is absent. A scheme comparing the possible outcomes of P being the necessary cause of N with P having no causal relation to N may help us monitor the application of this logic:

Figure 1- The Possible Outcomes of P Being the Necessary Cause of N with P Having No Causal Relation to N

If: P is the necessary cause of N.	If: P has no causal relation to N.
Then the following cases are possible: -Ps with Ns. -Ps with non-Ns. -Non-Ps with non-Ns.	Then the following cases are possible: -Ps with Ns. -Ps with non-Ns. -Non-Ps with Ns. -Non-Ps with non-Ns.
So, only the following sets are possible: N = {P_1, P_2, P_3, P_4,...}. Non-N = {P_1, non-P_1, P_2, non-P_2, P_3,...}. P = {N_1, non-N_1, N_2, non-N_2, N_3,...}. Non-P = {non-N_1, non-N_2, non-N_3, non-N_4,...}.	So, all the following sets are possible: N = {P_1, non-P_1, P_2, non-P_2, P_3, non-P_3,...}. Non-N = {P_1, non-P_1, P_2, non-P_2, P_3, non-P_3,...}. P = {N_1, non-N_1, N_2, non-N_2, N_3, non-N_3,...}. Non-P = {N_1, non-N_1, N_2, non-N_2, N_3, non-N_3,...}.





Source: Elaborated by the author (2021) based on Grünbaum (1983).

Now we are apt to understand Grünbaum's (1983) premises for argument 1:

> To support Freud's aetiologic hypothesis that P is causally necessary for N, evidence must be produced that being a P makes a difference to being an N. But such causal relevance is not attested by mere instances of N that were P's, i.e., by patients who are both P's and N's. For even a large number of such cases does not preclude that just as many non-P's would also become N's, if followed in a horizontal study from childhood onward! Thus instances of N that were P's may just happen to have been P's, whereas being a P has no aetiologic role at all in becoming an N.
>
> […] to provide evidence for the causal relevance claimed by Freud, we need to combine instances of N that were P's with instances of non-P who are non-N's. And indeed, since he deemed P to be causally necessary for N—rather than just causally relevant—his aetiology requires that the class of non-P's should not contain any N's whatever, whereas the class of P's is to have a positive (though numerically unspecified) incidence of N's. (Grünbaum, 1983, p. 332)

Yet an N–P instance is logically equivalent to a non-P–non-N instance: if, having all the documented psychoanalytic cases in the history of humankind at hand, we witness that *every* case of N is a P, we can be sure that every case of non-P is also a non-N. But, ordinarily, we are not in this abundant circumstance. So, let us suppose we have 30 cases of obsessional neurosis, and that all of them display a premature sexual episode that was repressed. If we want to present a sharper evidence for Freud's hypothesis, what should we prefer: 1 extra case of N–P or 1 odd case of non-P–non-N? The latter, Grünbaum (1983) would argue. For, here, although 1 more N–P may enhance the probability of a next non-P–non-N case, it is not a proper non-P–non-N case. And, since the latter is the necessary evidence for our hypothesis of necessary cause, it must be presented. The issue "is *not* merely to provide evidential support for 'All non-P's are (will be) non-N's', or for its logical equivalent, by some instances or other", but "to furnish evidential support for the (strong kind of) causal relevance claimed by Freud" (Grünbaum, 1983, p. 333).

But, in the Rat Man case, Freud fundamentally regarded a mere sequence of events as evidence for a causal link, enacting the fallacy of *post hoc ergo propter hoc*, Latin for "after this, therefore because of this". The antidote for this fallacy is Mill's Method of Difference: we must check if the outcomes of the presence of the hypothesized cause are amply different from the outcomes of its absence. In everyday life, I may forget to use the Method of Difference, for example, when I infer that coffee cures cold because my cold was cured days after I drank some cups of coffee; but have I checked whether or not the colds of everyone who did not drink coffee were cured as fast as mine? For the philosopher, psychoanalysts also omit this kind of question, thus committing a major logical mistake (Grünbaum, 1983).

Nevertheless, even if we had, say, 500 psychoanalytic cases of N–P and 500 psychoanalytic cases of non-P–non-N, would we have a proper evidence for Freud's hypothesis? Not quite. This takes us to argument 2. In Freud's hypothesis, P is the necessary factor for the advent of obsessional neurosis *rather than any other mental condition*; but that evidence confirms the hypothesis of P being the necessary factor for the advent of obsessional neurosis *rather than any other mental suffering*, for

> even the non-N's of this design are presumed to be afflicted by some neurosis other than N. [...] persons who have practically no neuroses of any sort are hardly available to analysts in sufficient numbers to carry out the putative retrospective determination of whether they were non-P's or P's. [...] as Mr. Blake Barley has noticed, [...] these combined instances lend credence only to the hypothesis that, within the class of neurotics, P is aetiologically relevant to N. But these putative combined instances do not support the Freudian claim of such aetiologic relevance within the wider class of persons.
>
> In short, the Freudian clinical setting does not have the epistemic resources to warrant that P is neurotogenic. (Grünbaum, 1983, pp. 337)

Alas, check-mate! This would be epistemically alarming *even if* we assume that the retrospective investigation of clinical psychoanalysis is not memory-corrupting, that is, even if we credit the existence of non-Ps in the infancy of non-Ns from the reports of adult patients.

2.3.2 The Thematic Affinity Fallacy: "Mere Thematic Affinities Are Not Evidence for Causal Links"

A reader of psychoanalytic cases tends to be amazed and disturbed with how much their authors argue for causal connections betwixt mental products that are disparate in time and space simply because they are similar in form and circumstance. Thematic affinities bewitch us – every one of us is aesthetically provoked by metaphors, for example – but not every one of us would find rational to have them as the premises of a scientific paper on psychopathology. Besides, although we could say Freud's tendency to take them soberly was born with his biological notion of primary process, this tendency has become an argument for the hermeneutic account of psychoanalysis; if it is an arsenal for the "scientophobics", to regard thematic affinities as scientific premises is likely to be absurd.

Apart from the *post hoc ergo propter hoc* fallacy and the sampling bias, Grünbaum (1984, 1989) identified in psychoanalytic inference a problem related to the thematic affinities one can witness in every published psychoanalytic case. To be short, he argued that since Freud used to infer causal links between events *only because they were similar*, he commits a univocal fallacy.

Grünbaum (1989) discusses everyday cases in which a causal inference from a thematic affinity could not be strong unless a specific background knowledge or evidence were considered (Grünbaum, 1989):

1. A professor infers a student to have committed plagiarism since the latter's paper agrees verbatim with an old encyclopaedia. The professor is justified from the background knowledge that an independent production of the encyclopaedia's exact wording is highly unlikely.
2. A tourist infers that another person was present in a desolate beach since there are shapes on the sand matching the shapes of left and right shoes. The tourist is justified from the background knowledge that foot-like shapes on the sand could hardly have been caused by the action of the wind.
3. Visiting a Frank Lloyd Wright's house was causally relevant to dreaming with a house just like it the night after, assuming that pictures of it were never seen beforehand.

In these three inferences, the background knowledge would tell us that one event *makes a difference* to the other. But a background knowledge can also tell us the opposite, thus illegitimating a causal attribution (Grünbaum, 1989):

1. Visiting *some* house one day is not causally relevant to dreaming with *some* house the night after, since houses are everywhere in our waking lives.
2. Similarly, historians should *not* look for a causal linkage in the facts that the friends Thomas Jefferson and John Adams died on the same day *and* that this day was the 50th anniversary of the Declaration of Independence, written by one and drafted by the other. Not even an *indirect* causal linkage, such that the demises were *effects of the same cause*, should be pondered here. Background knowledge would tell historians that the coincidence could only be explained by separate causal chains[8]. To invigorate his argument, Grünbaum (1989) cites the philosopher Elliott Sober (1987) who, rejecting Hans Reichenbach's and Wesley Salmon's eminent principle, defended that correlations are *not necessarily* evidence that a common cause is at play.
3. In support of this thesis, Sober (1987) presented a case in evolutionary theory, the inference by "cladistic parsimony": a similar characteristic between two species is evidence of a common ancestor when this characteristic is an "apomorphy", but not when it is a "plesiomorphy". Thus, a mere thematic affinity between two species does not justify a claim of common origin (Grünbaum, 1989).
4. Another of Sober's refuting cases is the rise, at a similar rate, of the cost of bread in England and the sea level in Venice. This affinity is barely a token of a common cause, let alone of a more direct causal linkage.
5. The last of Grünbaum's (1989) non-psychoanalytic example of a weak inference from a thematic affinity comes from the magical thought of a long bygone medicine: in the 16th century, Paracelsus professed that liver disorders can be cured with an herb shaped like a liver (Hacking, 1975). Today we are able to notice how irrational this kind of inference is, but Grünbaum claims that "its logical defects are no worse than those of causal inferences from mere thematic connections, which abound nowadays in some theorizing about human behaviour" (p. 489). At this juncture it is foremost psychoanalysis that Grünbaum has in mind.

---

[8]Having introduced the concept of common cause, Grünbaum (1989) gives one more example of strong inference:

> On the other hand, a common cause is indeed implicated in the familiar fact that the probability of a storm occurring soon after a sudden barometric drop is appreciably greater than the probability of a storm in general: A pressure drop over a wider area is the common cause of both the sharp barometric drop and the storm, which are thereby linked indirectly. (p. 488)

Freud's causal inferences from thematic affinities, says Grünbaum (1989), would use as premises *only* the thematic affinities. "Freud should surely *not* be faulted for asserting, in principle, that some mental events can be linked *both* thematically *and* causally, though he mistakenly claimed entitlement to *infer* the latter linkage from the former alone" (p. 491).

Freud would have inferred that an affect elicited by a childhood event is the cause of the Rat Man's obsessive symptoms *only* because both instances showed the same themes. The patient in question had obsessive fantasies that his father and a woman he loved would be subjected to a military torture involving rats, of which he had heard during military services. The patient also told of when he was punished by his *father*, a *military* man, for having bitten his nurse *like a rat* in his childhood, and the *anger* he felt at the time. Freud concludes that these affinities are tokens of a causal link. Grünbaum (1989) replied that they are not, for here there would be no background knowledge ruling out the hypothesis that the two father-military-rat-anger themes are parts of different causal chains – in other words, that the themes' repetition is a coincidence or happenstance.

From the argued missteps of Freud, Grünbaum (1989) draws "a two-fold moral for the human sciences": "1) Let us be alert to thematic connections, but beware of their beguiling causal pitfalls; *a fortiori*, 2) narratives replete with mere hermeneutic pablum are explanatorily bankrupt; at best, they have literary value" (p. 492).

2.4 THE FUTURE OF CLINICAL PSYCHOANALYSIS

For the sake of conclusion, let us recapitulate the points and the structure of the critique we have been committed to analyse and let us state, at last, what the author of this critique finds to be its implications for clinical psychoanalysis. According to Grünbaum:

1. The argument "if a psychoanalytic patient got better, then there has been, during his or her psychoanalytic process, a correct insight on the dynamics of his or her character and afflictions" (the Tally Argument) is *unsound*, since one of its premises, the thesis that this insight is necessary for that cure (the Necessary Condition Thesis, or NCT), is *false*. The premise/thesis is empirically false because studies have shown that the therapeutic result of psychoanalysis is not superior to the result of other psychotherapies (Smith et al, 1980; Rachman & Wilson, 1980; Strupp et al, 1977). Such fact would be well explained by the hypothesis that cures in psychoanalysis (and in other psychotherapies) are placebos.
2. Given the fact that other epistemological defences of clinical psychoanalysis do not share the eminence of the Tally Argument, its unsoundness has two epistemological consequences: the

impoverishment of the evidence for the Repression Aetiology – since something other than the lifting of repressions may cause a cure – and the incapacity of a therapeutic effect to serve as an index of truth.

3. Given this incapacity, we must consider the possibilities that interpretations made in clinical psychoanalysis are based on data contaminated by the suggestive moves of the analyst and/or are instances of logical errors.

4. The documented practice of clinical psychoanalysis shows that it is likely that cognitive suggestion pervades it. Historically, analysts tend not to consider their "invasion" of the patient's associations or their selective reinforcement as a problem in this respect. Analysts seem to promote expectancy effects in the patient; in any case, analysts have never developed a device to rule these effects out. This implies that the primary data of clinical psychoanalysis is unreliable, hence that its hypothesis confirmations may be spurious.

5. Even if we supposed that this alleged evidence is reliable, the documented practice of psychoanalysis indicates that its clinical inferences are weak, namely, that its clinical conclusions do not logically follow from its clinical data. These inferences exhibit the *post hoc ergo propter hoc* fallacy, the thematic affinity fallacy and a sampling bias (the use of a convenience sample).

For Grünbaum (1984, 1993), the upshot of the diagnosis that the clinical method of psychoanalysis is devoid of cogency is the assertion that it is unfit to test or prove hypotheses, especially causal ones. Therefrom, he also affirms that its classical hypotheses born out of the clinical context must be discredited until an experimental or epidemiological method – since they conform to Millian standards of scientific investigation – is used to confirm them. Along these lines, the clinical method of psychoanalysis would not be entirely disposable: although it would not be fit for testing hypotheses, it would be a competent way of *generating* hypotheses. According to the philosopher of Pittsburgh, then, in spite of every charge, the couch would still have an important *heuristic* role in psychological science.

3 THE RESPONSES TO GRÜNBAUM

3.1 INTRODUCTION

An academic reaction to the arguments just presented came fast and strong. A torrent of multifarious mentions, comments, sections, papers and books dedicated to taking a favourable or unfavourable stance toward Grünbaum's critique of psychoanalysis began to appear soon after the late-1970s papers where he first presented some parts of this critique (e.g., Flax, 1981); this torrent kept flowing for the more than four decades that elapsed since then. Given this and given the title of this chapter, I should state without further ado that it is impossible to give an exhaustive account of such responses within the boundaries of a monograph's chapter. What I intend in this chapter, thus, must be something humbler: to identify some[9] remarkable attitudes and strategies in Grünbaum's interlocutors and to get inspiration and background for my own responses to the philosopher.

I shall distribute the responses to Grünbaum into five groups. The response shall be considered *favourable* to the philosopher if it agrees with his conclusion that one cannot test causal hypotheses within the clinical method of psychoanalysis, and *unfavourable* if it disagrees in any way with this conclusion. Among the somehow favourable, there are two kinds of attitudes: one that accentuates his critique and brings more elements on its behalf and one that sees its conclusion as an invitation to test the psychoanalytic theory with experimental methods. Among the unfavourable, there are also two kinds of attitudes: one that agrees with the philosopher's premises but does not agree with his conclusion, and another that disagrees with both his premises and conclusion. The premises of Grünbaum I shall have in mind to label the opposing authors into the two subgroups comprise, basically, his understanding of how a reasonable induction looks like, that is, his model of scientific testing, as well as his exegeses of Freud and his depictions of the practice of clinical psychoanalysis. Finally, it is reasonable to discern a fifth group, where the responses accord and discord with the philosopher at the same time: accord with his conclusion but not for the reasons he furnishes.

[9]Unfortunately, I was unable to devise reasonable and clear criteria to select the responses to Grünbaum to be cited along this chapter. The reader may note, though, that the famous 1986 debate in the journal *Behavioural and Brain Sciences* (Grünbaum, 1986a), cited in the Introduction, was given great attention here, for both its representativeness and historical importance. I shall refer to it along the thesis as "the BBS debate".

It is important to observe that the same author may be present in more than one group, as the elements to be distributed are not authors, but arguments; anyway, no scholar should be compelled to hold equal attitudes and strategies in a debate as complex as this one is.

3.2 THE RESPONSES

3.2.1 For

3.2.1.1 "Yes, and Here is More": Stressing and Expanding the Argument

In the BBS debate, Roger Greenberg (1986) reports the findings of his and Seymour Fisher's book of 1977, "The scientific credibility of Freud's theories and therapy" (Fisher & Greenberg, 1977; 1985) to endorse Grünbaum's charges on the use of case-study material for the testing of hypotheses. In his *Foundations*, Grünbaum cites the "friendly critics of psychoanalysis" six times (Grünbaum, 1984). Those findings indicated, before *Foundations*, that the analyst's theoretical commitments are likely to bias and distort the patient's statements and the inferences drawn from such statements. The psychiatrists' findings also indicated, in line with Grünbaum's still-to-come dictum, that psychoanalysis can be tested in "objective" studies – but, contrary to the philosopher, that hundreds of such studies are surely "attesting to the merits of many Freudian postulates" (p. 241).

In the same debate, Edward Erwin (1986) shows discomfort with a friendly-critic compromise such as Fisher and Greenberg's. Three of his points in the debate are important here. (1) As both for the father of psychoanalysis and for most contemporary analysts clinical evidence is the major evidence for psychoanalytic theory, a denial of the uncogent clinic and an exclusive acceptance of experimental evidence falls short in demonstrating that Freudians are rational. (2) If the friendly critics assent to this fact and have only the modest purpose of presenting evidence, Erwin can reply to them that the experimental studies cited in support of the Freudian theory in Kline (1981), Silverman (1976), and also in Fisher and Greenberg (1977) actually "fail to provide strong support for any part of Freud's theory" (p. 235). (3) Faced with the reply that future experiments *might* vindicate the theory and the therapy, Erwin can rebut that, anyhow, we have no rational motives *now* (by that time) for thinking of them as true or effective.

Erwin (1986) thinks that the attitude before the arguments in *Foundations*, “a question that Grünbaum does not address in his book” (p. 236), should not be of a *suspension of judgment* on the epistemic worth of the theory and the therapy of psychoanalysis, but of *pure disbelief* in this worth. Although admitting with Grünbaum that incredulity must not automatically follow from lack of evidence, he tends toward a “tentative” belief that the bulk of the theory is false. Freudian theory, he says, has a hard time to explain some clinical and experimental findings, such as successful symptomatic therapies without symptom substitution, and overall, the success of behavioural and cognitive techniques. Moreover, he continues, many psychoanalytic hypotheses stand far from common-sense, being “neither trivial nor self-evident” (p. 236). He poses a rhetorical question: as for these hypotheses, should not the scarcity of evidence push us to a less patient verdict than Grünbaum’s, to a conclusion that it is rational to regard them as false?

There are replies that stress and expand Grünbaum’s arguments on the *inferential weakness* of clinical psychoanalysis.

Among the findings Greenberg (1986) cites from his and Fisher’s book to ratify Grünbaum’s conclusion are the empirical studies of illusory correlations – for example, the studies by Chapman and Chapman (1967, 1969) – demonstrating that clinicians tend to see significant relationships in the data even when there is none. This would be due to theoretical prejudices filtering and deforming the material.

Along the same lines, Barbara Von Eckardt (1986) points to a “theoretical bias” in Freud’s method of interpretation – a discussion that, she says, is under-elaborated in Grünbaum (1984)[10]. In Freud’s case studies, Von Eckardt (1986) argues, the fundamental principle to formulate interpretations from the exchanges with the patient is the coherence of these interpretations with a pre-existing theory[11]. This kind of reasoning would be theoretically oriented in an unscientific way because it would be circular, a *petitio principii*: there is not, in this conduct, a theory-independent way of checking whether the theory is true or not, for the data are “tinted” by the theory *in advance*. She concedes that the role played by theory in the *generation* of interpretations would not undermine the objectivity of psychoanalysis *if only* there were a procedure of consulting a theory-independent data of some sort to discard incorrect interpretations; this would be so even if the whole range of candidate interpretations lied within

[10]Grünbaum elaborates on a similar problem in the context of his charge of data contamination. Grünbaum’s “selection bias” and Von Eckardt’s “theoretical bias” both seem to be kinds of “confirmation bias”, but he discusses “selection bias” mostly as a medium of influence (see chapter 1).
[11]Von Eckardt (1986) invites us to check examples of this in Cioffi (1970), Timpanaro (1975) and Wolpe and Rachman (1963).

the scope of the same fostered theory. Therapeutic success, she writes, could have this role of a theory-independent check, but, with Grünbaum's debunking of the Tally Argument, the latter can undermine the charge of theoretical bias as poorly as it can eradicate concern over nongenuine productions. She concludes that "the problem of theoretical bias […] remains" (p. 263), thus being impartial as to a future solution to it.

Von Eckardt (1986) also calls attention to the problem of generalization as another that Grünbaum could have elaborated further in his 1984 publication. What she has in mind here is Freud's inferential leaps from a small sample to the entire human population, where the sample does not seem to have the properties to permit such a generalization. Even if we grant that inferences about a single case are possible, we should admit that the psychoanalytic enthusiasm in stating truths on human nature (or, at any rate, on psychopathological, sexual, social, etc., categories) from single cases is not self-justified. This mind-the-gap interpellation is indeed used as a premise for the recurring accusation that the Freudian theory is culturally biased (Von Eckardt, 1986).

Other replies stress and expand Grünbaum's arguments on *how the analysts' expectations may corrupt the patients' productions*.

Greenberg (1986) cites in this respect the studies of Robert Rosenthal to back once more, in the BBS debate, Grünbaum's purging of the clinical method from inside the territory of cogency. Rosenthal (Rosenthal & Rubin, 1978), a behaviourist, presented evidence for what was named "interpersonal expectancy effect" in the experimental context: the phenomenon in which the expectations of the experimenters cause them to interact with the research subjects in a way that increases the probability that these subjects will respond as anticipated (amazingly, even when the subjects are rats). It is a "self-fulfilling prophecy", and the "prophet" may not be an experimenter: research on these effects was extended to teachers and pupils, employers and workers, as well as to therapists and patients (Rosenthal, 1978). Expectancy effects are, thus, very likely in contexts where "observers enter into transaction with the object of their observations", contexts such as psychoanalytic therapy (Greenberg, 1986, p. 240).

A central ally in Grünbaum's arguments on suggestion, notably on the subtle, reinforcing suggestion that may occur in the clinical exchange, the analyst Judd Marmor (1986) replays in the BBS debate what had been put forward in *Foundations* (see p. 22 of this thesis). There he cites his own words from the 1974 paper "Psychoanalytic therapy as an educational process":

> […] patients of each school (of psychoanalysis) seem to bring out precisely the kind of phenomenological data which confirm the theories and interpretations of their analysts. […]

> What the analyst shows interest in, the kinds of questions he asks, the kinds of data he chooses to react to or to ignore, and the interpretations he makes, all exert a subtle but significant suggestive impact upon the patient to bring forth certain kinds of data in preference to others (Marmor, 1974b).

Furnished with both clinical and experimental evidence for this liability, Marmor (1986) claims that it openly shakes the "sacred cornerstone" of free association. He further reminds us that it is the "phenomenon" of transference itself that "renders the internal validation of the psychoanalytic method highly questionable" (p. 249).

Despite the convergences, Grünbaum's stance is a little different from Marmor's. Grünbaum defends the *risk* and the *high likeliness* of clinical suggestion; Marmor defends it as *flagrant* and *pervasive*. Further, the hypothesis of transference is not for Grünbaum proven while Marmor (1986) mentions transference as a "phenomenon" whose "existence" was "described" (p. 249). We can understand these divergences by keeping in mind that Marmor, although a harsh critic, is an analyst trying to correct and improve his practice, while Grünbaum is a philosopher of science interested in any science's burden of proof.

There are also replies that stress and expand Grünbaum's debunking of the argument that, if patients got cured, then we can be sure that true unconscious ideas and links were in them uncovered by their analysts – Freud's *Tally Argument*.

Hans Eysenck (1986), for example, thanks Grünbaum for having demonstrated, with his analysis of the Tally Argument, how the shortcomings of Freudian therapy must disturb the truthfulness of the Freudian theory. The behaviourist who wrote "Decline and Fall of the Freudian Empire" tells he has always considered self-evident that the failure of Freudian therapy to be more curative than natural remission or placebo (Rachman & Wilson, 1980), as well as the success along these lines of other psychotherapies, such as behaviour therapy, is "the clearest proof we have of the inadequacy of Freudian theory" (p. 236); nevertheless, he comments, a philosophical study of this linkage, such as the one found in *Foundations*, was very welcome.

But let us be fair: here Eysenck misunderstands Grünbaum's point. Grünbaum cites research showing that *all kinds* of psychotherapy that are more curative than natural remission have approximately the same curative power; this would make the Tally Argument unsound, the epistemology of clinical psychoanalysis having nothing else to discard the hypothesis of suggestion. But this is not the same as arguing that the failure of a therapy indicates the failure of the theory behind it; it is rather the same as defending the failure of an argument that links good therapy with true theory, which, *considering psychoanalysis' surmised epistemological monotony*, proves its uncogency even if its therapy works as fine as other psychotherapies. By proving that rival theories engender the same therapeutic results, those researches would indicate that a successful psychotherapy can come from an untruthful theory. To wit, that a theory may be false and yet may be able to manipulate, blindly but competently, some psychotherapeutic factors.

One may grant that a successful therapy is not inevitably evidencing a successful theory, but still insist that *a failing therapy is always evidencing a failing theory*, as well as claim that this is the key to understand Eysenck's point above. Indeed, Eysenck (1986) seems to be just arguing that, because Freud's theory includes the Tally Argument and a whole theory of how psychoanalytic therapy works (a theory of the therapy), harming this therapy would harm a significant part of his theory. And he concludes: "It is difficult to see how psychoanalysts can continue this argument, other than by having recourse to the hermeneutic principles which Freud himself would have rejected, and which, as Grünbaum points out, remove psychoanalysis completely from science" (p. 236).

3.2.1.2 "Yes, Let us Look Beyond the Clinical Context": Promoting an Experimental Psychoanalysis

Grünbaum's excursion into clinical psychoanalysis' epistemology has had among its endorsers an interesting kind: not the specialist in defaming the historical figure of Freud in the Cold Freud Wars, not the behaviourist, cognitivist or neuropsychologist concerned with how much the popular and scholastic wisdom may be mistaken over the scientific status of psychoanalysis, not the philosopher of science that had no reason to disagree with the Professor's careful and versed arguments – but the academic psychologist or psychiatrist that has seen an invitation in Grünbaum's conclusion that the experiment is the only cogent arena to test psychoanalytic hypotheses. "Yes, let us go on with this project of taking psychoanalysis away from the couch and into the lab – it is for the best", they replied from the 1980s massive hustle around psychoanalysis on.

In the BBS debate, we read that Gerald Klerman (1986) was convinced by Grünbaum that a scientific psychoanalysis is yet to be established by investigations independent of the clinical context; he suggests important problems to be faced by this expected experimental psychoanalysis: problems regarding *sampling*, the nature of pertinent *control groups*, the development of reliable and valid *measures of intrapsychic functioning*, such as defences, conflicts and self-representations, and the use of *quantitative measures*.

In the book where Grünbaum last published his arguments on psychoanalysis, a compilation of papers dwelling upon such arguments called "Philosophy, science and psychoanalysis: a critical meeting", Linda Brakel (2015) reports the experimental study that Howard Shevrin and colleagues, Brakel among them, published in 2013 and that famously provided neuroscientific evidence for the causal role of unconscious conflicts in producing symptoms – a fact later recognized by Grünbaum himself. Shevrin and colleagues had already undertaken an experimental study that was simpler than this 2013 celebrated one (Shevrin et al, 1992); aware of the study, Grünbaum manifested the conviction that it had presented qualified evidence for the existence of unconscious conflicts, but not for the role of such conflicts in producing psychiatric symptoms (Brakel, 2015). Then, in order to provide such evidence, the research team devised a new study, inspired by the 1992 one but "with a few important differences" (Brakel, 2015, p. 65).

Within a design where every subject undergoes every condition, Shevrin, Brakel, Snodgrass, Kushawa and Kalaida (2013) presented, to each one of eleven phobic subjects, visual primes consisting in their "unconscious conflict words"[12]; for the experimental condition, such primes were subliminal, while for the control condition they were supraliminal. Each prime presentation was shortly followed by a target presentation: the experimental target was composed of "supraliminal conscious symptom words"[13] while the control target was composed of "Osgood unpleasant words"[14]. Their ERP[15] brainwaves were measured under each of the four conditions assembled, specifically their brainwaves' alpha frequency – a frequency associated with cognitive inhibition. The experimenters' hypothesis was that "a causal relationship could be demonstrated if the unconscious conflict words produced a differential inhibitory effect on the conscious symptom words" (Brakel, 2015, p. 66). It was confirmed: alpha generated by the unconscious conflict words showed to be highly correlated with alpha associated with the conscious symptom. Moreover,

> this inhibition was observed *only* when the unconscious conflict words were presented as subliminal primes, not when these same words were presented as supraliminal primes. A further control solidified the findings: no results were obtained with either subliminal or supraliminal unconscious conflict word primes when the targets were the Osgood unpleasant words (Brakel, 2015, p. 66).

[12]Such words were derived from the following procedure: psychoanalysts interviewed the subjects and picked out the central unconscious conflict of each of them; next, for each the analysts selected a number of words they could agree upon as representing each unique core unconscious conflict (Brakel, 2015).
[13]In the procedure described in the footnote above, the analysts also chose an equal number of words to reflect each subject's conscious symptom (Brakel, 2015).
[14]Words with non-individualised negative meaning (Osgood, May & Miron, 1975; Brakel, 2015).
[15]The "event-related potential", a technique where neural responses to specific events are extracted from the EEG (electroencephalogram) (Luck, 2014).

This story ended, according to Brakel (2015), with Grünbaum laurelling their effort to solve the fuss he had started decades ago:

> Concerning this experiment and its attempts to gain empirical evidence for causally active unconscious conflicts, Shevrin reports (*Science Daily*, June 16, 2012, and personal communication) that Grünbaum sent him an email stating, "I am satisfied" (p. 66).

Almost ended, actually, because, in the *last* paper on psychoanalysis, published in that same book, he showed "he cannot be really satisfied regarding the causal powers of unconscious conflicts" (Brakel, 2015, p. 66).

Back to the BBS debate, Joseph Masling, Frederick Suppe, Morton Reiser, Horst Kächele, and less directly Greenberg and Marmor, were already promoting an experimental or epidemiological psychoanalysis as a response to Grünbaum's *Foundations*. Masling (1986) affirms his "only disappointment" with the book was how much it had disregarded the vast experimental evidence which until that time was showing itself relevant to the worth of the theory:

> In passing, Grünbaum notes that some experimental work has been done, but his primary purpose evidently was not to dwell at length on the several reviews of empirical data on psychoanalytic theory (Fisher & Greenberg 1977; 1978; Kline 1972; Masling 1983; Masling & Schwartz 1979; Silverman 1976). (Masling, 1986, p. 250)

The most important of these studies, he says, are the ones that, despite coming from the psychoanalytic theory, aims at clarifying topics the theory ill-recognized or was incorrect about; the ones testing how truthful the theory is, in their turn, would be the least interesting of them.

This last point fails to inform us that some psychoanalytic hypotheses are so basic that they are not recognised as psychoanalytic – they go unnoticed by the critics even when validated by experiments. The causal role of conflict in the genesis of psychopathology is one of these hypotheses. It was confirmed "in a wide variety of approach-avoidance studies in lower animals", writes Marmor (1986, p. 249), who also comments that Grünbaum skips a debate over such postulate. Anyway, already in 1986 there were "literally hundreds and hundreds of objective studies attesting to the merits of many Freudian postulates" (Greenberg, 1986).

By the way, by 1986 these extra-clinical studies had already had a great ethical and political role. The entrenchment of analysts behind presumed clinical prerogatives, and among these analysts an orthodox understanding of the causes of sexual orientation, despicably hampered the political deed of removing homosexuality from the DSM's[16] list of pathological

[16] The Diagnostic and Statistical Manual of Mental Disorders, authored by the American Psychiatric Organization (APA).

categories in the 1970s; this story is told by Suppe (1986). That it is told as a compliment to Grünbaum's *Foundations* is altogether symbolic. The story lies somewhat aside from the point of those extra-clinical studies' epistemic worth, but not so much. For it shows how nasty injustices and neglects can stem from an uncritical and absolute epistemic "exceptionality" in psychoanalysis – a euphemism for the irrational entrenchment the analytic community can sometimes enact[17].

In the BBS debate, two scholars tried to look into the future of experimental or epidemiological psychoanalysis, telling us what was being done of it at the moment and what we could still expect of it in the future.

The analyst Reiser (1986) was one of them. While he could not be fully sure of the probative futility of clinical data, he saw that "ultimate validation will have to come from extraclinical [sic] findings" (p. 255); nevertheless, he presented a model in which the clinical process could find its place in this search for truth. This place is basically the same Grünbaum (1984) had already reserved for such process – a heuristic place, namely, the mere but relevance-ensuring place of a hypothesis generator. The work of "recording, describing, arranging and studying the 'crude ore' data" of analytic processes could "refine psychologically framed questions and hypotheses, for example, about memory, anxiety and perception" in order to make them amenable to "experimental testing by methods of natural science" (p. 255) – by which he means the methods of both neuroscience and cognitive science, the two poles in the midway of which psychoanalysis would reside.

Along this midway theory we would have though to be aware of the conceptual immiscibility between the psychological and the biological realms, studying mental and

[17]The American Psychiatric Association (APA) ceased to consider homosexuality a disorder in 1973, removing it from the second edition of its diagnostic manual. From inside the APA, however, a group of neo-Freudian psychoanalysts headed by Irving Bieber and Charles Socarides disputed this and asked for a referendum vote of the membership, the majority of which voted once more for the depathologizing resolution. Suppe (1986) tells us that, in spite of this, by the time he writes, almost every analyst regards homosexuality as a mental disorder "rooted in faulty identifications, infantile fixations, or defenses [sic] against unconscious impulses such as castration anxiety exacerbated by the Oedipal conflict" (p. 261) and wants to have it treated on the couch.
What motivated the majority of the APA was exactly the extra-clinical evidence largely disconfirming the orthodox psychoanalytic position – position which, Suppe (1986) wants to make clear by citing Freud's famous response to a mother bedevilled by her son's homosexuality, was not shared by the father of psychoanalysis. But "virtually all psychoanalysts" (p. 262) despised this evidence: "Only in the consultation room does the homosexual reveal himself and his world. No other data, statistics, or statements can be accepted as setting forth the true nature of homosexuality", says Socarides (1970), as cited by Suppe (1986, p. 262). With Grünbaum, then, there could no longer be such hateful excuses for resisting extra-clinical testing (Suppe, 1986).

physical objects independently and comparatively when they come from the same phenomena; this is the approach upheld by the author and named a "dual track approach" (Reiser, 1986, p. 256), which would serve to "avoid the pitfalls of inappropriately psychologizing physiology or physiologizing psychology while at the same time allowing for generating intermediate concepts that could serve as 'transducers' between the two domains" (p. 256).

By the time, Reiser (1986) disclosed that a program of this sort was far from becoming mature, but also that some impulses to its reasonable and clear direction had already appeared. In the years that followed the BBS debate, a program named "neuropsychoanalysis" was strikingly developed (Fotopoulou, 2012). His statement that Grünbaum featured as a "sheep in wolf's clothing" (p. 256), given that *Foundations* could stimulate analysts to "respond constructively" (p. 256), was altogether correct.

Kächele (1986) also took a glimpse at the experimental future of psychoanalysis. He consents entirely with Grünbaum's conclusion that causal hypothesis on the origins of neuroses cannot be validated by the clinical enterprise, but only by way of largescale epidemiological studies, as well as developmental and twin studies – his examples are the works of Schepanck (1974, 1984) and Emde (1980). Kächele (1986) remarks that, given what can be done in this respect, "it may turn out that psychoanalysis as founded by Freud will undergo major changes or even dissolve into a new frame of reference" (p. 244).

3.2.2 Against

3.2.2.1 Non Sequiturs and Missing Premises

Accepting the Premises but Rejecting the Conclusion. And there are responses expressing discomfort with Grünbaum's arguments. Some, to be discussed in the next section, are rejecting his arguments as a whole. Others, appreciating his premises, reject only his transition from the latter to the conclusion that the clinical method of psychoanalysis cannot be probative. This position is permissible if his set of premises is deemed incomplete or his transition, a too-long leap, namely, a non sequitur. Perhaps the most representative instance of this kind of response to Grünbaum is found in Edelson (1986):

> whereas I do agree with Grünbaum that the canons of eliminative inductivism are the canons of scientific reasoning and method, that the case study investigations of psychoanalysis have involved serious violations of these canons, and that it is the responsibility of contemporary psychoanalysis to bring its investigations into conformity with these canons, I cannot agree with his conclusion that any reliance by psychoanalysis on data obtained from the psychoanalytic

> situation to support its causal inferences according to these canons is by the very nature of things doomed to failure and that psychoanalysis must turn instead entirely to epidemiologic and experimental research for such support.
> […]
> In short, I am not inclined to quarrel with Griinbaum's diagnosis, whatever quibble we may have over details, but I am prepared to question his remedy (pp. 232-233).

Brian Farrell's comment in the BBS debate (Farrell, 1986) is another good example of this position. He states simply that Grünbaum's criticism is *misleading* inasmuch as it attempts to convince the reader that the uncertainties and the unscientific features of the psychoanalytic method entail its worthlessness. It would be possible to assume that the clinical material tells us the truth, or part of the truth, about patients' disorders and the human nature – that it has "probative value, small or great" (p. 237) – even standing before "Grünbaum's anti-Freud argument" (p. 237). Farrell (1986) seems to argue that nothing prevents the *concrete* version of the psychoanalytic method to outpace the arguments about its validity that a psychoanalytic author *is able to give*.

When proposing a "minimal set of standards to be met by clinical case studies in order to be accepted in […] psychoanalytic literature" (p. 233), Marshall Edelson (1986) does not even try to relativize suggestion's contaminating power in the clinical situation. He takes it for granted, but thinks it might cease to be a problem if only analysts exposed which factors of a clinical situation could have influenced the production of the data favouring a hypothesis if the latter turned out to be false and presented "some argument […] for dismissing at least one such factor as a plausible alternative explanation" for the obtention of the favouring data (p. 234).

Fine and Forbes (1986), on the other hand, would regard Edelson's "at least one" demand as too bureaucratic. According to them, analysts should set aside the possibility of suggestion in cases it cannot be contextually "articulated", that is, "made specific, concrete and testable[18]" (p. 238). Comparing it to Descartes' demon, the philosophers interpret Grünbaum's charge of suggestibility as consisting in "a general, always-available doubt, independent of any specific mode of instantiation or mechanism of operation", in other words, as "just a place-holder for skeptical doubt, or fallibility" (p. 238). In sum, the philosophers think that suggestibility is an epistemological threat, all right, but also that it is senseless for a clinical analyst to worry about it in its intangible and imponderable versions.

On the same line of argument, Robert Holt (1986) affirms that the matter of how far the responses of any particular patient are contaminated by the analyst's gross or delicate suggestions "is an empirical matter" (p. 243), one that begs to be properly investigated. The

[18]Also, according to them, analysts should regularly check whether this picture has changed.

habit of tape-recording and transcribing sessions could, according to him, permit this controlled investigation of “the extent to which therapists […] ‘lead’ their patients or […] shape their productions by a kind of operant conditioning” (p. 243). “It is quite possible for a patient”, Holt (1986) remarks, “to be much affected by suggestion in one respect and for the analytic data to be adequately ‘clean’ in many other respects” (p. 243). Like Farrell (1986), the psychologist thinks Grünbaum throws out the baby with the bathwater: it would not follow from the problems of the traditional clinical method that the data from the psychoanalytic hour are useless.

This clamour for an empirical check on the charge of contaminating suggestibility is a popular line of reply in the BBS debate, as we can see in an eloquent passage by Donald Spence:

> Just because we are not always aware of the original texture of the clinical dialogue and just because it is perfectly possible for certain kinds of suggestion to appear does not mean that they present us with an impossible methodological obstacle, or that the effects of suggestion cannot be isolated, when and where they occur, and their influence kept separate from other aspects of the psychoanalytic process. As more verbatim transcripts are made available, it should become increasingly possible to identify instances of suggestion and to study, in turn, their effect on other parts of the treatment (Spence, 1986, p. 259).

Verbatim transcripts could, thus, inform analysts of the contaminated or uncontaminated nature of the clinical exchange and, in case some confounding influence is demonstrably active, such transcripts would also give analysts resources to neutralize (“isolate”) it; in other words, even if suggestion were active, a detailed inspection of previous sessions could make it controllable in following ones. At any rate, evidence for the occurrence of suggestion would not be simple, exhibiting three levels: evidence for the leading behaviour of the analyst; evidence for the alteration of the patient’s responses as a reaction to the analyst’s leading behaviour; finally, evidence that such are responses accepting the analyst’s “command” at face value (Spence, 1986). Suggestion is only an evil demon if inaudible and invisible, he seems to be claiming.

Spence (1986) shows that, although Grünbaum’s suggestion-argument is benefited by an omission of primary evidence, namely, of clinical transcripts, it is also weakened by this omission. The crux of the argument is the double-negative that analysts are *not* able to prove, from their non-controlled and private clinical encounters (and, of course, from the debunking of the Tally Argument), that *no* suggestion transpired in them. However, the very nature of the argument leads to its destruction, as the lack of primary evidence makes it equally impossible for Grünbaum to prove that suggestion is a common phenomenon in the clinic: “Until we have more adequate methods for ‘unpacking’ the clinical interchange and can find a new genre to take the place of the traditional case report, Grunbaum’s charge may remain moot – but it is certainly not proven” (Spence, 1986, p. 259). Lack of primary evidence also makes Grünbaum ineligible to affirm that the training given to analysts does not prepare them to be less directive, "schooled" as the latter would be "in a tradition which places an emphasis on minimal comment and redundant examples" (p. 259).

Around 1985, researchers interested in clinical psychoanalysis were devoting more and more attention to the empirical side of its possible biases. Apart from Edelson's abovementioned "minimal set of standards", a prominent expression of such a trend was the research undertaken by Lester Luborsky. By the way, Holt (1986), in the same piece presented above, indicated the work of "Luborsky"[19], as well as the work of "Dahl, Gill, Weiss & Sampson and Silberschatz"[20], as the ones demonstrating that tape-recording and transcribing sessions could allow for "carefully controlled, hypothesis-testing research" (p. 243). Luborsky contributed to the advancement of methods to reduce bias along the identification and description of "symptom-contexts" and "core conflictual relationship themes" appearing in psychoanalytic sessions, methods we shall succinctly portray in a moment. Responding specifically to the problem of contamination, Luborsky (1986) reports that his team would apply a "Relationship Apperception Paradigms Test (RAP)" in the patients *before* their treatments begin with the aim of registering the "core conflictual relationship themes" that are independent of the psychoanalytic encounters, thus allowing the discrimination between intrinsic characteristics of the patients and artifacts of the latter's compliance to the therapist' expectations. If "the core conflictual relationship themes are evident apart from the treatment as well as within the treatment", then researchers are entitled to conclude that "the pattern is characteristic of the patient rather than generated by the expectations of the therapist" (p. 248).

This a powerful response to the problem of contamination. Luborsky's following, "more general", response to the problem, however, is not. He claims that the suggestion hypothesis is not consistent with the fact that "predictions based on sessions have dozens of highly significant and substantial correlations with measures outside the treatment" (p. 248). I would observe, however, that it is perfectly *possible* for the analyst's influence to extend beyond the session and into the patient's life as a whole.

Be that as it may, as Allan Gauld and John Shotter (Gauld & Shotter, 1986) note, it is not *in spite*, but *in virtue*, of a probable action of suggestion within psychoanalysis' borders that we should keep an attentive eye on this domain: we "should stand back in amazement at the peculiar properties of interpersonal relations" evidenced by the phenomena of suggestion (p. 240). They assert that "a special study is required of such phenomena" (p. 240).

[19]Holt (1986) does not cite the author appropriately.
[20]Holt (1986) does not cite the authors appropriately.

In a more recent paper, Lacewing (2013b) does not try to deny the risk of pervasive suggestion in clinical psychoanalysis:

> given the nature of the clinical setting, the therapeutic relationship, the unconscious and unintended communications of the analyst, the importance of the analyst's interpretations for generating data (psychoanalysis would not get far if analysts did not propose understandings of their patients and observe the reaction), there is no obvious way to amend clinical methodology that would prevent the possibility of suggestion (p. 725).

But, unlike Grünbaum, he shows it does not follow from this risk that suggestion could never be traceable in the clinical context and that clinical research is therefore doomed. Suggestions would be comparable to the expectancy effects occurring in experiments, he argues, and experimenters already know how to protect their results from such effects: they check how much of their results are replicated by many independent experimenters, presuming that the higher their replication the lower the probability of them having been promoted rather than just observed. He proposes that the global community of analysts transpose this solution to their provinces by taking into account the results that are corroborated across a significant number of different and unrelated analysts, patients and institutions: where the results are corroborated, they could be confident that no bias, including suggestion, have occurred. He notes that, as the "psychodynamic model of the mind" is largely and independently corroborated but "metapsychological and etiological theories" are not, only that model should be deemed as based on uncorrupted evidence.

Lastly, let us present some arguments that, although admitting clinical psychoanalysis' inferential problems to be well represented under Grünbaum's pen, do not accept his conclusion that induction in that context can be nothing but weak.

We have mentioned that, in the BBS debate, Marshall Edelson proposed a list of minimal conducts the psychoanalytic community should expect to be expressed in its publications of clinical case studies[21], and we have seen which point of the list attempts to give a prudent channelling of the problem of contamination. It is worthwhile to go through the rest of the list, not only because it is presented in the famous debate, but also because it is a summary of Edelson's book-response to Grünbaum's *Foundations*, called "Hypothesis and Evidence in Psychoanalysis"; the book-response was published in the same 1984 in which Grünbaum's critique was still piping hot. Edelson (1986) expects authors of clinical case studies: to state in the clearest possible manner their hypotheses, conclusions or generalizations about the cases or treatments at issue; to separate facts and observations from interpretations, i.e., to distinguish

[21]He observes then that the list should be relevant specifically for that part of the community aiming to construct a scientific body of knowledge (Edelson, 1986).

"what can be observed without knowledge or use of the theory being tested from interpretations based upon the very theory such observations are being used to test" (p. 233), showing later how such observations can, not merely juxtapose with, but *explain*, the hypotheses about the case; to specify which potential observations could refute or weaken their hypotheses, to report at least some inconsistent or unexplainable observation and, in case their hypotheses are not discarded, to indicate grounds for keeping such hypotheses, with or without the original scope; to give some reason why their observations are evidence for their hypotheses rather than for at least one rival hypothesis; finally, to propose how extensively their hypotheses could be generalized to similar cases or treatments, and to justify this proposition. Edelson (1984, 1986) presumed such measures could save psychoanalysis from the bulk of most vociferous epistemological criticisms, especially the one by Grünbaum.

In a vaguer manner, Edelson (1986) also proposes some promising paths to be followed by wise and vigilant analyst-researchers to deal with the problems indicated by Grünbaum, among which are: to explore the methodological insights of single subject research designs; to cultivate the directions set by Luborsky's symptom-context method (see below) and Glymour's bootstrap strategy (Glymour, 1980, 1982); and to explore "methods recently developed in the social sciences for making valid causal inferences when data are nonexperimental or qualitative" (p. 233).

Moreover, Edelson (1986) rescues for the forgetful anti-Freudian the fact that the method of free association was actually sustained by a "complex set of premises about the mental apparatus" (p. 233), not by mere professions of faith. Grünbaum would have lost the opportunity of putting Freud in the dock for the right reason – the lack of cogency of such presuppositions. Analysts, in their turn, should be more curious about them, for

> theoretical presuppositions always underlie a procedure for obtaining data in science. What theoretical presuppositions about the mental apparatus do in fact underlie free association (regarded now solely as an instrument for obtaining data relevant to the assessment of the credibility of psychoanalytic hypotheses)? What empirical support exists independently of the use of the procedure itself for these presuppositions? We psychoanalysts have not even precisely specified what variables (e.g. suspension of conscious purposiveness) constitute the procedure itself; how to estimate the values these variables take (e.g. the degree to which such suspension is achieved); and how to determine just what properties of communications change, and in what way, as these variables take different values. The answers to such questions must be sought by psychoanalysis in the years ahead if it is to refute convincingly Grünbaum's claim that there is no way to differentiate among the mass of clinical material those communications having greater degrees of evidential value from those having lesser degrees (p. 233).

With his bias-averting methods, Luborsky contributed to clinically answer "to what extent […] the therapist's inferences [are] accurate about the parallel between the patient's current relationship patterns and past parental relationships" (Luborsky, 1986, p. 247) and whether "psychotherapy sessions [could] provide data that permit verification of the causes of symptoms" (p. 248). He developed one method for independent judges to measure, from sessions' transcripts, the patients' "core conflictual relationship themes", that is, their "transference patterns", and another – the "symptom-context method" – where again professionals and scholars "from varied theoretical backgrounds" (p. 248) would independently and blindly judge the contexts in which symptoms had been enacted, experienced or reported during sessions as well as control-contexts in which no symptom had emerged. Through this last method, one would be able to discern "the 'causes' of neurotic and psychosomatic symptoms" – that is, the factors "shown to have a high probability of being temporally associated with the development of a symptom when that high degree of association could be, and probably is, causal" (p. 248).

Luborsky (1986) tells us that, in a 1983 discussion in the Rapaport-Klein Study Group, Grünbaum recognized the relevance of the symptom-context method and of its results, but observed that the method would not be able to cover "Freud's grandest clinical hypotheses". Luborsky (1986) sees in this a sign that the method *could one day* become able to cover such hypotheses: "if a method is able to test a minor hypothesis it is only a matter of the investigator's ingenuity to provide it with a major hypothesis to test" (p. 248).

Like Luborsky's, Wilma Bucci's (1989) response to Grünbaum is a theoretical and methodological proposition to make inferences in clinical psychoanalysis more reasonable and reliable. In the paper "A reconstruction of Freud's tally argument: A program for psychoanalytic research", she notes that the Tally Argument was essentially the Viennese's attempt to harmonize two oppositions: the opposition between the mind's private and subjective nature and science's demand for mutually observable events, and the opposition between the analyst's investigative role and the analyst's therapeutic role. She then starts to introduce a "new approach to scientific psychoanalytic research" (p. 250), which would address such oppositions in a systematic way through a renewed, empirically-informed version of the Tally Argument.

Bucci's (1989) opening step in the paper is to reformulate psychoanalysis in terms of the "dual-code model of mental representation". According to this model, the mind registers things in two different systems or codes, the verbal and the nonverbal. The verbal is the one of language and logic: it comprehends lexical items and its processing is single-channelled, sequential and organized by logical rules. The nonverbal code, in its turn, "includes experience in any sense modality, representations of bodily experience, and representations of motoric activity" (p. 257) and follows distinct rules: the representations of this system are connected or categorized because they have similar structures or co-occurred in a certain time or place. As the model tells there would be bidirectional "referential links" between the systems, it permits a restatement of the goal of psychoanalytic therapy: to transpose the perceptual-motoric-emotional schemas of the nonverbal system into words (to "reach" them with the verbal system), and, backwards, to alter them through words. She concludes: "the special value of the psychoanalytic method, which rests on insight mediated through free association and interpretation, over other forms of treatment, is precisely in the effect of activity of the referential links" (p. 260).

With this in hand, Bucci (1989) puts forward her renewed version of Freud's Argument, the "Tripartite Probabilistic Tally Argument (TPTA)":

> 1. If the patient gets better, then it is probable that changes in the perceptual-motoric-emotional (PME) schemata of the nonverbal system have taken place.
> 2. If changes in the PME schemata of the nonverbal system occur following a particular analytic interpretation, then it is probable that referential connections to that system have been made.
> 3. If referential connections to the nonverbal system are made by an interpretation, then it is probable that the interpretation was accurate, i.e., tallied "with what is real" in the patient (p. 261).

This version would exhibit advantages in relation to Freud's. It would replace Freud's subjective notion of "correct insight" for the two mediating notions of "referential connections" and "perceptual-motoric-emotional structures", which can be "independently defined in terms of external and observable events" (p. 262). It would also be more realistic in that it proposes a probabilistic rather than a necessary relationship between the terms. Bucci (1989) defends that any initiative of investigating the variables of the TPTA by relating them to one another and to observables, be it clinical or experimental, is a direct or indirect test of psychoanalytic hypotheses.

At last, she sets out to show how the two mediating notions can be empirically defined and the TPTA, empirically supported. She cites the experimental research done by Paivio (1971, 1986) and others basing the measures of Referential Activity (RA) on scales rating "the qualities of *Sensory Concreteness*, *Specificity*, *Clarity*, and *Imagery* levels of speech" (p. 264), and also Von Korff's (1987) study bringing evidence that "the most positive treatment course occurred where both analyst and patient were high in RA and least successful where both were low, with dyads mixed on this dimension showing midrange results" (p. 269). Regarding the measure of perceptual-motoric-emotional structures (PME structures) "played out [by the patient] repeatedly in both verbal and nonverbal behavior in diverse specific forms" (p. 273), both in itself and in the framework of the TPTA, she cites hers (Bucci, 1988) and Leeds' (1988) systematic procedures to identify them, which involve condensations of "what the person says and does to its common elements across specific situations, people, and places" (p. 274) by different research workers at different increasingly generalising stages to "reduce as far as possible the effect of clinical intuition on the judgement process" (p. 274); and she cites her empirical proof (Bucci, 1988) that the repetitive PME structures "are more likely to emerge in segments of the free association that are high in RA, i.e., where the verbal material is specific, concrete, and clear" (p. 274).

In sum, Bucci's (1989) response to Grünbaum is that the Tally Argument is a hypothesis that can be clinically tested. However, as the probabilistic strictures are multiplied along the three propositions of the TPTA, the relationship between symptom alleviation and interpretation accuracy is likely to be very weak. She concludes that interpretation accuracy may be only one factor contributing to the therapy's efficacy or may even be irrelevant to it – and that the TPTA is not the most straightforward path to test the psychoanalytic theory.

3.2.2.2 Wrong Assumptions, Nonsense Argument: Rejecting Both the Premises and the Conclusion

We arrive at the already mentioned responses which discredit, not only Grünbaum's conclusion about psychoanalysis' clinical method, but also the premises upon which this conclusion stands. According to them, Grünbaum would rely on wrong ideas about this method. The ideas he would be wrong about, however, vary across the responses: he would be misinterpreting Freud's words; misrecognizing the potential of the method; misunderstanding the aims and tenets of both psychoanalysis and the scientific practice in general; etc.. A fair sample of this populous corner in the warehouse of responses to Grünbaum is hard, but here is an attempt.

To begin with, we have philosopher Peter Caws (1986) claiming he is not as sure as Grünbaum that psychoanalysis is like an edifice with (fragile) foundations. A better metaphor for describing psychoanalysis is, he would rather say, that of a scaffolding, therefore something without a foundation – and Freud, an honest thinker, have not dared to think it is more than this, a temporary and fragile construction which would in the future allow for a solid and lasting one. In Freud's inaugural movements of authorship, he pursued a neurophysiological model of the mind to be the foundation for his new clinical method and therapy, but later he explicitly settled for developing the latter without any foundation, as the evidence he needed was not obtainable back then (Caws, 1986). Although Freud remained defending that psychoanalysis "had scientific interest and should be regarded as a part of the science of psychology", "it has to be admitted that psychoanalysis as a full-fledged natural science has remained promissory" (Caws, 1986, p. 230). So, it was too easy for Grünbaum to demolish the foundations of clinical psychoanalysis, because, in fact, they do not exist – which would explain his "jovially combative" style, signalling victories "with triumphant exclamation points" (pp. 230).

Fine and Forbes (1986) also maintain that Grünbaum's victories are somehow artificial. Specifically, the authors maintain that his premise that Freud ignored some plausible explanations for patients' data only leads to his grand conclusion if coupled with an absurd principle – that, for a hypothesis to be supported, all of its rivals must be proved implausible. This would be a "methodological myth" (p. 238), for ruling out all of the competitors of a hypothesis would not be generally possible or even necessary in science. Moreover, the authors blame Grünbaum of misrecognizing that "there is no good and general theory of evidence (or support)" and that a theory of the sort is only assessable *comparatively*. Thus, Grünbaum could only have examined "the evidential structure of Freud's clinical argument" *in relation to* "similarly complex structures that form the clinical background for theory and therapy in, say, somatic medicine" (p. 238).

Caplan (1986) concurs with Grünbaum that the Tally Argument is the methodological linchpin of Freudian psychoanalysis. He also thinks, though, that to verify elements of the psychoanalytic theory in experiments and prospective investigations is to move away from its original spirit; the hypotheses supported through these highly controlled methods would be radically different from the hypotheses Freud and his early followers have formulated to account for the therapeutic efficacy of psychoanalytic techniques. Analysts are "welcome" to try to find data beyond the clinic but, "if they do, what they ultimately discredit or vindicate will in all likelihood have little to do with Freudian psychoanalysis" (p. 229).

One of the most influential objections to Grünbaum's argument about the Tally Argument was given by the philosopher Frank Cioffi in the BBS debate and in many other publications. According to Cioffi (1986), Grünbaum was absolutely wrong in interpreting that Freud had defended his epistemic practices primarily by appealing to therapeutic effects. To demonstrate this, he machine-guns examples of inferences and comments in Freud where the Tally Argument was either indiscernible or contradicted: Freud spoke of "countless successes" right after he had recognized problems in the seduction theory (hence he was aware that such successes were unrelated to truthful interpretations); he attributed developmental hypotheses to normal character structure, as well as to pathological conditions which he took as immune to therapeutic influence; "there was no therapeutic upshot to vouch for […] [the] veridicality" (p. 231) of many of his interpretations of dreams and errors; he took as evidence "extraclinical explanatory triumphs (Oedipus, Hamlet, totem and taboo) in which suggestion could not operate" (p. 231); his theory of infantile sexuality was based on direct observation, not on therapeutic reactions to interpretations making use of this theory; he more frequently than otherwise "expresses certainty as to the correctness of an interpretation in which there is no question of therapeutic effect, such as Dora's playing with her reticule, her pseudoappendicitis, her limp, and so on" and has not abandoned his certainties about the Wolf Man case after coming to know that the latter was once more deranged; etc.. In sum, Cioffi (1986) claims, Freud did not wait for therapeutic effects to follow the enunciation of aetiological and dynamic constructions in order to show confidence about such constructions. If Freud placed "cardinal reliance" on anything, "it was on what Grünbaum refers to as the 'consilience' between his inferences and the clinical and extraclinical data" (p. 230); he, for example, "believed in the overwhelming influence of the Wolf Man's primal scene not because of its therapeutic effect but because of its narrative power", that is, "its ability to confer intelligibility on the data" (p. 230).

While Cioffi defends that the lenses used by Grünbaum inadequately focus on the therapeutic dimension of psychoanalysis, Robert Woolfolk (1986) opines that his lenses should have captured more of it. Although aware it was not the aim of *Foundations*, Woolfolk (1986) expected to read there "some elucidation of the moral dimensions of psychoanalysis", as this "seems pertinent to a determination of the extent to which a natural science framework fits it" (p. 265). He emphasizes that psychoanalysis does not only describe how the psyche works, but also outlines a moral perspective; that, *qua* theory of personality and psychopathology, it is inescapably prescriptive and normative, reflecting and influencing standards of conduct and revealing itself as

> the contemporary cultural equivalent of descriptions of virtue and peccability. Viewed through this lens, psychoanalysis has much in common with Marxism, in that each not only provides an ambitious philosophical anthropology but also a theory of liberation. It is perhaps this aspect of psychoanalysis that makes it one of our most intriguing and influential cultural products (pp. 265-266).

During the BBS disputation, the same Shevrin who had been (and remains) advancing neuroscientific experiments designed to test the existence and nature of unconscious dynamic processes, experiments with which Grünbaum was allegedly "satisfied" (see above), does not see in his contamination argument the force to make us abandon the idea that the clinical context can be a source of knowledge. He points to the absurd consequences of the argument that contamination in this context is always potential and, with the resources in offer, untraceable, contending even that, although aiming at the psychoanalytic clinical method, "Grünbaum [...]

may have inadvertently hit reason itself" (p. 258). He imagines an experiment in colour perception where subjects are asked to tell the experimenters which colour was the disk they just saw: the experimenters would note a regularity between the light frequencies of the discs and the reports of the subjects' *qualia*. They would note this, however, only if they granted that the subjects' reports are accurate and that "there is a regular and invariant relationship between the meaning of words and their referents" (p. 258). Just like we admit that a subject states the truth when saying "I see blue", we could admit that a patient states the truth when talking about a specific thought at her sister's deathbed and that she has reproached and forgotten it until a moment ago.

> Why should this account not have evidential status as much as the statement, "I see blue," once we take into account that the capacity for reality testing in and out of the transference is not limited to experimental subjects? Moreover, if you deny reality testing to neurotic patients, you cannot maintain it for experimental subjects. If Grunbaum is altogether right about the analytic clinical method, then there is no relying on reasonableness in any context (Shevrin, 1986, p. 258).

Shevrin (1986) observes that, as trusting that patients can accurately describe their own experiences is not the same as trusting they can truthfully explain such experiences, and as Grünbaum fails to draw this distinction, the stance of his argument is suspicious to an unprecedented degree.

Jim Hopkins's also very influential response to the Millian philosopher was that inferences in clinical psychoanalysis cannot be measured by the Millian rule, but rather by the rule of common-sense reasoning (Hopkins, 1988). The procedure by which the mind of an analyst infers the cause of a behaviour is comparable to the procedure by which any human mind infers the cause of a behaviour: by transmitting the behaviour's apparent content to the description of a mental entity (a wish, belief, feeling, etc.) and by recognizing in the latter a cause. In both situations, the features of the cause are guessed from the features of the effect. If a person drinks water, we infer the person had the wish to drink water; if a patient schedules a session when he/she was assigned to watch over his/her kid, an analyst infers that the patient wishes to avert some feeling involved in that task[22]. The difference between common-sense and clinical psychoanalysis would amount only to a difference in theoretical and technical sophistication; the analyst would pay attention to all kinds of behaviours (including dreams, symptoms, etc.), would infer motives from large clusters of behaviours, would seek coherence among all of the patient's probable motives, would presume mechanisms of symbolization and distortion, etc.. Hopkins (1988) argues that, if common-sense inferences are cogent regardless

[22]The first example comes from Hopkins' (1988), the second from myself.

of not using Millian canons – as far as they are inferences to the best explanation – so is psychoanalysis; thus, that Grünbaum is wrong in ruling out thematic affinity as a mark of causal connection.

In a paper exploring the advantages of Hopkins' direction, Michael Lacewing (2012a) reflects on the claim, made by some authors, that common-sense inferences are implicitly Millian (he cites Kelley, 1967, 1973), as the background knowledge with which they become reasonable could be seen as a sort of "nonconscious statistical tracking that matches Mill's canons" (p. 203). Grünbaum's opponents who, like Hopkins, approximate psychoanalysis and common-sense but who, unlike him, agree with that Millian characterization would defend that

> nonconscious tracking of covariance across patients over time creates a significant body of background knowledge as to the causes of neurotic behavior and symptoms. Psychoanalysts can also contrast neurotic and 'normal' behavior, because the causes of the latter are given by analysts' common-sense background knowledge concerning motives (which acts in place of a non-neurotic control group) (Lacewing, 2012a, p. 203).

Common-sense psychological inferences could not be considered *exclusively* Millian, though, for two reasons. The first is that the concept of desire would be empty if it did not authorize Hopkins' simpler method. If "the existence and content of desires […] [were considered] logically independent from behavior *in toto*", then we would allow

> the *possibility* that people in general may have desires that never have any influence on their behavior. (It is important to remember that desires may influence behavior in many more ways than merely leading one to seek their satisfaction, e.g., avowals, daydreams.) But I can think of no (widely accepted) philosophical theory about the nature and content of desire that would not reject this possibility. If we ask, 'What is the desire to drink?', we can provide no answer without referring to the behavior to which it is typically causally relevant. What makes a mental state a *desire*, rather than some other type of state, is precisely the pattern of its relations (causal and/or normative) to behavior and other mental states (Lacewing, 2012a, p. 204).

The second reason is that the particular actions that would call for a *token* explanation could not be illuminated by a Millian tracking, as the latter could only be applied to *types* of causes and effects. Lacewing (2012a) concedes, though, that common-sense inferences *embed something like* the Method of Difference, and this *precisely because they are inferences to the best explanation*; he cites the defence made by Rappaport (1996) and Lipton (2004) that inferences to the best explanation are "sensitive to precisely the sorts of covariations that Mill's methods also seek to identify" (p. 205). Hence, he is not convinced by Hopkin's strong claim that the method of homogenising effect and cause, the "method of interpretation", is completely autonomous.

Convinced that common-sense psychology must be *partly* autonomous, though, and that psychoanalytic inferences are based on it, Lacewing (2012a) provides reasons for why such inferences would not inherit a relevant share of the faults of common-sense. He discusses the existence of the "correspondence bias" in psychoanalysis, a bias which researchers have demonstrated to be heavily committed by common wisdom (e.g., Jones & Harris, 1967). It would consist in "the tendency of observers to infer from a subject's behavior a corresponding disposition, even in situations in which the behavior is highly constrained" (p. 206). The philosopher cites evidence (Fletcher, 1995; Gawronski, 2004) indicating that this bias is reduced or eliminated when the observers in question are encouraged to process information carefully, when they have a high degree of "social intelligence" and when they lack a strong need for explanatory closure ("negative capability"), and defends that such conditions prevail in the clinical setting. He concludes that "developing this line of argument further could take us toward a defense of psychoanalysis as a form of expertise in inferring the motives[23] of others within the clinical setting" (p. 209).

In a book dedicated to Freud's theory of dreams and in a couple of papers, the philosopher Michael Michael (2008, 2015, 2019) gives one general and three specific responses to Grünbaum. In the tradition of Hopkins and Lacewing, he defends both that non-Millian Inference to the Best Explanation is a strong form of inference and one routinely employed in science and that it is the standard according to which psychoanalytic inferences should be assessed. Specifically, he discusses Grünbaum's attacks on the exploration of free association to search for the causes of the dream content, on psychoanalysis' method of dream interpretation and on Freud and Breuer's evidence for their theory of repression, and he argues that such attacks rest on a "too quick and far too peremptory" (Michael, 2015, p. 139) dismissal of Inference to the Best Explanation. The philosopher is in line with the many who think that Grünbaum "has misunderstood both Freudian and scientific reasoning" (p. 118) and thus that he has failed to demonstrate that psychoanalysis is a pseudoscience. (Michael's argument is peculiar, though, in that he also thinks that psychoanalysis is *not a mature science either*, lacking as it does reliable instruments for the intersubjective checking of an important share of its typical evidence).

According to him, Grünbaum has been blind to the richness of Freudian reasoning because "he is beholden of a philosophical theory of explanation in which all genuine explanations either deductively entail or make statistically probable that which they explain" (Michael, 2015, p. 130). However,

[23]All in all, he disagrees with Grünbaum's conclusion that causes are impossible to reach in the clinical setting; but he claims that only proximal causes (particular motives, typical structures of motivation) are reachable there, and that clinical psychoanalysis would have to be aided by extraclinical research to find out about distal causes of behaviour (childhood events, the early formation of structures such as the superego, etc.).

> what often matters in scientific inference is not establishing a logical or statistical relationship between the hypothesis and the evidence, but understanding well enough how it might be that that hypothesized fact would give rise to that evidence. This is especially so in the historical sciences, like geology, paleontology, and evolutionary biology, which often involve explaining a unique set of data through a unique set of causal events (Michael, 2015, p. 139).

In a paper on Freud's method of dream interpretation, Michael (2008) faces the two charges which were directed to this method primarily by Glymour (1983), but also unreservedly endorsed and promoted by Grünbaum in many of the latter's works (see Chapter 1, section 2.2.): that the method allows analysts to defend any interpretation they wish and that it presumes, without any reason, that the products of free association have caused the elements of the dream when the only certain fact is the inverse, that is, that the elements of the dream are invoking the associations. Michael (2008) argues, nonetheless, that both charges are unfounded. By analysing Freud's interpretations of his own dream about Otto's illness, he demonstrates that, if the legitimate inferences to the best explanation involved there are characteristic of the method, it does not open a space, in the first place, for a presumption of reverse causation:

> Freud's interpretations solve two explanatory problems: (1) why did the dream contain those elements, and (2) why does the material uncovered during free association have numerous connections with different elements of the dream. It is in also solving (2) that the Freudian explanation of (1) is superior to other explanations. For it is a surprising coincidence that an association has striking connections with two or more different elements of the dream. Indeed, it is a surprising coincidence that any item has striking connections with two or more different elements of the dream. The more striking the connections, the bigger the coincidence, the less likely that this happened by chance. Freud's interpretations explain this coincidence by positing that the item in question had a causal influence on the content of the dream (Michael, 2008, p. 63).

Nor does it open a space, in the second place, for the proposition of numerous interpretations:

> […] the numerous connections uncovered during free association, along with other factors such as the emotional significance of the associations, the distinctiveness of the dream images, and the historical proximity of events, may logically compel a certain interpretation. The evidential base here is so rich that it is unlikely that there are numerous equally plausible explanations, so the 'many interpretations' problem does not arise (Michael, 2008, p. 63).

In another paper on Grünbaum's reproof of Freud and Breuer's theory of hysteria, Michael (2019) shows, contra Grünbaum, that the evidence reported in the doctors' famous book comprises a reasonable support for their twofold hypothesis that an unconscious memory is the cause of many hysterical symptoms and that this memory is unconscious because it has been under continuous repression (a hypothesis which Michael called "FB"). As we know (see Chapter 1), Grünbaum raised an alternative explanation for the evidence that a hysterical symptom dissipates when a trauma with features akin to the symptom's features is remembered and told, and the strangulated emotion related to the trauma is expressed: the analyst's exploitation of thematic plasticity conjugated with the participation of placebo-effects (the "P" hypothesis). However, Michael (2019) demonstrates that FB provides a more virtuous explanation than P because FB explains *all the relevant facts* about the symptom – why the symptom exists at all, why it began when it did (some time after the traumatic event), why it was thematically related to the traumatic event, why it disappeared when it did and in the circumstances that it did – while P explains *only this last fact*.

> My suggestion is that the explanatory loveliness of FB, and the relative explanatory unloveliness of alternative hypotheses, would have played a role in Freud and Breuer developing the high degree of credence they had in FB, whether or not they were able to articulate these grounds. If so, and if, as many contemporary philosophers of science hold, such explanatory loveliness constitutes valid rational force, then Freud and Breuer can be deemed to have been reasoning rationally (p. 48).

Brakel (2015) would argue that, even if Grünbaum were right about what Freud and Breuer avowed to be the evidence for their aetiological hypothesis, and about many other things under Freud's pen, such accurate readings would not be able to "fixedly define and determine psychoanalysis" (p. 59) as a clinical theory.

> To the extent psychoanalysis is and can remain a viable clinical theory, both analytic clinicians and theoreticians must be able to allow for modifications in the initial conceptualisation Freud provided – expanding some aspects, retracting others, all while retaining that which is essential. Grünbaum allows no such flexibility. As such, many of his arguments have a straw man opponent, vitiating the overall importance of his critique (pp. 59-60).

Brakel (2015) shows Grünbaum's sliding from bibliographical fidelity to theoretical inflexibility exactly in his dismantling of Freud's arguments about the evidence for the Repression Aetiology: although his dismantling is undertaken through the very premises of the arguments in question, he fails to consider crucial understandings on causality that have become available since such arguments were written. The Aetiology can be confirmed, according to Freud, with the evidence that the undoing of a repression was responsible for the disappearance of a symptom; but, says Grünbaum, as that evidence would not properly exist (none of Freud's arguments would rule placebo out), and as it was recently countered (all psychotherapies would perform equally well), then we would have proof that unconscious contents and conflicts do not have causal powers. But we know today, Brakel (2015) argues, that the eradication of a primary cause does not necessarily make its effect disappear, at least when organisms are at issue. For example, "eradicating beta hemolytic strep bacteria – the essential cause of rheumatic valvular heart disease – will do nothing to improve this cardiac disease once it has been established" (p. 67). She gives yet another example:

> Suppose a lung abscess is primarily caused by microbe A. Even if an antibiotic kills many of these initial causative agents, the infection can persist. How? With fewer populations of microbe A around, microbe B numbers can multiply, settle in, and perpetuate the infected abscess (p. 67).

Thus, Brakel (2015) swaps Grünbaum's values. While Grünbaum thinks Freud's "principle of therapeutic evidence" would become perfectly reasonable if one managed to rule out the placebo hypothesis, she thinks the principle is mistaken. On the other hand, she thinks that this fact would not (or should not) be important in psychoanalysis, for Freud's Repression Aetiology is *correct* anyway, as the 2013 experiment conducted by her and her colleagues demonstrates (see above in this chapter). It follows for her that merely knowing what has been repressed cannot not be curative in itself. However, again, she questions if believing in Freud's literal arguments is really a pre-requisite to being a psychoanalyst. She refuses to treat the matter of psychotherapeutic action as an *either-or* matter, what is curative having to be *either* insight *or* something unrelated to insight: "I along with any and every psychoanalyst will assert that both contribute, in variable ways, differently for different patients, and even for any particular patient at various stages in the treatment" (p. 68).

Finally, Brakel (2015) criticizes the grünbaumian and majoritarian attitude of taking psychoanalytic clinical theory to be the core of psychoanalysis; she defends that taking this core to be instead the psychoanalytic *general theory of mind* brings many advantages and precludes many of the problems raised by the philosopher. Proposing that this general theory is composed of five basic presuppositions[24], she argues that the theories deriving from this core, including the "clinical theory", are indeed "not only much more variable, […] [but also] much less reliable", and that the philosopher does not actually discredit the *foundations* of psychoanalysis when he discredits, for example, the sexual nature of pathogenic repressions or the curative effect of free association.

The doctoral monograph of Maximiliano Azcona (2016) is one of the few in Latin America to extensively face the "grünbaumian turn" in psychoanalysis. Throughout almost 180 pages, the Argentinian psychologist concludes that Grünbaum's approach is unsatisfactory for three reasons: his analysis and reconstruction of Freudian arguments is partially or totally mistaken; his argumentation is biased by the fact that his philosophical point-of-departure is not adequately spelled out and is unfairly taken as universal; his Millian methodological pretensions are limited and flawed by their own nature, but become even more so when employed in disciplines such as psychoanalysis.

---

[24]"Of the five, there are three assumptions, one methodological tool, and one corollary" (Brakel, 2015, p. 122): the assumptions of psychological continuity, psychological determinism and of a dynamic (psychologically meaningful) unconscious, the methodological tool of free association, and the corollary that there exists "two formally different types of mentation" (Brakel, 2009, as cited in Brakel, 2015, p. 126), namely, primary and secondary processes.

Azcona (2016) demonstrates, from textual evidence, that the Tally Argument was never for Freud a central epistemological point, as from the first psychoanalytic works Freud would have admitted the existence both of spontaneous remission and of partial successes in other therapies. Moreover, he would have never been dependent on the value of psychoanalytic therapy to appreciate the value of psychoanalytic knowledge: he would have taken many human products, such as jokes, taboos, religion, literature, bungled actions, etc.., as confirmatory support for his hypotheses and, in clinical cases, he would have disclosed that many therapeutic disappointments have guided his theoretical maturations. Azcona (2016) also appeals to a logical dimension in claiming that a causal relationship between cure and approximate truth is nonsensical regardless of the ingenious considerations on this relationship found in Grünbaum.

Like Michael (2008), Azcona (2016) observes that Grünbaum ignores the specific context of the method based on free association in stipulating its validity. This method would be suited to explore a multifarious, non-linear and motivational kind of causation, and would involve an abductive inference from an accumulation of converging evidence (an inference to the best explanation). Finally, the psychologist also criticizes the use of Millian canons to assess causal knowledge, not only in psychoanalysis, but in general; for him, such canons evoke a grossly simplified and unrealisable picture of the scientific method and, in any case, they have been originally thought for application to the domain of physical, experimentally assessable, phenomena. Psychoanalysis could only be judged beside the standards of astronomy, history, evolutionary biology, geology, comparative anatomy, and the other sciences with which it shares heritage (Azcona, 2016).

This section could not be complete without some defenders of the hermeneutic version of psychoanalysis that have remained immaculate after Grünbaum.

Matthew Hugh Erdelyi (1986) agrees that the Tally Argument "does not have the power, even in principle, to validate insights without extra-clinical corroboration" (p. 234), but, as any analyst would admit that false memories can be therapeutic and that true insights can be countertherapeutic, he argues it is Grünbaum's *enthusiasm* with the Argument's representativeness that deserves criticism. He argues that, in any case, the fall of the Argument would not have the catastrophic consequences for psychoanalysis announced by the philosopher because the practice would not be reduced to the "retrospective discovery of etiologies" (p. 234). (Erdelyi seems to understand this expression as just discovery of the pathogenic facts transpired in childhood and engraved in memory, and he seems to interpret Grünbaum as taking this discovery to be the greatest goal of psychoanalysis.). Instead, psychoanalysis would have become over time “a vehicle for insight” (p. 234), for the awareness and the working-through of thematic patterns in the patients’ here-and-now.

Erdelyi (1986) is thus adopting a hermeneutic version of psychoanalysis. For him, the Freudian project is to uncover meanings, and meanings would have a special nature: they would not be causes, nor would they be caused by something. In detecting that the context of an unfamiliar word establishes its meaning, one does not establish that the context has caused the word; one laughs at a joke, understanding its meaning, without knowing its cause. The psychoanalytic meanings are “not deterministic but determinate relationships, ones involving coherences, interconnections, patterns” (Erdelyi, 1986, p. 234). He agrees that free associations may not retrace the cause-and-effect chain but, once again, his hermeneutic perspective causes him not to attach much importance to grünbaumian imputations. Free Associations would “provide the contextual ecology from which meanings may be derived”, functioning after all “as microscopes/telescopes of the mind” (Erdelyi, 1986, p. 234).

In the same spirit, Anthony Storr (1986) contends that Grünbaum is “flogging a dead horse” (p. 260): the idea that Freud was a scientist or that psychoanalysis ever was or could be scientific would be found only in the head of a few antiquated, fundamentalist psychoanalysts. For him, Freud was, like Jung and Adler, one of the geniuses of semantics who were able to elaborate a coherent and persuasive narrative out of episodes of their patients’ lives. As it would be perfectly possible to make sense of people’s lives from different points-of-view, this enterprise could “never be either completed or scientifically validated” (Storr, 1986, p. 260). He cites authors for whom the detached and deterministic mindset which is needed for learning about the external world would not be exactly suited for the understanding of human motives.

Not long ago (in philosophy-years, at least), in the second edition of his book “For and against psychoanalysis”, Stephen Frosh (2006) rediscussed a worrisome objection raised against the hermeneutic account of psychoanalysis in the context of Grünbaum’s criticism, showing that this account, if refined, could still be the feedstock for a response to the philosopher. On the one hand, he tells us that, “increasingly, the idea that psychoanalysis seeks to express the ‘truth’ about an individual is falling into disuse” (p. 62) and comments that this fact could free psychoanalysis “from the invidious position of making claims to absolute knowledge” (p. 63). On the other hand, he concedes that this sort of truth-relativity would have a disturbing implication: it would leave “poorly articulated and slippery” (p. 69) the criteria for evaluating a psychoanalytic interpretation.

The classical hermeneutic criteria for evaluating any interpretation, he argues, are not enough. Clinical wisdom could discriminate among psychoanalytic interpretations excelling in theoretical coherence, inner consistency and intelligibility, as well as embrace interpretations running against such criteria; moreover, most psychoanalysts would argue that "there may be perfectly good 'narrative truths' that pass patients by or are actively refuted by them, because their truth lies in the complex of negative personal attributes and feeling states to which they draw attention" (Frosh, 2006, p. 68). For him the value of a psychoanalytic interpretation would be rather in how they affect the patient: a good interpretation would be "one that improves the flow of free association, leads to deepening affect and a better relationship with the analyst" and "is connected with symptomatic and relationship change" (p. 58). On the analyst's side, an interpretation should be risked when the impact of a patient's words in her or his emotions can be registered and reflected upon. Such criteria would be broader than the ones that are simply empirical, for they would require the articulation of subjective and intersubjective processes, including recognition of the place and the emotional states of the "researcher".

> Such criteria can be rigorously pursued and their adoption does not mean that permission is given for mysticism or for total nihilism in the articulation of theories. But their pursuit requires skills that are traditionally located outside science as narrowly conceived: interpretive and literary skills, as hermeneuticists and postmodernists might both suggest; also, self-reflective skills by which the investigator, whether clinician or researcher, demonstrates that her or his own feeling states have been engaged with and understood (Frosh, 2006, p. 78).

Frosh (2006) is here departing from the principle that "it is not possible to divorce reason from emotion in the way postulated by positivistic science"[25] (p. 77), and it would also not be advisable to divorce the two elements, at least when it comes to comprehending "data" and formulating theories about the human world. He expresses opposition to Grünbaum, therefore, in that the intersubjective processes between analyst and patient would *distort* the facts; for him, it would "rather supplies the necessary conditions under which significant facts can be known" (p. 60).

In a way, Frosh (2006) is insisting upon the fruitfulness of both hermeneutic psychoanalysis and the Tally Argument – but upon a more complex version of them. Insight would for him be legitimate not only when it is clear that the patient has been emotionally transformed but also when, before that, analyst and patient experience a whole process of coming to know self and other through emotional and other intersubjective entities. As for hermeneutic psychoanalysis, it remains for him the case, having been an analyst himself, that psychological "reality" is impossible to define "because it fluctuates and is reconstructed continually as it is enacted and produced in different contexts", but also that "some 'readings' of the unconscious are more forceful than others" (p. 72).

[25]A principle, by the way, with which mainstream neuroscience agrees today.

3.2.3 For and Against: Accepting the Conclusion but Rejecting the Reasons Leading to it

Finally, we have the meagrest corner of the room of responses to Grünbaum, where there are arguments accepting that the clinical-psychoanalytic method is unsuitable for the testing of psychological hypotheses, but rejecting the premises Grünbaum makes use of to get to this conclusion.

Although accepting Grünbaum's conclusion, Greenberg (1986) reminds us that nowhere Freud gives us the elements to decide whether his treatment had superior results to other treatments or to no treatment at all. Concerning clinical cases, Freud

> describes in detail only four […] he had seen using psychoanalytic treatment. Of these only one showed any evidence of significant improvement [the Rat Man]. It is both striking and curious that Freud chose to demonstrate the usefulness of psychoanalysis through the presentation of largely unsuccessful cases. Indeed, we were forced to conclude that Freud never presented any data, in statistical or case study form, that demonstrated that his treatment was of benefit to a significant number of the patients he himself saw (Fisher & Greenberg 1977; 1985). (Greenberg, 1986, pp. 240-241)

Plus, toward the end of his life, Greenberg (1986) argues, Freud becomes less and less prone to linking theoretical truth with healing and to ascribing durable and powerful benefits to his treatment, a tendency that had its notable apex with the 1937 paper "Analysis Terminable and Interminable". Indeed, the scholar finds puzzling that, despite the Tally Argument, Freud warns us many times of the dangers in analysts having too much an interest in patients' cures, namely, of the dangers of a *furor therapeuticus*.

The psychologist Hans Strupp (1986) believes that psychoanalysis would have been less a victim of controversy had Freud confined himself to the therapeutic enterprise, refraining from his ambition to develop an abstract, "universal theory to account for all aspects of human behavior" (p. 260), and that psychoanalysis' future would lie in the observation of recorded "transactions between patient and therapist in the here-and-now" (p. 261). He agrees with Grünbaum that the NCT is false because other psychotherapies perform equally well, but thinks that the therapist's influence should not be reproachfully called "suggestion", nor a "placebo effect", "thereby implying that there must be 'special' psychological mechanisms in formal psychotherapy" (p. 261). He believes that all psychotherapy can do is to "persuade the patient to give up unrealizable wishes" and to "promote cognitive changes in how the patient views himself and the world" (p. 261), and that the distinction between specific and common factors in psychotherapy, between its technical and interpersonal factors, is a pseudodistinction. He argues that therapeutic successes cannot arbitrate about the correctness of any theory of psychotherapy because therapeutic change can be attained through a miscellany of undeterminable means. Be that as it may, psychotherapy for Strupp would be based on psychological influence and we would learn little about its character “when we attribute it to ‘suggestions’ or ‘placebos,’ which are merely synonyms for lack of scientific knowledge” (p. 261).

Strupp (1986) also asserts that Freud’s attempts to explain “neurosis choice” through specific “pathogens” was fruitless, not because his method was flawed, but more simply because a person’s difficulties in living would be “always multidetermined and almost inextricably woven into a person’s ‘character’ as it has evolved over many years” and because “the term ‘neurosis’ is itself of questionable scientific value” (p. 261).

The philosopher Anthony Derksen (1992) also criticizes Freud’s Tally Argument but with premises that are different from the ones raised by Grünbaum. He maintains that the Argument is far from sophisticated – in a reference to Grünbaum’s rhetorical affirmation about it[26] – because, “supported, the argument is acceptable but nothing special” and, “unsupported, it is just a direct denial” of the three grünbaumian charges[27], that is, a *petitio principii* (p. 95). The Argument may come down to the medical truism that the cause of an illness must be annulled for there to be a cure; analogously to the claim that the disease caused by bacterium X cannot be overcome by removing bacterium Y, the Tally Argument’s trivial core is “that the patient must become aware of the *right* unconscious pathogen, for without the right insight the pathogen would remain intact” (p. 94). The philosopher argues there is nothing methodologically wrong here – but also nothing impressive. In the case where the psychoanalytic theory about the causes of mental illness is not supported, however, the Argument would have the same methodological sophistication of modern-day evangelists:

> Consider modern-day evangelists who preach that only knowledge of God provides real happiness. We may now question whether it is knowledge of God that makes their followers “really” happy. Perhaps it is just the feelings of being loved and being part of a group that bring about the happiness. We can challenge the evangelists along familiar lines: How do they know that it is not suggestion which makes their followers speak of real knowledge of God? How do they know that their happiness is not just a placebo effect caused by the new group life? And lastly, how do they know that it is the knowledge of God that causes their real happiness? That is, the evangelists face the same three charges that also troubled Freud (p. 94).

---

[26]“The epistemological considerations that prompted Freud to enunciate his Tally Argument make him a sophisticated scientific methodologist, far superior than is allowed by the appraisals of friendly critics like Fisher and Greenberg (1977) or Glymour (1980), let alone by very severe critics like Eysenck" (Grünbaum, 1984, p. 128).

[27]Derksen (1992) splits Grünbaum’s critique into three major charges in much the same way as I do in this doctoral thesis. He calls them “the Causality Charge”, “the Suggestion Charge” and “the Placebo Charge” (pp. 75-76).

If the psychoanalytic theories of illness and therapy are revealed as uncertain as religious theories of the same kind, Freud's argument could do nothing but beg the question – that is, just assume that his theories are true without further qualification. Thus, Derksen (1992) sees "no reason to give Freud the stature of a sophisticated methodologist" (p. 95).

Derksen's argument is very similar to the one given by Paul Wachtel in the BBS debate (Wachtel, 1986). There Wachtel (1986) perplexes before Grünbaum's opinion that the Tally Argument testifies to the sophisticated methodologist in Freud, contesting that the Argument is, unlike what the philosopher of science defends, utterly question-begging, hence "an instance, not uncommon in Freud's writings, in which overreaching assertion gets the best of [...] [his] methodological superego" (p. 264). With the Argument, Freud would have proved what should be proven – his clinical theory, which would have the NCT as an empirical consequence –, by assuming the NCT to begin with. To assume that cure will come for the patient only if the analyst's formulation tallies with what is real is untenable unless one is already departing from psychoanalytic theory. Wachtel (1986) alludes to a passage in *Foundations* in which Grünbaum argues that, *if* the analytic treatment has the therapeutic monopoly that the Tally Argument has given to it, then it can take credit for the recoveries of its patients without statistical comparisons with control groups; then, Wachtel (1986) asks: "but how can one know if it has such a monopoly unless one does the comparisons?" (p. 264).

3.3 CONCLUSION

In this chapter, I have presented some remarkable attitudes and strategies in the responses to Grünbaum's critique of psychoanalysis and distributed such responses into five groups. Among the ones who agree with his conclusion that it is impossible to test causal-psychological hypotheses in the clinical-psychoanalytic context, there are the ones who compliment and complement his arguments, the ones who see them as endorsements of experimental research on psychoanalytic hypotheses and the ones who disagree with his premises. Among the ones who disagree with this conclusion, I have distinguished those who admire his premises from those who question them. This distribution may be useful for philosophers of psychoanalysis interested in Grünbaum's work. It would become much more useful, for sure, if it included a systematic identification of kinds of attitudes and strategies within each group. This is a worthwhile project that I may pursue in the future, but the function of this chapter was just to show how diverse the debate in question is and some attitudes and strategies that influenced my own response to the German-American philosopher.

Before heading to my responses, however, we should discuss some of the methodological, theoretical and epistemological principles that guided them.

4 SYNTHESES AND PROPOSITIONS, ORIGINS AND HORIZONS

4.1 INTRODUCTION

In the two preceding chapters, we have understood Grünbaum's diagnoses about the epistemological foundations of clinical psychoanalysis and we have gone through the main lines of reply to such diagnoses. There comes the moment to give to his arguments my reply, which must answer the question of whether the practice of clinical psychoanalysis can rise to the status of a reliable and reasonable context of psychological research. This reply shall be presented in the approaching three chapters, each of them discussing a different set of problems. In the chapter immediately following the one you are now reading, I shall be concerned with the business of inferring causes from data born in the clinical context, a data that is thereby not subject to experimental controls. The problem of data contamination by suggestion shall be tackled in the fifth chapter. We shall finish our trek by elaborating on the kind of link that can exist between truthful interpretations and therapeutic effects. This third chapter, thus, is a connecting route; in it, I shall present the methodological, theoretical and epistemological background of my forthcoming reply.

Along the responses we have assembled in four groups in the second chapter, we should still note a couple of general approaches: a thorough reading of Freudian and Grünbaumian texts, followed by an urge to reinterpret them and point to what has been ill-interpreted in them; and another approach in which the authors, besides purely inspecting and contemplating the issues raised by Grünbaum, somehow seek to see these issues from another perspective, to redirect them, and even to solve them. I shall label these approaches, respectively, analytical and propositional-synthetical, and discuss their features. With this, I shall question the usefulness of discriminating only between normative and descriptive approaches in the philosophy of science, for both approaches tend to exhibit features that drive them into the domain of what I am calling the analytical approach.

The approach of my reply shall be propositional-synthetical. This problem-solving approach presupposes the establishment of, so to speak, an origin – a synthetical model of the specific scientific practice to be mended, regulated, adjusted, if needed – and a horizon – a synthetical model of scientific practice in general in consonance of which the epistemologist can propose the measures to mend, regulate, adjust the specific practice. In this chapter, we shall establish a minimal version of psychoanalysis' theory and clinical method, as well as a minimal set of principles characterizing what I take to be a proper scientific conduct. Toward the end of the chapter, we shall think about what kind of scientific knowledge we can expect from clinical psychoanalysis.

4.2 A PROPOSITIONAL-SYNTHETICAL APPROACH

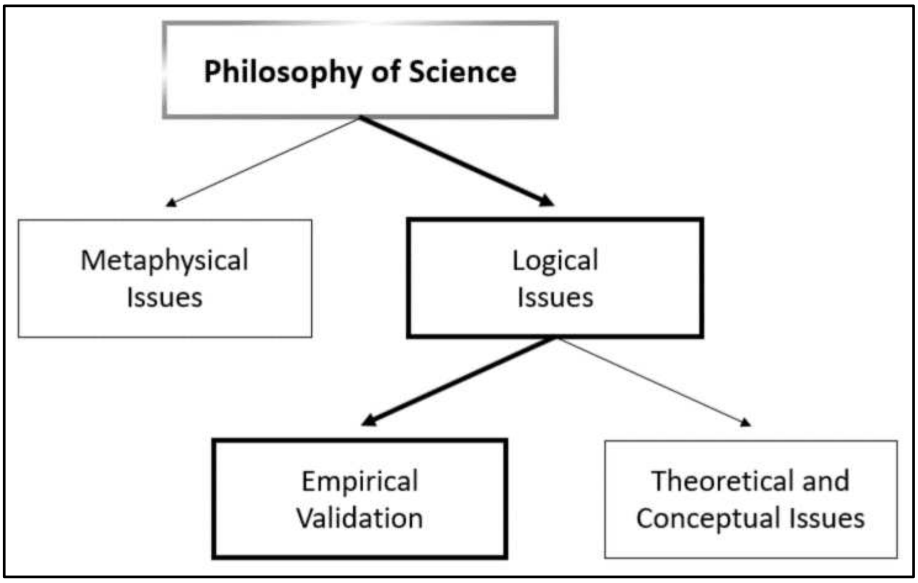
Before exploring kinds of approach in the philosophy of science, we should note that the branch may exhibit different kinds of *concerns*. More broadly, part of the philosophers of science have concerns of a *logical* kind, and the remainders, concerns of a *metaphysical* kind. Questions addressing the conditions under which scientific knowledge is or ought to be produced are very different from questions addressing the ultimate nature of this knowledge. Questions about reasoning and method are logical, while questions about whether there is something out there independent of our minds and theories, and about how much we can reach it, are metaphysical. By stating this contrast, I am endorsing a point Ian Hacking presented in his “Representing an Intervening: Introductory Topics in the Philosophy of Natural Science”:

> […] rationality has been one of the two issues to obsess philosophers of science. We ask: What do we really know? What should we believe? What is evidence? What are good reasons? Is science as rational as people used to think? Is all this talk of reason only a smokescreen for technocrats? Such questions […] are traditionally called logic and epistemology. […]
> Scientific realism is the other major issue. We ask: What is the world? What kinds of things are in it? What is true of them? What is truth? Are the entities postulated by theoretical physics real, or only constructs of the human mind for organizing our experiments? These are questions about reality. They are metaphysical. (Hacking, 1983, pp. 1-2)

The academic work you have in hands lies within the logical branch of the philosophy of science, for it seeks to unravel the rationale that enables one to appraise certain scientific or quasi-scientific hypotheses. But part of such a rationale has an altogether non-empirical character – there are studies of internal consistency, of intertheoretical entailment and consistency, and of conceptual evaluation and innovation, for example (see Kukla, 2001) – and I am interested, rather, in how certain hypotheses would be generated, corroborated, supported, confirmed, refuted and changed by *empirical data*. Thus, we must constrict further the place of this work within the philosophy of science: it is interested in the relationship between the empirical and the theoretical; in other words, in matters of *empirical validation*.

Below, we see a figure summarizing the links between the different kinds of issues addressed in the philosophy of science and indicating which of these kinds the work you have in hands is interested in:

Figure 2 - The Issue of Empirical Validation in the Philosophy of Science in Relation to the Other Kinds of Issues Addressed by This Branch



Source: Elaborated by the author (2021).

Allow us now to talk about approaches. The most famous taxonomy of approaches in the philosophy of science is related to the most debated issue of the discipline since the second half of last century: what should be the role of data from the history of science in the models of general and particular scientific practices. The distinction between works departing from science's actual *modus operandi* throughout history and those attached to a demarcation project is well established and still useful in the discipline. The approach of the former is usually called *descriptive* and the approach of the latter is usually called *normative*.

The normative philosopher of science inspects the actions and publications belonging to a particular epistemic practice, reconstructs their logic and judges this logic according to prior canons of validity, soundness, strength and cogency. The aim of such a philosopher is to state if and in which respects the practice in question is good or bad science, or even no science at all. The aim of philosophers of science we name descriptive, on the other hand, is not to judge, but to make intelligible a particular, purportedly scientific, practice, illuminating and unfolding this practice's contingent epistemic regime.

McMullin (1970) argues for a similar distinction in his taxonomy of approaches to the Philosophy of Science (PS). To him, there is an External PS, in which one looks "outside science itself to some broader context" deriving thus "a theory of […] how [scientific inquiry] […] should proceed" (p. 24); and an Internal PS, in which there is "a careful 'internal' description of how scientists actually proceed, or have in the past proceeded" (p. 26). Defending this last approach, Lebrun (1977) affirms that each science must be considered in its singularity and appreciated, not as a set of truths, but "philologically". He affirms that "it is possible to confer [to a science] the status of a text" (pp. 137-138).

McMullin and Lebrun were in the 1970s living the aftermath of Thomas Kuhn's theses in "The structure of scientific revolutions", published in 1962. One the most cited academic books of all time, it has demonstrated that history – that is, actual cases of scientific practice and development – contradicts many of the ideas cherished by the philosophers of science. "The Structure…" concluded from such cases that – to use Hacking's (1983) summary –

> There is no sharp distinction between observation and theory.
> Science is not cumulative.
> A live science does not have a tight deductive structure.
> Living scientific concepts are not particularly precise.
> Methodological unity of science is false: there are lots of disconnected tools used for various kinds of inquiry.
> The sciences themselves are disunified. They are composed of a large number of only loosely overlapping little disciplines many of which in the course of time cannot even comprehend each other. […]
> The context of justification cannot be separated from the context of discovery.
> Science is in time, and is essentially historical. (p. 6)

The book contradicted every common thesis of the movement that was dominant at the time. For logical empiricists like Rudolf Carnap, Carl Hempel and Karl Popper, only logic should account for the relationship between the empirical and the theoretical in the sciences: the history of the sciences could be a source of good or bad *examples*, but not of insights. Although the logical empiricist's project was declared dead (Passmore, 1967), its influences are still found in the works we are here calling "normative".

Imre Lakatos and Larry Laudan elaborated models of science's practice and development that could accommodate Kuhn's approach and insights to the bright indications of science's rationality. Lakatos (1999) agreed with Kuhn that Popper had given a poor description of how sophisticated science operates and evolves. He argued that, in the history of science, a research program – a chain of theories with a common "hard core", that is, a common set of founding hypotheses – is (and should be) prolonged not simply because it has historically resisted refuting attempts but because it has exhibited a *growing power to predict with exactitude novel and unexpected facts*; he found that, in history, the "hard core" of the theories with such power is *justifiably* armoured by the updating of its auxiliary hypotheses. Influenced by Kuhn, Laudan (1978) renounced the positivist idea that science progresses as far as it is able to accumulate knowledge throughout the critical and invigorating moments of its history. For him, it does not matter if such moments are origins of radical transformations, for scientific progress is characterized only by the *growing resolution of empirical and conceptual problems*.

Kuhn was the spark to the methodological scruples we can witness in recent epistemology. His book was followed by an intense debate on how difficult could the alliance between history and philosophy of science be. Ronald Giere (1973) downgraded it to a "marriage of convenience"[28] (p. 283) in his review of a 1970 volume of the "Minnesota Studies in the Philosophy of Science" dedicated to the promising alliance. He concluded:

> […] none of the contributors [in the volume] attempts, or even cites other attempts, to show in a systematic way just how philosophical theses about theory choice may be supported by historical case studies. To raise this issue is not necessarily to hold dogmatically to a distinction between the descriptive and the normative. On the contrary, I would argue that all norms have their roots in facts. The general problem is to show that philosophical conclusions may be supported by historical facts and just how this comes about. Until this is done, the historical approach to philosophy of science is without a conceptually coherent programme (p. 292).

In his "More than a marriage of convenience…", Burian (1977) defended that the complex relationships between reconstructed theory and real theory must be kept in mind by the "logicist philosopher of science" if this philosopher wishes model and phenomenon to be similar. According to him, theories are often expressed in mutually inconsistent forms in textbooks, articles, etc., and a good reconstruction will only be achieved by facing this inconsistency, that is, the theories' historical character.

Decades after his cautious paper, Giere delivered a less cautious one. He concluded that the philosophy of science "should be transformed into something like the theory of science" and that it "should be naturalized" (Giere, 2012, p. 61). To wit, he concluded that philosophers should construct a theory of how science works just like a chemist would construct a theory of how a molecule works, and that they should lose all hope to find "autonomous standards for the practice of science" (p. 61). He noted that, since 1973, the interaction between history and philosophy of science became more and more substantial and that philosophers of science were no longer writing "about black ravens and shadows of flagpoles" but taking examples from "real science" (p. 63). In the same paper he holds, though, that his perspectives about the relationship between the two disciplines "are similar to what they were at the beginning", for each remains exhibiting "different goals and methods" (p. 63). He argues that:

> If, as Kuhn had suggested, history of science can serve as evidence for philosophical claims about science, we have a circle. Before one can use the history of science as evidence one needs a philosophical account of what constitutes evidence. […] And this problem follows after the

[28]The first footnote of Giere's "History and Philosophy of Science: Thirty-Five Years Later" is somewhat Freudian: "As a sidelight, I should mention that my choice of the marriage metaphor might have been influenced by the fact that I was at that time in the midst of a divorce" (Giere, 2012, p. 59).

obvious prior question of how factual claims could ever be evidence for normative claims[29] (Giere, 2012, p. 60).

Although today the borders between normative and descriptive works are more permeable, it is possible to cite some epitomes of each approach. Grünbaum's criticism of Freud is on the normative side, along with the logical empiricists already cited. "Laboratory Life: The Construction of Scientific Facts", by Bruno Latour and Steve Woolgar, and "Constructing the Subject: Historical Origins of Psychological Research", by Kurt Danziger, are remarkable examples of historical-sociological approaches apart from Kuhn's "Structures".

Some responses to Grünbaum's critique of psychoanalysis do not fit exactly into this normative-descriptive dichotomy, even though they usually make use of normative and descriptive works in the philosophy of psychoanalysis. They aim to go beyond the latter, because they do not just demonstrate the internal coherences of psychoanalysis or provide criteria to conclude that psychoanalysis is good, bad, or pseudo-science; they also redirect or even solve the problems raised by Grünbaum. The authors of these responses portray epistemological issues around the psychoanalytic method and indicate adjustments and fresh perspectives with the goal of making this method reasonable; Grünbaum's critique is only a trigger and an organizer in the process. Recent examples are Michael (2008, 2015, 2019), Lacewing (2012a, 2012b, 2013a, 2013b, 2018), Lynch (2014), Brakel (2015), Kaszubowski (2016), Azcona (2017).

These philosophers of psychoanalysis pay attention more to some parts of the pertinent data than to others, that is, they are *not committed to a perfect balance between all of the pertinent data*, like normative and descriptive philosophers usually are. Moreover, instead of sticking to what the data *explicitly indicates*, as normative and descriptive philosophers tend to do, they try to see what it *implicitly offers*. This relative freedom serves the purpose of solving the problems exhibited by the epistemic practices of psychoanalysis. For the methods and rationales of psychoanalysis, they propose concepts, dialogues, emphases, etc. – some less innovative, some more. Their intellectual energy is channelled to a *problem-solving attitude*.

In the collective response to Grünbaum in the *Behavioural and Brain Sciences*, Robert Holt defends such an attitude:

> […] it is neither desirable nor important to put psychoanalytic theory on an imaginary prisoner's block and try it in a methodological court, seeking a verdict of innocent or guilty. Obviously, it has serious shortcomings. The more interesting question is what can be done about it. Can the theory be salvaged? Can we not find ways of using the data yielded by psychoanalytic treatment, which have impressed so many as uncommonly rich and revealing? If Grünbaum could be

[29]Note that here he is no longer certain that "norms have their roots in facts" (Giere, 1973, p. 292).

> persuaded to apply his formidable talents to these questions, all who still hope for a scientific psychoanalysis would be even more in his debt (Holt, 1986, p. 244).

In the same spirit, Michael Ruse stresses there that Grünbaum only shows what is wrong; that he inspires but does not answer an all-important question – “where do we go from here?” (Ruse, 1986, p. 256). Grünbaum would show that “Freud was no Newton of the mind”, but he should answer to the suggestion that “Freud may nevertheless have been its Kepler” (Ruse, 1986, p. 257).

I should, thus, oppose such philosophies of psychoanalysis to purely normative and descriptive ones. From now on, I shall call the latter “analytical”, because all they aim is to thoroughly characterize the epistemic practices of psychoanalysis and/or the concept of science, and the former, “propositional-synthetical”, because, although they make use of such accurate and exhaustive accounts, they also raise the question “where do we go from here?”; because they try to reach syntheses, instead of just analyses, and to propose diversions and solutions, instead of verdicts.

The spirit of the “propositional-synthetical approach” can also be seen in Quine’s (2013) description of the process of explication, even though the latter is applied to concepts and the former to methods:

> [In explication,] we do not claim synonymy. We do not claim to make clear and explicit what the users of the unclear expression had unconsciously in mind all along. We do not expose hidden meanings [...] we supply lacks. We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions (p. 238).

The work of the “analytical approach” is confined to data: the history of science, models of rational science, psychoanalytic publications or reports, and first-hand observations of psychoanalytic practice. The analytical approach adopts a passive stance toward the data, the “given”; the approach does not add to the “given”, does not change it. Because of this limit, there is energy left to be as encompassing and faithful as possible.

In most of the defensive responses to Grünbaum, we see an analytical approach. In most of them, the scholar either shows how Grünbaum did not characterize Freud or post-Freudians properly – how he ignored this or that text, how he did not put the figure in its proper ground, etc. – or else accepts Grünbaum’s “Freud” but see nothing to be adjusted in it – from considering psychoanalysis an exceptional enterprise in the history of science, or something along such lines. This is to be Freud’s barrister. Propositional-synthetical philosophers, on the other hand, are not quite the barristers of Freud; nor are they his prosecutors or judges. The work of the propositional-synthetical approach is like the work of an assistant – of a social assistant, to insist with the judicial metaphor.

Now that we have this new dichotomy in hands, I can state that my responses to Grünbaum in the three following chapters shall not be analytical, but propositional-synthetical. In order to resolve, dissolve or divert the logical problems of clinical psychoanalysis, I shall explore the following general strategies (among others): to check whether scientific research traditions with objects, methods and problems similar to those of clinical psychoanalysis could provide it with helpful insights and tools; and to characterize and inspect the alternative hypotheses Grünbaum claims to undermine psychoanalytic method and theory, and then to question if they are actually able to do it.

I would like to argue a little more about the usefulness of this new dichotomy. I have affirmed before that purely normative and descriptive approaches are both analytical. But why should we find it a good idea to put Grünbaum and entrenched Freudians on the same side? Why put such dissonant approaches under the same label?

The classical normative-descriptive dichotomy may lead us to think that the only tasks left to philosophers of science are either to harshly judge or else to covertly bless the science under study. I, on the other hand, think that the task of all philosophers is also to *affect* their object of study. We can, for sure, benefit from art, be ethical and political, make sense through language, acquire knowledge, do science, etc., without a philosopher; but making the foundations of such actions clearer enables us to handle and refurbish them, enables us to maintain in sight the horizons we have chosen to maintain. In putting normative and descriptive approaches together on one side and a propositional-synthetical approach on the other, I am calling attention to an artificer's attitude that is undeniably valuable to philosophy.

Above, the task to covertly bless the science under study would be the task of a descriptive approach. This is a bold suggestion; a descriptive epistemologist like Lebrun (1977) would tell us that "a text does not have to be relativized (or, inversely, allegorised)" and that "it asks only to be read and reread, as game cards, opened and shuffled innumerable times" (p. 141), in an attempt to convince us that one does not judge if one only unfolds the specific decisions, backgrounds, networks, etc., standing for a science's rationality. However, to do that and nothing else is to imply that everything is fine with that rationality. Although it declares itself even-handed, the purely descriptive epistemology is ultimately an endorser; it is as judgemental as the purely normative epistemology. This is another reason to put descriptive and normative approaches together: they are two sides of the same coin.

A last point in favour of the new dichotomy. We may think that the criticism present in the normative philosophy of science has a propositional tone, but this is not quite the case. An example: when Grünbaum concludes that the clinical method cannot test psychoanalytic hypotheses, hence that we should test them through experimental and epidemiological methods, would not he be solving a problem? Grünbaum sets his Freud, sets his epistemological canons, compares both, and concludes that one is very far from the other. The existing methods that satisfy Millian canons could welcome the objects of psychoanalysis (psychopathology, sexuality, etc.), he thinks (but he also thinks it is highly likely that they will falsify Freud's specific hypotheses). This is not to be propositional, though: the method is simply discarded after its problems are identified.

Before the approach and strategy I have chosen can be operated, some questions must be answered.

If our approach is problem-solving, to which state must we conduct our science so we can say we solved its problems? Before that, how can we be sure that its claimed problems are actual ones? And how much can we "solve the problems" of the clinical method of psychoanalysis without *disfiguring* it? What is *the* clinical method of psychoanalysis, by the way? When mentioning sciences and scientific research traditions similar to psychoanalysis, and when questioning Grünbaum's undermining hypotheses, I am assuming a definition of both "psychoanalysis" and "good scientific practice" that I have not yet disclosed. A propositional-synthetical approach to the epistemology of the clinical research in psychoanalysis must have an *origin* and a *horizon*. Otherwise, one does not know what one is attempting to solve, and how one could solve it.

Above, I have hinted at what a "barrister" of Freud in a response to a critic like Grünbaum consists in. Both Grünbaum and the barristers of Freud consider Freud the ultimate authority to state what psychoanalysis is. They should concede, though, that a science is more than the text of its founders; it is the interest for an object, that is, for a set of phenomena, and a quest for methods that can inform us of true and content-rich statements, laws or theories related to this set (Schurz, 2014). A science is a project; it should have principles that should be looked for in the text of its founders, but even the most basic of such principles could be disputed. Nevertheless, to accept such a bold claim is to accept the burden of thinking which principles can be disputed without disfiguring the science in question. If not *strictly* in Freud and other classical authors, where is, after all, psychoanalysis' core?

The same can be argued as regards a definition of "good science". The discussion above revolving normative and descriptive philosophies of science were already a hint to the demarcation dilemma. Is this definition in the actual practices that are called sciences or in rational standards rooted in logical systems and intuition?

Needless to say, there will always be a dispute over what the cores of both psychoanalysis and science are; in both matters, the historical and the logical are always interacting. The only honest, scholarly way to find an origin and a horizon to our propositional-synthetical approach without being too formalist (that is, too *analytical*, as the term was used here) is to come up with *minimal* definitions of both that respect the diversity of existing definitions and at the same time *allow for the malleability necessary in problem-solving operations*. In other words, we need to be a *soft descriptivist* for a definition of "psychoanalytic" and a *soft normativist* for a definition of "scientific".

Such definitions shall be presented along the sections of this chapter[30].

4.3 A CLINICAL PSYCHOANALYSIS

4.3.1 The Clinical Method

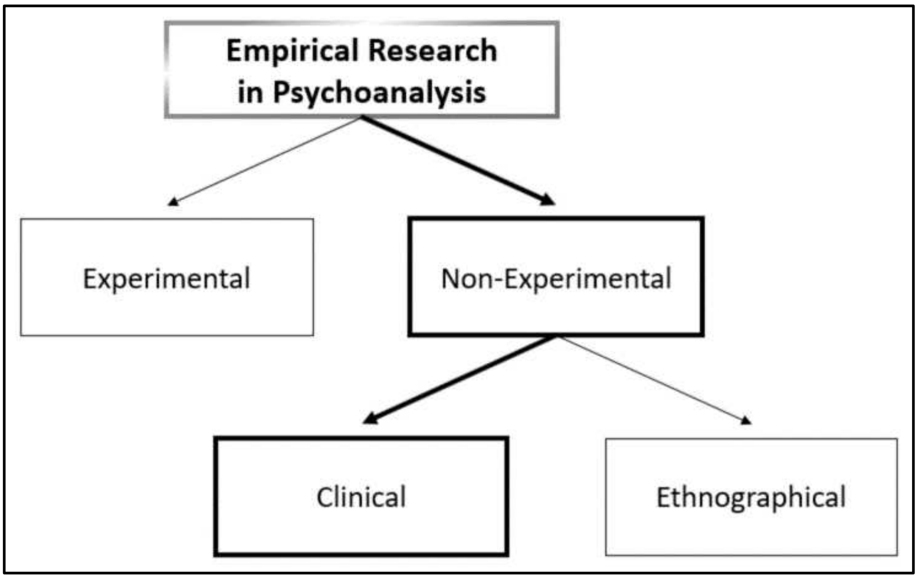
Research in psychoanalysis can be philosophical or empirical. This is a philosophical research in psychoanalysis about one kind of empirical research in psychoanalysis – the clinical research, a research that makes use of the clinical method. This taxonomy may seem amiss for psychoanalytic authors who consider the term "empirical" and even "research" as equating the term "non-clinical". But, historical whims aside, this attitude betrays the concept of "empirical research" – a stringent search through an interchange with the sensible world. As such, it can only be contrasted with the "philosophical research", which is one undertaken only through mental operations. By the way, it would be wiser to put the term "conceptual research" within the scope of the latter, not confusing it with "clinical research", for concepts are not determined by empirical data; they cannot be taken as true or false, but only as fertile or infertile (Kukla, 2001).

---

[30]They should, without a doubt, be questioned. Why would they be good enough? Where did they come from? They may come from a too personal pondering of logical and documental supplies, so they may be *too* soft or even miss the core. Nevertheless, my concern here is only over whether they are explicit enough to enable a qualified debate with this thesis in the future.

Empirical research in psychoanalysis can be experimental or non-experimental. The former consists in a use of the experimental method (whose elements shall be described in a moment) to test or advance psychoanalytic theory, and may have as starting points both therapeutical and psychological questions[31]. In psychoanalysis, clinical research is, of course, non-experimental. Another kind of empirical research in psychoanalysis that makes use of non-experimental methods is the ethnographical kind; the so-called "applied psychoanalysis" can be included in it, as well as any research over the culture of psychoanalytic communities.

Figure 3 - The Clinical Research in Psychoanalysis
in Relation to the Other Kinds of Empirical Research in Psychoanalysis



Source: Elaborated by the author (2021).

Clinical research has a special place among the other kinds because it has predominated in psychoanalysis' history and has given birth to its classical hypotheses, which, as expected, pervade any one of its empirical and philosophical researches. As there are some psychological and medical hypotheses whose factors cannot, in an ethical or practical sense, be reproduced in an experiment, the clinical context is a sometimes-indispensable wellspring of knowledge for psychology and the rest of the health sciences. Despite this projection of the clinical research in psychoanalysis and elsewhere, its epistemological vigour is frequently put under dispute.

The concept "clinical research in psychoanalysis" suggested here asks for some elucidation. It is, first, a kind of research that generates and tests *hypotheses*, which, unlike concepts, can be true or false (or, more accurately, probable or improbable), and thus *must be*

[31]Ranging from neuropsychological to anthropological questions.

limited by the reality outside of our mental operations. (It is *empirical*, after all.) Second, these hypotheses refer to utterances and circumstances of *clinical cases*, in which there are *suffering subjects* and *no experimental control*. Thus, it should also not be confused with "clinical" researches in health sciences like pharmacology (the so-called "clinical trials"), since the latter present an experimental design. From this, it does not follow that a clinical research in psychoanalysis must, to deserve this name, present a total lack of control over conditions and conclusions; actually, we should suppose that the lack of cogency presented by case studies can be solved with some control moves. The possibility of being an "idiographic nomotheticist" (Kächele, Schachter & Thomä, 2009, p. 22) will be analysed throughout the chapters.

But what does an experimental control imply that would be out of our definition of clinical research? One may identify four things. Experimenters always manipulate the factor under study; and, so that no nosy factor is able to play a part in the outcomes, they randomise the subjects' selection and assignments and standardize every conduct and condition. Most importantly, maybe, experimenters also are generally uninterested in deviations once they infer statistical significance. The clinical research, on the other hand, is mostly an *ex post facto* inquiry, does no randomisation and *almost* no standardization – *almost*, as it may have standard ways to collect data –, and, naturally, it has an idiographic approach – even though it can infer regularities from the cases it studies.

These distinctive features of clinical research are explained by its belonging to a context of illness and care. The subjects of this kind of research are always suffering at some level and asking for help with higher or lower freedom – they are patients. Given this simple condition, it follows that: 1) subjects willing to be part of a clinical research can only come up by convenience or opportunity, hence there can be no random selection of them and no random assignment of them to artificial groups; 2) in the clinic, the causal mechanism of the phenomenon at study (the illness or discomfort) has already occurred, hence there can be no manipulation of hypothesized factors; 3) the most urgent aim of the clinic is the care, not the research, hence there can be no rigid standardization of the conditions of the setting and of the conducts of the researcher; 4) a suffering, especially a mental suffering, always manifests itself in a singular manner, hence the necessity of an idiographic spirit in the clinic.

It would be unethical to publicly convoke tormented persons to a research (1), to produce an illness (2) or to satisfy our diligent curiosity in detriment of the patient's mental health (3). And it would be scientifically dishonest to report only phenomena that can fit in an eminent nosology and forget the others (4).

Clinical research has, then, an *ex post facto*, or after-the-fact, design. In this design, "the investigation starts after the fact has occurred without interference from the researcher" (Silva, 2010, p. 465). The dependent variable (the fact or effect) is chosen and analysed "retrospectively in order to identify possible causes and relationships between […] [it] and one or more independent variables", whereupon "the researcher can eventually adopt a prospective approach, monitoring what happens after that" (Silva, 2010, p. 465). It is duly undertaken when "the phenomenon occurred naturally", when "it is not practical to manipulate the independent variables" or "the control of independent variables is unrealistic" or yet, importantly, "when such manipulation of human participants is ethically unacceptable (e. g., delinquency, illnesses, road accidents, suicide)" (Silva, 2010, pp. 465-466).

Ex Post Facto is one among some nonexperimental designs, the distinctive component of which is that "the groups already exist and the experimenter [sic] cannot or does not attempt to manipulate an independent variable" (Lohmeier, 2010, p. 911). Although participants are not assigned to groups in such designs, the researcher

> can usually still determine what is measured and when it will be measured. So despite the lack of control in aspects of the experiment that are generally important to researchers, there are still ways in which the experimenter [sic] can control the data collection process to obtain interesting and useful data. (Lohmeier, 2010, p. 911)

The atom of clinical research is the case study. According to Yin (2018), a case study is an empirical study of a contemporary phenomenon (the "case") "in depth and within its real-world context, especially when […] the boundaries between phenomenon and context may not be clearly evident" (p. 50). As a result, a case study must cope with "many more variables of interest than data points [cases]", hence benefiting from a "prior development of theoretical propositions to guide design, data collection, and analysis" and depending on "multiple sources of evidence, with data needing to converge in a triangulating fashion" (p. 50). Researchers may make a case for a case study when desiring to understand complex social phenomena over which they have little or no control (Yin, 2018).

Finally, we are able to compare the experimental and the clinical methods:

Table 3 - A Comparison Between the Experimental Method and the Clinical Method

	Experimental method	Clinical method
Selection and consent	The experimenters randomly select the subjects, who must then consent to being part of the research before it starts.	The subjects are patients; they were suffering and came to the clinician asking for care. The patients' data will only be used for a public research with their consent at any time of their caring process.
Assignment to groups	The experimenters randomly assign the subjects to two or more groups in order to compare the outcomes of the presence or absence of the factors under study.	In a multiple-case research, one may group subjects according to their traits after single-case researches are finished.
Manipulation of factors	Factors are hypothesized and manipulated.	Factors to be investigated are suggested by the initial data of the given subjects (patients). Relevant factors are discovered throughout the research. Factor manipulation is limited or impossible.
Researcher's conducts	The experimenter's conducts are standardized.	The clinician's conducts vary, but general rules of conduct may be applied, especially rules relating to data collection.
Setting conditions	The setting conditions are standardized.	The setting conditions may vary within a general format.
Generalization	Only collective tendencies are important.	Idiography is the most important task, but regularities may be inferred from them.

Source: Elaborated by the author (2021).

It must have become pellucid that a matter of picking the most cogent method between those two must not be posed whatsoever. Despite the iterative recognition of the experimental method as the highest accomplishment of science, one cannot merely replace the clinical inquiry by the experiment. Some research questions, especially some psychological research questions, can only be put forward in a clinical context. We have cited mostly ethical reasons for this and now we should add a practical one. Although experimenters are usually averse to introspective data, an experimental interview can, in theory, access memories and fantasies, but the fast and fussy experimental encounter will likely foster social inhibitions and make this access defective. Only the long, careful and open relationship between the researcher and the subject can foster the confidence necessary to permit that both lay hold of inner life's truth.

To demonstrate this bold claim, let us take the arguments of the science communicator, psychologist and sexologist Jesse Bering in the article "Oedipus Complex 2.0: like it or not, parents shape their children's sexual preferences", published in *Scientific American* (Bering, 2010). The article presents a case of a "rubberphile" man from the same man's handwritten retrospective report; according to the sexual autobiography, the fetish started, to the best of the man's knowledge, "when, at the age of seven, he'd stumbled upon his mother's glistening white rubber bathing suit hanging on a clothesline on the back porch", an event that "coincided with his first becoming aware of that strange stirring on his loins". Bering (2010) laments that reports such as these are, "alas, little more than anecdotes", adding that there are nonetheless ways to study the relationship of child sexuality with adult sexuality in the laboratory.

He cites Thomas Fillion and Elliot Blass' classic experiment with rat-mothers and rat-pups: Fillion and Blass (1986) altered in three distinct ways the scent of three rat-mothers in their nursing phase to test the hypothesis that this would distinctly affect the sexual behaviour of the pups fed by them when the latter grew to a mating age. Compared to the other rats, the male rats once fed by a rat-mother with a lemon-scented vagina ejaculated faster when mating with an also lemony female rat and had problems to have an orgasm with female rats with a native scent. Bering (2010) underscores, though, that similar studies have not been done with humans and that "it's unlikely that we'll ever know for certain whether or not these data have any analogues with human sexuality", since

> it would be a real challenge to find mothers willing to tinker with their child's development in this shame-ridden domain. Turning one's son into a fetishist with an unhealthy attraction for reproductive organs that smell like Lemon Pledge may well be going above and beyond the call of scientific duty […].

Yet the psychologist comes back to his enthusiasm with the laboratory by citing experiments that demonstrate related mechanisms in humans without involving such a fetish brewing. In Fraley and Marks' (2010) experiments, students were asked to inform how much they were sexually attracted to images of strangers. In experiment number one, the students considered the strangers preceded by the subliminal image of their opposite-sex parent more attractive than strangers preceded by the subliminal image of another student's opposite-sex parent. In a second experiment, the strangers considered more attractive were the ones whose faces had been digitally combined with the students' own faces. In the last experiment, the participants to whom it was falsely informed that their faces had been combined in the images felt less sexually attracted to these images than participants to whom nothing was told.

Theoretical discussions in sexology and evolutionary psychology were hugely enriched by those experiments. Nevertheless, clinical investigations would be unparalleled complements to them. Only clinical data can help us to address some deep explanatory questions in sexology, psychopathology and the psychology of social life: what is the mental mechanism involved in a phenomenon from these provinces, what are the fantasies, wishes and mental schemes involved, why a result rather than any other, etc.. For instance, why did the sexual desire was lower when one was aware that it was a desire for a *Doppelgänger*? What are the unconscious mental schemas taking us to have the image of a parent as a turn-on? Which representations exist between an aroma and an orgasm? (If only animals similar to rats could talk…).

Bering (2010) himself acknowledges that experimental science must go beyond some of its limits to study complex and delicate objects such as childhood sexuality, for the latter, and especially its causal relationship with adult sexuality

> [...] is an elusive topic to study, at least in any rigorously controlled sense. It's also an area of research that a prudish society – or at least one that views an individual's sexuality as appearing out-of-the-blue with the first pubescent flush of hormones (or, alternatively, as unfolding in some highly innate, blueprinted sense impenetrable to experience, e.g., "the gay gene") – prefers to look away from, in spite of its centrality to the human experience. Unlike, say, studying children's acquisition of language, examining the precise developmental pathways to adult sexuality is more or less impossible. That's not because it's empirically impossible but rather only because childhood sexuality is one of those third-rail topics that gets zapped by the electric fencing of university ethics boards and is therefore at risk of always remaining little understood.

Clinical research, therefore, must (and does) offer much more than interesting narratives. It is not a matter, here, of picking the most cogent method – most of the scientific community picking the experimental – but of finding ways to refine both.

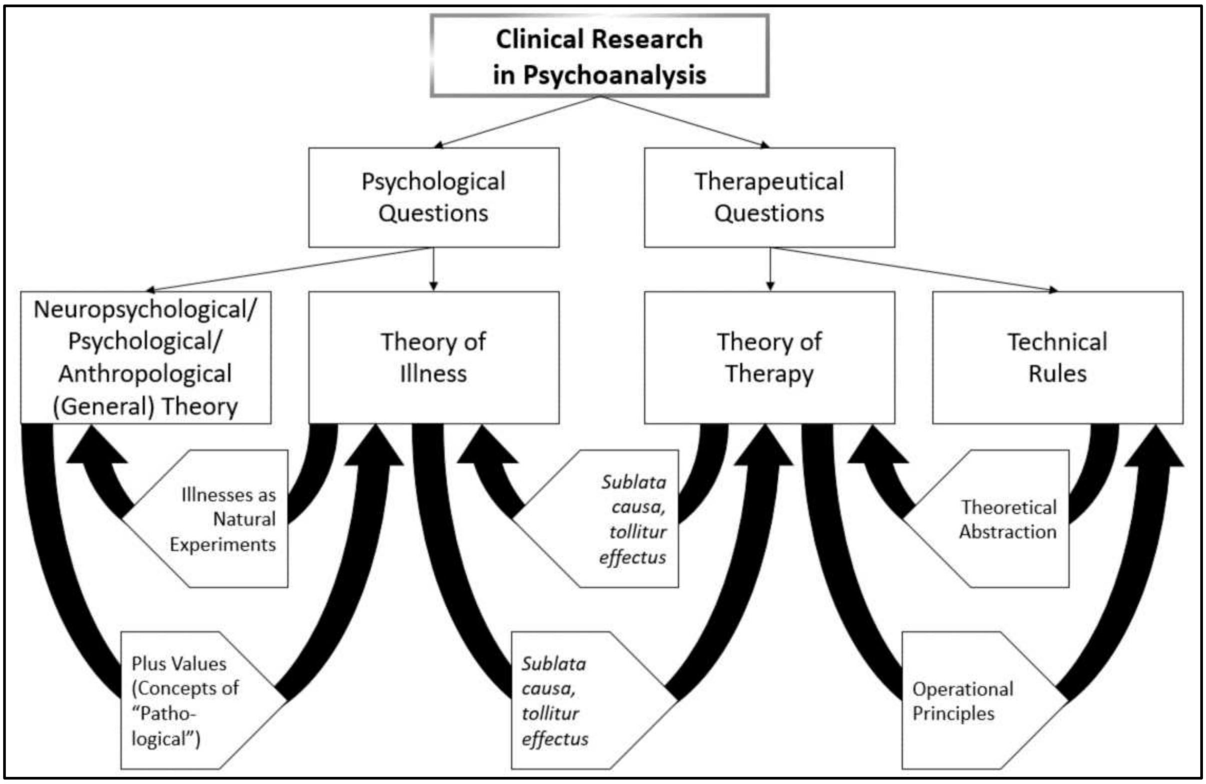
Psychoanalytic clinical research is interested in answering two kinds of questions, the psychological and the therapeutical; it is interested in answering how the mind works and develops and how psychotherapy works or could work better. The first kind has fostered and continues to foster Neuropsychological/Psychological/Anthropological (General) Theories and Theories of Illness, all labelled psychoanalytic. From the second kind it has followed and it continues to follow Therapeutical Theories and sets of technical rules.

Thus, what we call psychoanalytic theory can be broken up into three different kinds of theories. A General Theory comprises facts about how human minds in general work and develop, both in individual and collective levels, how such minds relate to the human brain, what is the ultimate nature of the human mind, etc.. A Theory of Illness (or Illness Theory) comprises both facts and values: values establishing which products of the human mind are undesirable and/or wrong, in other words, what constitutes a mental illness; and the causal and phenomenological facts about such illnesses. Finally, a Theory of Therapy (or Therapeutical Theory) is about which elements and processes would make an undesirable and/or wrong state-of-mind become a beneficial one.

The three theories are different from each other, but they are also connected to each other. An Illness Theory is a General Theory plus concepts of "pathological" and/or "vicious". One is entitled to transform the facts of an Illness Theory into the facts of a General Theory if one accepts the claim that illnesses are like natural experiments – that ill mental functions are amplifications or depletions of normal mental functions. Needless to say, this transformation is very familiar to psychoanalysts. One can transform an Illness Theory into a Therapeutical Theory and vice-versa through the simple logic of the Latin phrase "*sublata causa, tollitur effectus*" – "the effect ceases when the cause is removed". If x is the cause of the illness, the illness ceases when x is removed; if the illness ceases when x is removed, then x is the cause of the illness. A rough example: if repression is the cause of hysteria (Illness Theory), hysteria is cured when repression is undone (Therapeutical Theory). Finally, for one to transform a Therapeutical Theory into a set of technical rules, one must explicitly or implicitly make use of operational principles, of principles defining how to make therapeutic factors concrete (e.g., defining which conducts of an analyst could help patients undo their repressions and which conducts could not). A set of technical rules can also be transformed into a Therapeutical Theory through theoretical abstractions.

The carriage- or dress-shaped figure below summarizes the possible results of the clinical research in psychoanalysis and the conditions for one kind of result to be transformed into the other.

Figure 4 - The Results of the Clinical Research in Psychoanalysis and Their Interconnectedness



Source: Elaborated by the author (2021).

I finally have the elements to fully state the nature of this thesis: it is a philosophical research about the clinical-psychoanalytic research that aims at psychological, not therapeutical, knowledge.

Some universities currently undertake empirical research in psychoanalysis that epitomizes our definition of "clinical" and that, moreover, is not confined to therapeutical questions. For instance, the Department of Psychoanalysis and Clinical Consulting of Ghent University has in the last 6 years published the following studies: "A metasynthesis of published case studies through Lacan's L-schema: transference in perversion" (Willemsen et al, 2015), "Interactions between obsessional symptoms and interpersonal dynamics: an empirical single case study" (Cornelis et al, 2017), "Content matters, a qualitative analysis of verbal hallucinations" (Moernaut, Vanheule & Feyarts, 2018) and "Working through childhood trauma-related interpersonal patterns in psychodynamic treatment: an evidence-based case study" (Van Nieuwenhove, Truijens, Meganck, Cornelis & Desmet, 2019).

On the other side of *La Manche*, the Department of Psychosocial and Psychoanalytic Studies of the University of Essex also undertake clinical research in psychoanalysis. Two recent publications of one of the Department's staff can ratify the social and scientific importance of this kind of research: "'When it comes to HIV, that's when you find out the genuinity of that love': The experience of disclosing a HIV+ status to an intimate partner" (Smith, Cook & Rohleder, 2017) and "Navigating the relational psychic economy of disability: The case of M." (Watermeyer, Hunt, Swartz & Rohleder, 2019).

Scholars from these two universities, as well as three scholars from the University of Leuven, are developing the Single Case Archive, "a tool that allows the quick identification of specific sets of cases according to specific interests of researchers". It started in 2013 "with the release of a preliminary online archive containing about 500 psychoanalytic cases". The spokespersons of the project affirmed it came to life from a dissatisfaction with the "nomothetic paradigm", in which large groups of subjects are assessed in a standardized way: "the human psyche is far too lively and (re-)active to be captured by standardized tests" and, if psychology "doesn't proceed by studying subjects one by one, it doesn't seem to proceed at all" (SCA, 2021).

Although in Brazil academic psychoanalysis is still predominantly philosophical, it also presents some instances of clinical research. In the Federal University of Rio de Janeiro (UFRJ), Julio Verztman is currently coordinating a "psychoanalytic study of compulsions in patients diagnosed with OCD and related illnesses" using a method of "multiple case studies"; this method was formerly used by him and his colleagues to study stigma in patients with social phobia and the relation between melancholic patients and patients with autoimmune diseases (Verztman, 2009, 2021). More encompassing, there is also the Academic Association of Research in Fundamental Psychopathology ("*Associação Universitária de Pesquisa em Psicopatologia Fundamental*") discussing propositions enlightened by psychoanalytic case studies. Among the published books of the Association is a rich study on panic attacks, written by a scholar of the University of Campinas, Mario Pereira (Associação Universitária de Pesquisa em Psicopatologia Fundamental, 2020; Pereira, 2003).

4.3.2 The Axioms of Psychoanalysis

Freud had a project to erect an encompassing psychological theory from clinical investigation. He dubbed a theory of this sort a "*meta*psychology", for it should be an account of what happens *underneath* or *beyond* the prime object of 19th century psychology, consciousness. However, many factors have provoked a splitting of this project. A myriad of psychoanalytic spinoffs has appeared throughout the 20th century: "a whole array of differing theoretical positions, with contrapuntal, at times passionate, defenses by its continuing adherents" (Wallerstein, 2006, p. 304). In a Foreword to Hinshelwood's "Research on the couch", the then president of the IPA, Charles Hanly (2013), affirmed that one of the two great problems in contemporary psychoanalysis is this "co-existence of diverse theories, at least some of which appear to be inconsistent with each other, without any reliable means of determining which is the most probable", a problem worsened by the formation of schools that prefer controversy to enquiry (p. xi). In a recent work on metapsychology as a foundation for psychoanalysis, Boag (2017) conveys the problem of pluralism with the help of some scholars:

> Rather than stimulating growth, pluralism, for many, has meant that psychoanalysis has disintegrated into a confusing plethora of conflicting approaches. There may be anywhere between twelve to twenty psychoanalytic schools (Frank, 2000; Novick & Novick, 2002), and rather than benefitting from constructive dialogue, each school generally remains insulated and prevented from development that would otherwise occur in a more interactive and rigorous discipline (Frank, 2000; Green, 2005; Rangell, 1988). […] Fonagy and Target (2000) go even further, viewing the "increasing fragmentation of theory" as potentially fatal for the future of psychoanalysis: "This fragmentation, euphemistically discussed as pluralism, could on its own spell the demise of psychoanalysis" (p. 408). (p. 1)

But the theoretical divergences in this Great Metapsychology Debate (Wallerstein, 2006) may distract us from more radical divergences involving the possibility, origin and essence of psychoanalytic knowledge. Positions on these three topics are brutally philosophical and could never be pondered the same way a theoretical proposition could; they are not mere divergences on what the data are implying or on which conceptual constellations are the most useful to organize the phenomena in question.

Psychoanalysts may verge into differing *metaphysics* for their notions of truth and reality (in their case, of *mental* reality). We may roughly say that, in the Freudian legacy, there is a confusion of tongues between anti-realists, usually also relativists, and all the others, usually critical realists. For relativists, the notion of truth is somehow harmed by the complexity of reality: if ultimate truths on our messy reality cannot be stated, then there are "local truths", the "correspondence to reality" turning into "what is held to be truth in a cultural or personal context". They insist that "there can be no framework-independent vantage point from which the matter of whether the thing in question is so can be established" (Baghramian & Carter, 2020). For anti-realists, there is no such thing as a reality outside of the subjectivity of the knower or, if there is, either it does not matter – since it cannot constrain our spirit – or it is impossible to grasp in some or in any way. Spence (1982) is a nice specimen of this post-modern combo, as the following passages suggest:

> […] the associations of the patient have no one-to-one correspondence with his memories and dreams, much less with his unconscious thought. […] the complex visual scene represented by a dream or an early memory can probably never be completely realized by language. (p. 28).
> Interpretations are persuasive […] not because of their evidential value but because of their rhetorical appeal; conviction emerges because the fit is good, not because we have necessarily made contact with the past (p. 32).
> […] an interpretation can be seen as a kind of artistic product, and as such, it becomes possible to consider its effects on the patient as a kind of aesthetic experience. What might be called its "beauty" need have no necessary relation to its "truth". (p. 37)

Psychoanalytic work does not have truth as a horizon, Spence supposes, because its therapeutic or emancipatory dimension is what matters most (see also Habermas, 1968, and Renik, 1998). A slight variation of this kind of position is an analyst that, from the historical fact that the idea of truth can serve the established powers, especially in the sphere of sexuality and mental health, concludes that there are (or there should be) sundry and limited psychological "truths"; Foucault and Butler would be influences here. Lacanian analysts, in their turn, although having the truth of a patient as horizon, hold that this truth is inherently impossible to be reached and articulated. Lastly, there are the Winnicottians, strongly influenced by phenomenology, considering that a substantial psyche like the one postulated by Freud far outweighs our immediate apprehension of the phenomena known as mental; in this line, a metapsychology would be only a tool to help us make sense of clinical data, disposable afterwards (see Fulgencio, 2003, 2013).

Opposingly, Freudians who are critical realists presuppose that the human psyche exists independently of the whims and limits of the ones who decide to study it *and* that the fact that this real psyche is eerily complex does not prevent them to get glimpses of it through theory. Surely, they think, these glimpses must always be questioned, for the whims and limits of the human psyche of scientists are also real enough – hence the "critical" in "critical realist". Simanke (2009) takes this stance:

> […] the reality of the mental can be (or must be) assumed independently of the fact that each one of the explanations proposed by psychoanalysis be true or not, which, evidently, can only be decided through the investigative practice itself, and not through an epistemological position-taking that precedes and conditions this practice. That is, to assume realism does not mean to accept that the mental be like psychoanalysis affirms it is, but only to admit that there be something that is mental, to which psychoanalysis refers in a more or less adequate manner, this reference having to be perfected by the development of the knowledge produced by psychoanalysis (pp. 117-118).

For the author, we must attribute reality to the entities, mechanisms and processes hypothesized by metapsychology "so that the causal role they play in the psychological explanation can be justified" (p. 118).

Bell (2009) inspects the attraction some psychoanalysts have for relativism and also takes a critical realist stance:

> That psychoanalysis has as its primary concern inner experience, which is complex and many-sided, may make it appear to belong within such a postmodernist framework. This is, I think, profoundly mistaken. Firstly, there is no necessary connection between complexity and relativization of truth. Secondly, psychoanalytic explanation, whilst asserting the enormity of the factor of subjectivity in determining any human being's understanding of his world, at the same time maintains a commitment to objectivity. In fact, this tension between the importance of subjectivity and the struggling for objectivity is an essential tension of our subject (p. 335).
> Where there is no conception of truth, its various counterparts, namely deceptions, lies, and misrepresentations also lose their foothold. What a depth psychology or depth social explanation might look like without such concepts is very hard to understand. (p. 332)

However, his deep quarrel did not customarily surface in history; its implications did. Splitting the Freudian legacy into a *bad* psychological theory – a speculative rendering that reduced the human soul to physics and biology – and a *good* clinical method – closer to phenomena and plastic enough to appreciate human singularity and autonomy – is an old way of saying that the legacy should be anti-realist (see, for instance, Politzer, 1928, and Habermas, 1968). Further, it is an old way of saying that the legacy is also, naturally, anti-naturalist. The stance that humans are radically apart from the rest of nature, and that psychoanalysis should be decanted from its biological sediments, have wedded well with the anti-realist love for the clinical method (Simanke, 2002, 2009). Indeed, everything hypothesized as *real* in the original *metapsychology* has a neurological character (Simanke & Caropreso, 2011). Nowadays, the deep quarrel appears deceptively as a mere quarrel of method choice: an experimental psychoanalysis, sympathetic to cerebrums and statistics, opposing a clinical, hermeneutic, idiographic one (Sandler, Sandler & Davies, 2000; Luyten et al, 2006; Wallerstein, 2009; Desmet, 2013).

The line-ups of the last paragraph seem, though, perfectly contingent. A metapsychology seems to guide investigation in the clinic (Mackay, 1989; Boag, 2017), and the clinic, in conjunction with experiments and ethnographies, seem to be able to enrich a metapsychology by generating, supporting and refuting hypotheses (Hinshelwood, 2013). A clinic can be naturalist, if only meaning, singularity and autonomy are part of nature (see Mackay, 1989, and Simanke, 2013), and a metapsychology can also be anti-naturalist, as shown by Lacan (Simanke, 2002). This research presupposes that the clinical method can be as stringent as the experimental and the experimental method, as limited as the clinical. But, all things considered, it seems impossible to dissolve the psychoanalytic schools' disagreements concerning reality and truth; that is where their dialogues are truly hindered.

Considering the plethora of "metapsychologies" and the radical divergences as regards the possibility, origin and essence of knowledge in psychoanalysis, it becomes urgent first to state what I shall understand as psychoanalysis in the chapters to come. Let me thus state the set of ideas I take to be most basic in psychoanalysis. This set is composed of three theoretical hypotheses (T) and two clinical-methodological rules (CM):

T1- There is a dynamic unconscious: some of our conscious mental products can only be fully explained by postulating unconscious motives.

T2- This implies conflict within mental life: some motives were always or became unconscious because, if they were made conscious, they would conflict with the subject's self-conservation, self-image and/or sociability; they can only become conscious, thus, if transformed.

T3- This dynamic unconscious is dominated by a primary process: unconscious motives are born through archaic representation-binding laws. In primary process, representations are bound only according to their sensory similarity and experiential contiguity.

CM1- For the analyst, the rules of neutrality, abstinence and free-floating attention: the analytic relationship must be controlled by the analyst in order to prevent epistemic disturbances and biases.

CM2- For the patient, the rule of free-association: the rules above enable the patients to tell the analyst everything that comes to their consciousness after the analyst has invited them to do so.

A dynamic unconscious, a conflictual psyche, a primary process, the purpose of epistemic control, the promotion of free-association. In my view, these hypotheses and rules would work approximately as axioms in psychoanalysis. Every other of its eminent hypotheses – the Oedipus Complex, the Castration Complex, the psychosexual phases in development, the transference, the many kinds of defences, the pathological effects of repressions, etc. – and methodological rules – the rule to analyse transference and resistance, for instance – could be made a consequence of them (with the right extra premises).

I believe that if one contradicts or changes one of such "axioms" one is no longer talking of psychoanalysis (of course, CM1 and CM2 are crucial only for a *clinical* psychoanalysis).

4.4 SCIENCE UNIFIED

Let me now present what I shall consider the indisputable definition of "good science" at which clinical psychoanalysis must aim. I endorse Gerhard Schurtz's (2014) "supreme epistemic goal of science" and his "minimal epistemological model" that explicates this goal. The supreme epistemic goal of science (G) is "to find true and content-rich statements, laws, or theories, relating to a given domain of phenomena" (p. 19). The more consequences a statement has, the richer it is. In order to clarify what a content-rich *truth* means, Schurtz (2014) presents the five basic epistemological assumptions (E1-E5) of scientific practice that would lead to it:

E1- A Minimal Realism: There is a reality independent of the epistemological subject. The possibility of knowing this reality in its entirety "is left open and cannot be answered a priori, but only based on the factual epistemic success of science" (p. 22). A sentence that is true has a *structural* agreement with part of this reality. This notion of structural correspondence "does not imply any intrinsic relationship or mirroring between language and reality. It also allows that theories which are strictly speaking false may be true in an approximative sense, or close to the truth" (pp. 22-23).

E2- Fallibilism and Critical Attitude: There is no infallible operation to reach true statements. We can never be entirely sure of the truth of a scientific statement, "but we can consider their truth to be more or less probable" (p. 23).

E3- Objectivity and Intersubjectivity: It follows from E1 that a statement's truth is independent of the beliefs and values of the epistemological subject; in other words, that it is objective. But, as only subjects make statements, the only way to approach objective truth is via intersubjectivity. The agreement of many mutually independent and competent researchers on the hypothesis enhances the chances of it being true. So, a statement is scientific if and only if it is possible, in principle, to convince any person of its truth.

E4- A Minimal Empiricism: If an object is scientific, it is in principle accessible to observation and experience. Observation and experience are the primary arbiters in the search for truth, for, without perception, reality cannot be reached at all. It is a minimal empiricism because

> it is not claimed that all scientific statements must be traceable back to observations via a chain of definitions, or must be provable by observations. Scientific theories can and should contain theoretical concepts, by which they can talk about things that lie beyond what is observable to human senses. The decisive requirement of E4 is merely that statements about the unobservable must have observable consequences by which they may be tested. (p. 24)

E5- Logic in the Wider Sense: It is only when we know the consequences of a hypothesis that we can test it empirically, and the consequences of a hypothesis can only be known by using logical systems. In this assumption pinpointed by Schurz (2014), I would further specify that scientists may not be confined to systems of *classical* logic; for some, a *non-trivial* logic is enough.

This minimal epistemological model "makes as few assumptions as possible, but as many as necessary, in order to explain the possibility of objective science" (Schurz, 2014, p. 24).

In the last section, I have concluded that part of the dialogues inside psychoanalysis is *irremediably* hindered because what is "reality" and "truth" for some of them is not commensurate with what is "reality" and "truth" for the rest of them. Hence, solutions to the differences between relativists and critical realists shall *not* be discussed in this thesis. Indeed, by endorsing G and E1-E5, I am immediately assuming a critical realist stance for psychoanalysis. But why a critical realist and not a relativist stance? One, because there is an inherent contradiction in relativism: to affirm "there are many truths" is to affirm one truth. Two, because I take dialogue – or intersubjectivity, or yet the possibility of public recognition of a statement – as a high moral value. Not to have it as a value is to welcome the risk of solipsism and social futility.

We can thereby testify that, in a work of philosophy of science concerned with logical questions, it is impossible to evade metaphysics altogether. (The recognition of this fact, by the way, was partly the cause of the decay of Logical Empiricism.) Following Schurz (2014), I reiterate, though, that assumptions G and E1-E5 are far from rigid and fussy in their metaphysical demands. They make space for a range of metaphysical stances. The idea of a mostly logical work in philosophy of science, which shall be instantiated by this thesis, can be held after all.

4.5 WHAT KIND OF SCIENTIFIC KNOWLEDGE COULD CLINICAL PSYCHOANALYSIS PRODUCE?

4.5.1 What Would Clinical Psychoanalysis Distinctively Investigate?

Before we start to tackle the problems Grünbaum has delineated and inspected with great wit, we must still state with the most clearness why it is worthwhile to discuss and advance the epistemology of the clinical research in psychoanalysis – otherwise all of our efforts could be revealed as vain. From all we have seen about the specifics of the clinical method, is the knowledge this method is able to produce precious as regards some of our strongest social needs? What exactly would we be able to obtain of epistemic capital from the application of the clinical method through psychoanalytic theory? What kind of results would this research program be able to produce?

For this, we must target a concrete set of projects of clinical research in psychoanalysis, be they actual (as those cited on pp. 81-82) or potential (as those we can devise by considering urgent social and theoretical problems). Keeping this set in mind, we must answer three questions. Would the clinical research in psychoanalysis have privilege over a specific domain? In other words, would there be kinds of phenomena that it would have more resources to investigate than other research programs? Second, which kinds of hypotheses would the research program have to conceptualize in order to integrate the "clinical concrete" to the "extraclinical abstract and/or remote", as well as the idiographic nature of clinical results to the nomothetic nature of scientific results? Third and last, what do we talk about when we talk about causation in deep psychology?

Let us consider the following list of topics that have been or could be investigated by the clinical method of psychoanalysis: patterns of therapeutic relationship (transference) between perverse patients and their analysts; interpersonal dynamics in obsessive patients; recurrent themes in the verbal hallucinations of schizophrenics living in Western Europe; common fantasies surrounding the disclosure of a HIV+ status to an intimate partner; childhood sexual memories in rubberphiles; the causal relationship between the COVID outbreak and a higher frequency of dreams with paranoid themes; the causal relationship between an unconscious need in males to "take revenge" upon a rejecting mother and their murderous fantasies and actions; gender-related conflicts in bisexual cis people; common fantasies active in the minds of refugee children with Resignation Syndrome.

Which dimensions of human life there appear in this list of research projects? The list displays, basically, topics related to the dimensions of *personality* (which may include the dimension of *identity*), of *sociability* (that is, the dimension comprising phenomena of aggression, tolerance, power, care, etc.) and of *sexuality*. The Freudian cannon teaches us that this triad of human dimensions is interrelated by nature, like branches with a same root; nevertheless, if one wishes to keep a dialogue with common-sense, it is useful to recognize them as distinct from each other. On the same line, it is exclusively from a commonsensical perspective that we may affirm that *psychopathology* is a central topic for clinical psychoanalysis. The list seems to show topics absorbed by the pathological state of such dimensions, but we have seen that, according to the Freudian paradigm, the pathological mind is just a hyperbole of some aspect of the normal mind. At any rate, would this triad of dimensions, both in its pathological and its normal states, be the *privileged domain* of clinical psychoanalysis *qua* research program?

The answer is no. A clinical behaviourist program, for example, could well study personality, sociability and sexuality (and, of course, psychopathology)[32]. What is crucial here is not the dimensions themselves but the peculiar way clinical psychoanalysis investigates them. Proceeding with the example, the "pure-blood" clinical behaviourist would never investigate, like the clinical analyst, mental representations, and "less than never" unconscious ones; following a specific anthropology and epistemology, the behaviourist study of motivation is rooted in the identification of the environmental conditions the organism, with its innate dispositions and its history of environmental conditioning, finds itself in. Let us devise, thus, a candidate to be the privileged domain of clinical psychoanalysis which is better than the mentioned triad: clinical psychoanalysis would be the research program studying the form, content and origin of *unconscious motives*[33] in any human dimension; its method, in any case, would be specially fit for the study of unconscious motives.

Clinical psychoanalysis began with the interest in a specific kind of mental phenomena: mental phenomena that are unfamiliar and astonishing, and, at the same time, that show to have little or not fully to do with congenital factors or with immediate environmental constraints and pressures. If one's mental phenomenon does not make sense, it must be explained by something we are not conscious of. If this something cannot be exclusively ascribed to one's genes, to one's gestation or to one's immediate situation, then it must be a structure of ideas and affects that surreptitiously have been putting one's mind in motion – an unconscious motive. The method of clinical psychoanalysis was developed to study such surreptitious structures.

Clinical analysts are interested in mental phenomena that (apparently) do not make sense, in actions that (apparently) go wrong, in mind-legs (and some actual legs) that stumble, deviate or paralyse while walking; in unusual, rare or surprising behaviour. An analyst, as much as any scientist, wishes to explain such phenomena precisely because the latter do not make sense according to the state of the art, including the art of common-sense. Here we concur with Rubinstein (1975), who observes that

---

[32]By the way, an anecdotal comment: a professor of Behaviourism I had, worried about the low popularity the behaviourist program held among psychology students, was frequently bothered by the fact that the attractive issues of psychopathology and the psychology of art were immediately associated to the Freudian tradition.

[33]To work with another contrasting example: clinical cognitivism may also study psychological motives, but not so much the unconscious ones and not so much the remote origin of such motives. However, the differences between the programs could well be considered mild, if we take clinical psychoanalysis to be interested as much as clinical cognitivism in mental schemas, as we shall defend in a moment.

> Obviously we do not try to explain everything we see, hear about, etc. Only about certain things do we feel that they have to be explained. According to Hanson (1958, p. 86; 1971, pp. 39f.) and Toulmin (1961, pp. 44ff.), among others, in science generally, unless for one reason or another we are indifferent to it, we feel the need for an explanation if we come up against an unexpected and in this sense puzzling event. As Sherwood (1969, pp. 10ff.) has pointed out in an excellent discussion of this question, that is true for psychoanalysis also. In this case unexpected events are behaviours that deviate from what is typical of a person *as such*, and/or as a member of the human species, and/or of the various groups and subgroups of this species to which he belongs. Evidently we must know what is typical, i.e., expected, or else we could not possibly classify something as unexpected. Sherwood (p. 18) speaks here about "a presumption of knowledge" (p. 7).

In his book-reply to Grünbaum, Edelson (1984) expresses a similar point:

> In a classical Newtonian particle system, a physical object remains in a state of constant motion unless a force intervenes to change its motion. It is change in its motion that requires explanation. Analogously, psychological entities or the psychological structures formed of them make sense unless some "force" (a causally efficacious factor) interferes. It is that (or the extent to which) a psychological entity does not make sense that requires psychoanalytic explanation (p. 94).

According to the author, the decision over whether a psychological entity or structure is senseless or senseful does not depend on any knowledge of psychoanalytic theory; the notion of psychological senselessness is, thus, pre-theoretic or non-theoretic in relation to such theory. A psychological entity or structure does not make sense "if it has no apparent relevance with respect to achieving any wished-for state of affairs, and no apparent relation to other psychological entities, and a subject cannot prevent himself from producing it" (Edelson, 1984, p. 95). Psychoanalysis would be distinctively implicated in the explanation of such senseless entities or structures as long as no physical defect or trauma, no biological fault, no innate psychological deficiency and "no specific feature of some external environment or specific information or lack of information in that environment is either a necessary[34] or sufficient condition for, or cause of, the fact that the psychological object or structure [in question] does not make sense" (p. 96). In his attempt to spell out the domain of psychoanalysis, Edelson (1984) claims that his adumbrated determination of the theoretical predicate "… is a Freudian system"

> suggests a view of psychoanalysis as a science of the imagination. In this view, psychoanalysis is not primarily a general psychology, encompassing all human capacities or "behaviour" (Edelson, 1977). Psychoanalysis studies action insofar as action flows from, or is determined by, imagination. Its primary interest is not in accounting for rational action, but in accounting for psychological entities or structures that do not make sense, by tracing the vicissitudes of

[34]We shall see in a moment that we can still talk about unconscious motives even when non-psychological factors are necessary but not sufficient for the occurrence of a behaviour – so, to be honest, we would fully agree with Edelson's formulation only if the three words "…either […] necessary or…" were taken away of the cited excerpt.

> instinctual wishes in successful, unsuccessful, interrupted, aborted, or distorted wish-fulfilment (Edelson, 1984, p. 106).

Later on in the same chapter, he observes that philosophical discussants of psychoanalysis frequently ignore that the latter is all about "the ubiquity and the power of the fantastical" (p. 107).

We have concluded that the dimensions of personality, sociability, sexuality and psychopathology are not private assets of clinical psychoanalysis, and that, if the latter does have a privileged domain, it is the domain of unconscious motives. Nevertheless, such domain is easily found in the four dimensions, which for this reason should at least be called clinical psychoanalysis' *sphere of investigation*.

If a specific social group wants to know about the unconscious motivation for pathological, sexual, social and characterial behaviour, about its form, content, and origin, if to know about it is considered useful, good, just, etc., by the group – only then must the research program of clinical psychoanalysis be supported and developed.

4.5.2 What Would It Mean to Entertain a Hypothesis in Clinical Psychoanalysis?

We find a recent demarcation inside the multitude of hypotheses we label psychoanalytic in Michael Lacewing's paper "The Science of Psychoanalysis" (Lacewing, 2018). According to the philosopher, the psychoanalytic theory is actually a group of three: a clinical theory, a metapsychological theory and an aetiological theory.

The clinical theory would be the corner of psychoanalytic theory with postulations of "typical structures of motivation, their typical effects, and their manifestations in the consulting room, for example, the existence and nature of defence mechanisms, such as repression, projection, and so on, and their clinical manifestations in conflict, compromise, resistance, and transference" (pp. 96-97), also including "an account of the causal role of these mental structures in the manifestations of mental illness and character traits, and a theory of how therapy works" (p. 97).

It should be kept apart from the metapsychological corner, which contains manifold and avowedly rival accounts "of the fundamental structure or organization of the mind, […] how such structures function, and their relation to mental illness, character, and mental health" (p. 97) – for example, Freud's id, ego and superego and Klein's depressive and schizo-paranoid positions –, as well as from the aetiological corner, where we find hypotheses over "the causal origins and typical development of the structures falling under clinical and metapsychological theory, especially in relation to childhood experience" (p. 97).

To justify this partition, Lacewing (2018) argues that, while clinical hypotheses can be inferred from clinical data alone, *many* metapsychological and aetiological hypotheses cannot. Metapsychological and aetiological hypotheses would stray far from the actuality of the clinical exchange, a shred of proper evidence for them having thus to come from extraclinical research. Only when the clinical theory is unburdened of the vestiges of the other two theories – and the philosopher grants this can be largely done, save with some of its aspects related to clinical technique –, the import of clinical evidence to support it can be regarded as scientifically rigorous (Lacewing, 2018).

I shall not adopt this manner of organizing and predicating the psychoanalytic warehouse of hypotheses, for two reasons. I believe, first, that it is impossible to separate what Lacewing (2018) puts under the scope of clinical theory from what he defines as a metapsychological theory. (However, I consent that extraclinical research serves the latter theory well; coherently, I do not take a use of extraclinical results in the inference of the clinical theory as making it any less clinical). The second reason is that a central aim of the monograph you are reading is to demonstrate that it is both relevant and possible to test aetiological hypotheses – not exclusively but at least mainly – through clinical data.

It is in the approaching Chapter 4 that I shall attempt to demonstrate both that it is possible to strongly test aetiological hypotheses from the couch and that it is impossible to refrain from metapsychological hypotheses in the clinical context. In fact, I have commented on the last point earlier in this chapter and now I am about to do it once more by bringing Benjamin Rubinstein's argument that "clinical theory becomes more interesting and also more satisfactory […] when properly integrated with an acceptable extraclinical theory" (Rubinstein, 1980, p. 427). According to the Finish-American analyst, analysts tacitly assume that the rules used in clinical inference reflect the independent existence of certain mental events because, were it otherwise, they would have to regard such rules as just one of many possible ways to organize clinical data. Analysts would assume the existence of such – unconscious, purposive, etc. – events in their interpretations, "neural events by another name" in his anti-dualistic account and, thus, we would have to admit that "the postulates of the clinical theory either are at the same time *hypotheses* of […] [an] extraclinical[, basically neurophysiological,] theory or are readily transformed into such hypotheses" (p. 439-440).

If clinical hypotheses are just expressions of extraclinical hypotheses about the functioning of an organically-rooted mind – that is, of metapsychological hypotheses –, then it is not fruitful to oppose the latter to the former. I agree with Rubinstein (1975) that the significant distinction in psychoanalysis, and perhaps in all human sciences, "is not between why and how things happen, nor between motives and causes, nor between explanation in terms of meaning and causal explanation, but between explanations involving a *person* and explanations involving the *human organism* this person is also" (p. 5). A more useful contrast to guide the classification of hypotheses in psychoanalysis is, thus, the contrast between the *general* and the *particular*, not, like Lacewing seems to defend, between the "clinical concrete" and the "extraclinical abstract and/or remote". (From this perspective, Lacewing's classes would each be crossed by both general and particular hypotheses.)

Inspired by Rubinstein (1975), I defend that the hypotheses entertained in clinical psychoanalysis are of the following three kinds: *core* and *categoric* hypotheses, which are general hypotheses with differing scopes, and *particular* hypotheses, which are valid for only one person. Each kind has its special regime of justification, but all of such regimes are based both on clinical data and on extraclinical hypotheses, as shall be argued in the following chapter.

This is one of the standpoints from which the notions of core, categoric and particular hypotheses could be defined in the context of clinical psychoanalysis: the *core* hypotheses of psychoanalysis are supposed to be *valid for all imaginable human minds*; the *categoric* hypotheses are supposed to be *valid for the minds of a group of people sharing a feature* (that is, of a category); and the *particular* hypotheses are supposed to be *valid for only one person* (a patient, in most cases). This is the standpoint of the *scope* that would stand under each kind of hypothesis. For example, "the nature of the attachment to caregivers during childhood determines the nature of attachment and sexuality in adulthood" is a core hypothesis, "paranoiac symptoms are defences against homosexual impulses" is a categoric hypothesis and "the childhood event where the Rat Man bites his nurse and gets beaten by his father is one of the causes of his obsessive symptoms in adulthood" is a particular hypothesis.

I have mentioned above that the inspiration for such classification comes from Rubinstein (1975). When he proposes his notions of *general*, *special* and *particular* clinical hypotheses, he seems to define them from this "scopic" standpoint. He defines particular hypotheses as the explanations of particular psychological phenomena coming from the interaction of more or less general and/or abstract statements – the general and special hypotheses, the psychoanalytic theory *per se* – with certain clinical data. Particular hypotheses would be the ones presented to patients during treatment and would be "generally referred to as interpretations" (p. 4). The psychoanalytic theory guiding the formulation of such interpretations would be composed of three *general* clinical hypotheses – a "motivational"[35], a "situational"[36] and a "genetic"[37], each one with many sub-hypotheses –, as well as of many *special* clinical hypotheses, some of which would "represent more or less extensive generalizations of particular clinical hypotheses" and some of which would "have been taken over directly from common-sense psychology" (p. 27), for example, hypotheses over "the common meanings of particular dream symbols" and "the dynamics of [...] hysteria" (p. 28).

Perhaps the reader has realized that I have changed the term "general" to the term "core" and the term "special" to the term "categoric". I do that not only because I consider the alternative names more informative of what they refer to – after all, Rubinstein's' "special" hypotheses are as general as his "general" ones. I also think that, by following the connotations of such alternative names, we are better able to extract the consequences of Rubinstein's notions and of the framework from which they came.

The use of the expression "core" was borrowed from Lakatos' work (see p. 67), where "hard core" refers, as "core" does here, to the most central or basic hypotheses of a research program. Nonetheless, in Lakatos' rhetoric the sense of "core" is more logical and historical than scopic. The hard core of a research program would be for him its hypotheses that have endured for contingent reasons throughout its historical variations. The hard core would also be the set of hypotheses that scientists have systematically attempted to save despite instances apparently refuting them (through an inquisition of the auxiliary hypotheses involved in such refuting instances), and that *for a good reason*: the set of hypotheses have presented a growing power to predict with exactitude novel and unexpected facts.

At least the difference between core hypotheses of psychoanalysis on one side and categoric and particular hypotheses of psychoanalysis on the other can be understood also from historical and logical standpoints. Historically, hypotheses such as "Dora was in love with Mr. K", a particular one, and "female sexuality is determined by envy of male sexuality", a categoric one, are controversial, and have been contested by many; the idea that "wishes [can partly] cause dreams", on the other hand, has remained credible throughout the generations. In a logical domain, too, core hypotheses, such as the latter and "the nature of the attachment to caregivers during childhood determines the nature of attachment and sexuality in adulthood", have resisted

---

[35]The hypothesis "that all activities in which a person engages, whatever their specific nature, are motivated even if on the face of it they may seem not to be" (Rubinstein, 1975, pp. 10-11).
[36]The hypothesis "that within limits people respond to external situations in more or less specific ways" (Rubinstein, 1975, p. 11).
[37]The hypothesis "that the presence in a person of certain motives and response dispositions may be explained, at least in part, by events early in that person's life" (Rubinstein, 1975, p. 11).

refuting evidence for having demonstrably been, not exactly predictive in Lakatos' sense, but at least explanatorily and practically valuable, while many categoric and particular hypotheses have not, for the opposite reason.

All in all, there are three standpoints from which the notions of core, categoric and particular hypotheses could be understood. From a scopic standpoint, core hypotheses are the ones valid for every human mind in the world, categoric hypotheses, the ones valid for the minds of a segment of humankind, and particular hypotheses, the ones valid for just one person. From a historical standpoint, core hypotheses are the ones that have been kept by psychoanalysis' devotees despite a long unfolding of empirical and non-empirical moves around the science, representing thus its *identity*, while categoric and particular hypotheses are the ones that as a rule have been historically questioned. Finally, from a logical standpoint, core hypotheses are the ones without which the clinical research in psychoanalysis would lose its explanatory potency, while categoric and particular ones can be discarded without harming this potency. Needless to say, these three senses of core, categoric and particular not only are consistent with each other; they also complement each other well.

To answer the heading of this section – what does it mean to entertain a hypothesis in clinical psychoanalysis – we should say that it means to entertain those three kinds of hypotheses. To do clinical psychoanalysis is mainly to interact with patients in a certain way and to infer particular hypotheses with the aid of core hypotheses (we shall see in the next chapter why it is not a good idea to be aided by categoric hypotheses here). But, by the moment one wishes to use this practice as a research device, it is because one is interested in knowing more about a group of minds with something in common, because one is interested in inferring *categoric hypotheses*.

Although a clinical research in psychoanalysis may generate and support the three kinds of hypotheses, its social and scientific impact is as great as its ability to generate and support categoric hypotheses. Such are its final aims, because a particular hypothesis taken in isolation, as it refers to just one person, cannot help to explain or treat other persons; and because core hypotheses are more richly, easily, and reliably supported by extraclinical research. Again, arguments over how the three relate to each other and how each one of them is scientifically inferable shall be presented in the next chapter.

4.5.3 What Would It Mean to Entertain a Cause in Clinical Psychoanalysis?

Clinical researchers in psychoanalysis wish to find the causes of symptoms, dreams, parapraxes, and the like, and their theory tells them such causes are psychological. But what is a psychological cause? What does a hypothesis that a symptom (or dream, parapraxis, etc.) is psychogenic imply? Would clinical researchers in psychoanalysis be too bold in holding that a behaviour is caused by the unconscious? What about the environmental and biological factors related to behaviour? What about the role of agency and free will in its determination? In sum, what kind of causal knowledge could clinical psychoanalysis produce? Such are fair questions, and tackling them could help us realize what exactly the explanatory ambitions of our clinical researchers are.

Perhaps the rawest idea non-academic persons get from psychoanalysis is the idea that certain events in someone's childhood cause this someone to become mentally problematic in adulthood. We know, though, that one of the first great scientific insights Freud had was that mental suffering comes not purely from what hit us or from what people did to us, but from our fantasies around such events. (Independently of Freud, it is nearly impossible to find a scientist defending that our minds are like a billiard table or a carbon paper – it is clear that we compute the world exhaustively before we do anything in and with it.) But how would a fantasy around a childhood interaction be able to influence adult behaviour? Time, space, people and demands have changed – why would this fantasy remain significant?

This fantasy would remain significant because our minds would drain some of its specific and concrete features, making it suitable to process inputs and deliver outputs not exactly identical to the original ones. Our minds would generalize and abstract this fantasy in such a way that it would become what psychoanalytic theory calls a *complex*, something similar to what the cognitive sciences call a *memory schema*.

According to Laplanche and Pontalis's "Vocabulary of Psychoanalysis", a complex is an

> organized group of ideas and memories of great affective force which are either partly of totally unconscious. Complexes are constituted on the basis of the interpersonal relationships of childhood history; they may serve to structure all levels of the psyche: emotions, attitude, adapted behaviour. (Laplanche & Pontalis, 1988, p. 72)

As to the notion of memory schema, there is more than one definition; all of them, anyhow, are comparable to this definition of "complex". The notion was introduced by Frederic Bartlett in 1932, but it was considered vague and was shunned by mainstream experimental psychologists until computer sciences revived it in the mid-1970s (Rumelhart, Smolensky, McClelland & Hinton, 1986 as cited in Wagman, 2000; Bucci, 2021). According to Bartlett (1932, p. 201, as cited in Wagman, 2000, p. 197), a memory schema is an "active organisation of past reactions, or of past experiences […] operating in any well-adapted organic response"; this kind of response is possible "because it is related to other similar responses which have been serially organised, yet which operate […] as a unitary mass". According to the author, this is "[…] the most fundamental of all the ways in which we can be influenced by reactions and experiences which occurred some time in the past" (Bartlett, 1932, p. 201, as cited in Wagman, 2000, p. 197). In 1980, Rumelhart (1980, p. 34, as cited in Wagman, 2000, pp. 146-147) defined it as a "data structure for representing the generic concepts stored in memory"; this structure consists in the "network of interrelations […] among the constituents of the concept in question", which could comprise "objects, situations, events, sequences of events, actions and sequences of actions". A schema makes us go beyond our observations: "once we have determined that a particular schema accounts for some event, we may not be able to determine which aspects of our beliefs are based on direct sensory information and which are merely consequences of our interpretation" (Rumelhart, 1980, p. 38, as cited in Wagman, 2000, p. 198).

Wilma Bucci's notion of "emotion schema" is even closer to the meaning of "complex". An emotion schema would be a *kind of memory schema* "formed through the repeated occurrence of a set of subsymbolic sensory, somatic, and motoric processes in relation to certain events of life" (Bucci, 2021, p. xxxv); she termed this set the "affective core" of the schema. Differently from other kinds of memory schema, interpersonal interactions would have a central role in its formation. The researcher illustrates her notion thus:

> The child feels a conglomerate of bodily experiences when her mother abuses her verbally or physically; she probably does not call her feelings shame or anger or terror, but she registers the painful experiences in relation to her mother in her memory schemas. In her mind, and in her memory, she may then also turn away from the image of her mother as the source of the painful activation, to avoid acknowledgment that her caretaker upon whom she depends for life is also the source of danger to her life. She may also seek another source, in other people or in herself, to account for the painful feelings. The organization of her emotion schemas, including the disconnections as well as connections, are based on repeated experiences of the pattern; she comes to see the world of other people in a particular way, to expect such interactions, perhaps to see them when they do not occur; perhaps to act in a way as to bring them about, or in such a way as to avoid them (Bucci, 2021, pp. xxxv-xxxvi).

I think that, under the pen of any analyst, a psychological cause invariably includes something like the notion of memory schema – which is arguably just a more recent name for the classical notion of complex[38]. But the memory schema is a mediating element: it comes

[38]It is no surprise that, throughout her career, Bucci has aimed to demonstrate the blatant interfaces between psychoanalysis and the cognitive sciences (e.g. Bucci, 1997).

from the fantastical dimension of a certain event of early life and it is activated only by inputs that somehow resemble aspects of such an event. Therefore, clinical researchers in psychoanalysis find the cause of a behaviour only if they find: 1) the representation of a primary external and/or corporeal event (a primary fantasy); 2) the schema ensued from this primary event; 3) the features of the immediate circumstance that activate this schema (the triggers)[39]. One cannot fully characterize a cause inside clinical psychoanalysis unless these three elements are considered: primary cause, schema and triggers.

Another peculiar character of psychological causation that should interest us here is its constant *shuttling*, so to speak: a primary fantasy is a cause suitable to be transformed by its own effects, which are other representations, just because both are in the brain and functionally close to one another. A fantasy is nothing but a kind of memory, and we know that memory is liable to be updated/corrupted by striking new information from external and internal worlds. The new information may change the original fantasy-derived schema[40], which then may change the original fantasy. This shuttling is not exclusive to the mind – hit by billiard ball *a*, a previously still ball *b* starts to move, hits the corner of the table, returns and touches ball *a* again, converting itself from an effect to a cause – but, there in the mind, it happens all the time.

In psychoanalysis, this shuttling-updating potential of mental entities is framed by Freud with the German word *Nachträglichkeit*, which means “deferred action” or “afterwardsness”; the notion was usually present in his remarks on mental temporality and causality (Laplanche & Pontalis, 1988). It refers to the fact that “memory-traces may be revised at a later date to fit in with fresh experiences or with the attainment of a new stage of development”, gaining thereby “not only […] a new meaning but also […] psychical effectiveness” (Laplanche & Pontalis, 1988, p. 111). Recognizing the notion inside psychoanalysis dissolves the prejudice that the theory would reduce “all human actions and desires to the level of the infantile past” and that it would thus defend “a linear determinism envisaging nothing but the action of the past upon the present” (Laplanche & Pontalis, 1988, p. 111). Freud points out “from the beginning that the subject revises past events at a later date (*nachträglich*), and that it is this revision which invests them with significance and even with efficacity or pathogenic force” (Laplanche & Pontalis, 1988, p. 112). According to Freud, usually a memory-trace is not traumatic at the outset, but becomes traumatic when fresh circumstances give it a new meaning.

[39]The “triggers” would be the same thing Freud knew as the “precipitating” or “releasing” cause, “the one which makes its appearance last in the equation, so that it immediately precedes the emergence of the effect” (Freud, 1962a, pp. 135-136).

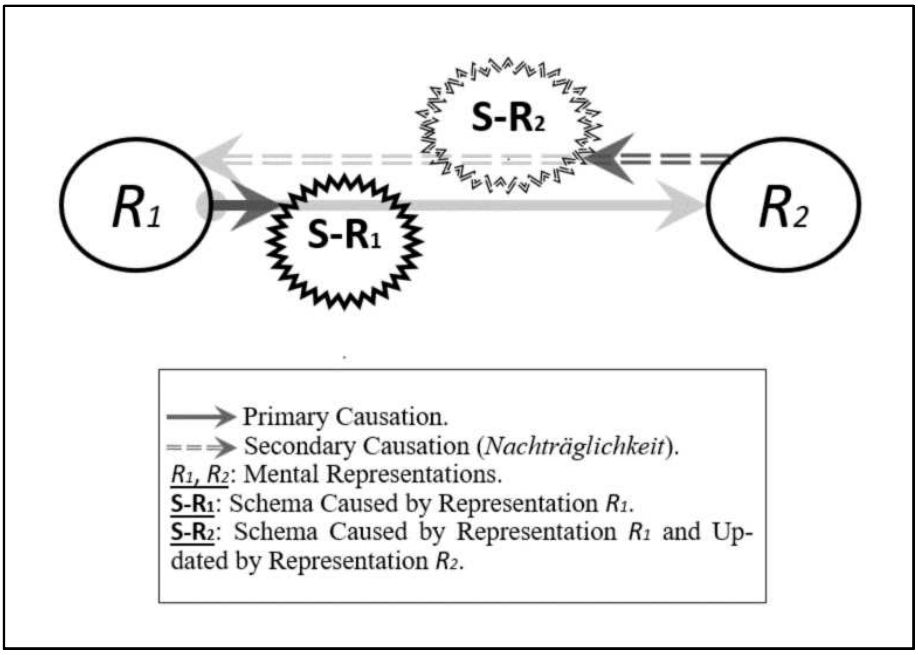
[40]Indeed, Bucci defines memory schemas as “organized representations of knowledge of all types that are activated *and altered* [emphasis added] by new experience […]” (Bucci, 2021, pp. xxxv).

Less gloomingly, the notion of *Nachträglichkeit* could also be used in the context of healing, abreast of the notion of *Durcharbeiten* (working-through). Indeed, it seems that most hypotheses about the mechanisms of psychotherapy count on the mind's potential to re-arrange and re-transcribe representations and schemas (e. g., Lane & Nadel, 2020). Still in a wider picture, Karl Friston's model of the mind, cherished by the recent neuropsychoanalytic movement, is all about shuttles and updates:

> Friston presupposes that the brain's aim, like that of the organism as a whole, is to maintain homeostasis and resist the entropic forces of chaos and homogenization. To do this we – along with our fellow living creatures – need information about the environment, our place within it, and the likely outcomes of our actions. The past shapes our futures: based on prior experience, we make "top down" predictions about our sensory and interoceptive input, based on a model of how they were created. The discrepancy between these top-down predictions and the actuality – and accuracy – of bottom-up sensations is "prediction error." Via perception and action, these unavoidable "errors" are "minimized" by converting prior beliefs into posteriors (i.e., the newly assigned probability after the relevant evidence, the observed data, is taken into account). This process of Bayesian inference simulates past experience and ensures posterior beliefs align with newly sampled data (Holmes & Nolte, 2019, p. 2).

Considering the points of this section, the path between representations could be thus depicted:

Figure 5 - The Causal Link Between Mental Representations



Source: Elaborated by the author (2021).

So, in reflecting about the causes of their patient's behaviours, it is not enough that analysts identify primary cause, schema and triggers; they must also count that today's primary cause, schema and triggers will not be the same as tomorrow's if some of the patients' new experiences end up reconsolidating the schema. Before that, they must keep in mind that such determining elements indeed have not remained the same along the patient's life.

The last point we should discuss regarding the idea of causation in clinical-psychoanalytic research is how it would include the ideas of necessity and sufficiency. Does the claim "this psychotic crisis in subject *s* was caused by early event *e*, by the cognitive-affective script that it has engraved, and by a related feature present in the circumstance of the crisis" mean that, every moment these three factors are present in *s*, *s* shall *inescapably* have a psychotic crisis? Does it mean, moreover, that a psychotic crisis in *s* could not have occurred unless these three factors had been present? Of course, the answer is "no" for both questions. A psychotic crisis may be induced solely by non-psychological elements, such as LSD ingestion and a state of high fever. Further, even when there are reasons for hypothesizing psychological factors for a psychotic crisis, there are also reasons for hypothesizing that such factors are not alone; for example, a vast amount of evidence shows an important role of genes in the development of psychotic syndromes.

This is one example, but there are countless others. It is hard to find a phenomenon of interest to clinical psychoanalysis that can be explained only at a psychological level, under a psychoanalytic light. It is hard to find a fantasy-schema-trigger tangle that could be either necessary or sufficient for a behaviour. Hard, perhaps impossible. So, as today we are not able to know the whole truth about the causes of every phenomenon of interest to psychoanalysis, it is prudent to affirm this unnecessary and insufficient nature is ascribable to the causes of every such phenomenon until proven otherwise. Yes, indeed impossible, if we think carefully. For, when it comes to human behaviour, there is always the action of purely "terrestrial" and purely "celestial" causes; there is always the biological level, with its particular powers (see the example of psychotic crisis in the paragraph above), and there is always the level of agency and free-will (if one does not take it to be an illusion). Standing in the middle of the two, psychological causes can be modulated by the ape and by the angel within us; and we could say that, sometimes, representations and schemas are uninvited to the biological happy-hour of human behaviour.

But how can something not be necessary or sufficient for a behaviour and still be considered its cause? We need an idea of causation that accommodates this "weakness". The classical paper by John Mackie, "Causes and Conditions", has not only formalized this idea to us but also indicated that it is present beyond psychology and the human sciences (Mackie, 1965). He teaches us that, when we talk of the cause of a phenomenon, at stake is not the idea of a necessary and sufficient condition, but the idea of an *INUS condition*, that is, an *Insufficient but Necessary part of a condition which is itself Unnecessary but Sufficient* for the occurrence of a phenomenon.

He illustrates the argument with the case of a certain house where fire has broken out; experts are able to investigate the house and conclude that the fire was caused by an electrical short-circuit in a certain place. Mackie (1965) then asks "what is the exact force" (p. 245) of their conclusion. For the experts are not claiming that the short-circuit was necessary for the fire: a short-circuit in another place, or the overturning of a lighted oil stove, or many other things, could have also set that house on fire. Nor are they claiming it was sufficient for the fire: no fire could have occurred if inflammable material were not present near the short-circuit, or if an automatic sprinkler had prevented it from spreading. Thus, what do the experts mean by "the short-circuit caused the fire"?

> At least part of the answer is that there is a set of conditions (of which some are positive and some are negative), including the presence of inflammable material, the absence of a suitably placed sprinkler, and no doubt quite a number of others, which combined with the short-circuit constituted a complex condition that was sufficient for the house's catching fire – sufficient, but not necessary, for the fire could have started in other ways. Also, of *this* complex condition, the short-circuit was an indispensable part: the other parts of this condition, conjoined with one another in the absence of the short-circuit, would not have produced the fire (Mackie, 1965, p. 245).

The short-circuit was, then, an INUS condition for the fire. Mackie (1965) gives a formal definition of his notion:

> *A* is an INUS condition of a result *P* if and only if, for some *X* and for some *Y*, (*AX* or *Y*) is a necessary and sufficient condition of *P*, but *A* is not a sufficient condition of *P* and *X* is not a sufficient condition of *P* (Mackie, 1965, p. 246).

We could cite other examples. Scientists say that smoking causes lung cancer. Nonetheless, it is possible that the problem never develops in smokers; alternatively, it is also possible that non-smokers come to develop it. At bottom, scientists mean that smoking is an INUS condition for lung cancer. The same is true for a claim that alcohol ingestion causes car accidents. Alcohol ingestion causes an accident when conjugated with other conditions, for example a high amount of it, tiredness, a busy road, etc.. And an accident can occur without drunk drivers: because drivers were using the mobile, because it was raining, because the car had mechanical problems, etc.. Last but not least, in clinical psychology, if one says that having been a victim of sexual abuse causes depression, one would also be implying that sexual abuse is an INUS cause of depression: not every depressed person has been abused, and not every abused person becomes depressed.

The "basic intuition" condensed in Mackie's acronym "is shared by researchers from many disciplines", including legal, epidemiological and economical ones:

> Legal scholars, for example, have advocated a relation called NESS (Wright 1988), standing for "necessary element of sufficient set," which is a rephrasing of Mackie's INUS condition in a simpler mnemonic. In epidemiology, Rothman (1976) proposed a similar criterion – dubbed "sufficient components" – for recognizing when an exposure is said to cause a disease: "We say that the exposure E causes disease if a sufficient cause that contains E is the first sufficient cause to be completed" (Rothman and Greenland 1998, p. 53). Hoover (1990, p. 218) related the INUS condition to causality in econometrics: "Any variable that causes another in Simon's sense may be regarded as an INUS condition for that other variable." (Pearl, 2009, p. 314).

In psychology, as well as in other human sciences, the notion is extremely helpful. Two pressing questions haunt the principle of mental determinism: the existence of agency and free-will on one side, and on the other the existence of non-mental (biological or physical) causes of mental phenomena, for example drugs, genes, famine, earthquakes, car accidents, etc.. If we take the psychological cause we ascribe to a behaviour to be an INUS cause, we respect the two questions and the complex nature of the mental domain; we recognize the metaphysical and ethical issues the domain raises without having to resolve them. On top of it all, we remain determinists. This is a very comfortable place for a science.

According to Mackie (1965), "the claims implicit within our causal assertions can be related to the forms of the evidence on which we are often relying when we assert a causal connection" (p. 245), and

> One possible account of general statements of the form "*S* is a necessary condition of *T*" and "*S* is a sufficient condition of *T*" – where '*S*' and '*T*' are general terms – is that they are equivalent to simple universal propositions. That is, the former is equivalent to "All *T* are *S*" and the latter to "All *S* are *T*." […] Whether an account of this sort is adequate is, of course, a matter of dispute; but it is not disputed that these statements about necessary and sufficient conditions at least *entail* the corresponding universals. (p. 253).

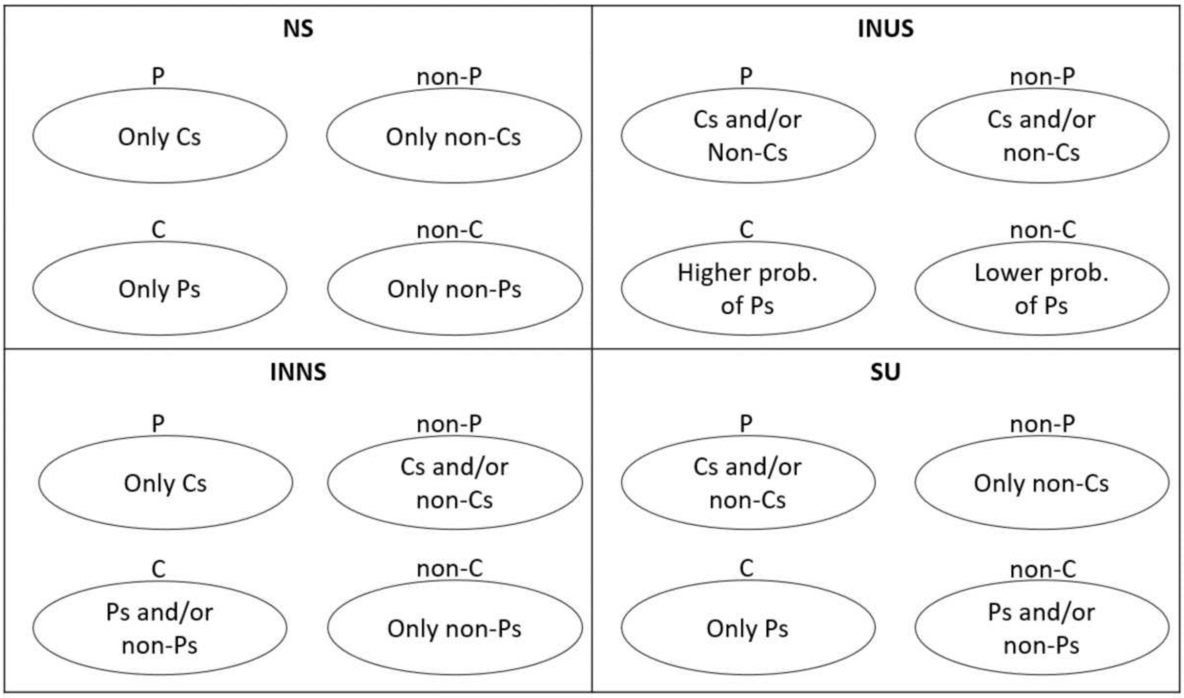
We should present what would evidence for an INUS causation be like, in other words, how much and how an INUS causation can determine the pattern of the data related to it. Let us go back to the examples of the burned house and the lung cancers. If we take a representative sample of houses where fire has broken out and another representative sample of intact houses, will we find in the former a higher number of instances of short-circuit in comparison to the latter? Not necessarily, since many things may cause house fires and many short-circuits may not cause house fires. Likewise, if we take a sample of people with lung cancer and compare it to a sample of similar people without lung cancer, we will not *necessarily* find more smokers in the first. However, if we gather samples according to the INUS cause, instead of according to the effect of the INUS cause, we find a somehow delimited pattern. In a sample of smokers, we will find more people with lung cancer than in a sample of non-smokers; as an INUS cause of lung cancer, smoking *increases the chances* of an organism developing lung cancer. The presence of short-circuits also brings more risks of fire than their absence.

We can remember, here, of the Method of Difference and of its use in experiments. For

> the Method of Difference does not require the utterly unrealistic sort of assumption used in what von Wright calls the simple case – namely, that the supposed necessary and sufficient condition is some single factor on its own – but that the much less restrictive assumption used here will still yield information when it is combined with nothing more than the classical difference observation. It is worth noting also that the information thus obtained, though it falls far short of what von Wright calls absolutely perfect analogy, that is, of a full specification of a necessary and sufficient condition, is information of exactly the form that is implicit in our ordinary causal assertions, both singular and general.
> But can observations of the kind required be made? A preliminary answer is that the typical controlled experiment is an attempt to approximate to an observation of this sort (Mackie, 1965, p. 256).

If we, inspired by Mackie (1965), play with the notions of necessary and sufficient conditions, we go beyond the two kinds of causes we have mentioned so far. We arrive at four kinds, in total: *NS Causes* are conditions which are both necessary and sufficient for the occurrence of a phenomenon; *INNS Causes* are conditions which are an insufficient but necessary part of a both necessary and sufficient set of conditions for the occurrence of a phenomenon; *INUS Causes* are conditions which are an insufficient but necessary part of an unnecessary but sufficient set of conditions for the occurrence of a phenomenon; *SU Causes* are conditions which are sufficient but unnecessary for the occurrence of a phenomenon. Mutation in a specific gene is an NS cause of the Tay-Sachs Disease; in Freud's NCT, a truthful interpretation is an INNS cause of a therapeutic effect; rain is a SU cause of a person's wet hair. Each kind of causation has a specific manner of determining the data related to it, as we can see in the figure below.

Figure 6 - Evidence for Each of the Four Kinds of Causation



P is a set of instances in which the phenomenon in question occurs; *Non-P* is a set of instances in which the phenomenon in question does not occur; *C* is a set of instances in which the cause in question occurs; *Non-C* is a set of instances in which the cause in question does not occur; *Ps* are instances in which the phenomenon in question occurs; *Non-Ps* are instances in which the phenomenon in question does not occur; *Cs* are instances in which the cause in question occurs; *Non-Cs* are instances in which the cause in question does not occur.

Source: Elaborated by the author (2021).

In extensively exploring which meanings the causal assertions of the psychoanalytic clinical research could have, we may develop the impression that its social usefulness is reduced to producing *causal knowledge* about mental phenomena. But its social usefulness is already blatant when it meets simpler tasks. *Phenomenological knowledge* is not causal, but it also depends on empirical data of the kind research in clinical psychoanalysis has privileged access to; and the usefulness of this knowledge is smoothly evoked. For example, no one would deny that as important as knowing the causes of PTSD is knowing how subjects diagnosed with it feel and think during crises, or how their symptoms develop along their lifetimes. *Conceptual knowledge*, too, although not founded on empirical data, is inspired and substantiated by them; psychoanalytic cases have always been sparks to revolutionary concepts whose influence is found everywhere.

4.6 CONCLUSION

In this chapter, I have presented which methodological, theoretical and epistemological premises shall guide my impending responses to Grünbaum's critique of psychoanalysis.

Inspired by other responses to this critique, I announced that I shall be concerned more with redirecting and solving the problems delimited by Grünbaum than with indicating errors in his interpretation of psychoanalytic and scientific standards. I named this approach propositional-synthetical. This problem-solving approach required that I defined an origin and a horizon for my study – a picture of clinical psychoanalysis to be preserved along the problem-solving moves, and a picture of the good scientific practice to be aimed at along the problem-solving moves.

I discussed a definition of clinical psychoanalysis through definitions both of the clinical method and of the most fundamental ideas of psychoanalysis. The clinical method was described as an empirical and non-experimental research method in psychoanalysis and elsewhere. I argued that this method produces a unique kind of knowledge and that experimental knowledge cannot replace it, but only supplement it. We have seen that the clinical research in psychoanalysis deals with both psychological and therapeutical questions, and we have seen the four kinds of results that can come from such questions – a General Theory, a Theory of Illness, a Theory of Therapy and Technical Rules – as well as the relationships between them. With these elements in hand, I stated that this doctoral thesis is a philosophical study about a specific kind of empirical research in psychoanalysis – the *clinical* research aimed at *psychological*, not therapeutical knowledge. Finally, faced with the problem of pluralism in psychoanalysis, I stated the three theoretical hypotheses and the two clinical-methodological rules that I take to be the origin of everything else in the discipline, the core of its core without which it can no longer be called psychoanalysis.

Then, I shortly presented a minimal definition of good scientific practice that I think should continually serve as a horizon to clinical-psychoanalytic research. I suggested that Schurz's "supreme epistemic goal of science" – "to find true and content-rich statements, laws, or theories, relating to a given domain of phenomena" (p. 19) – and his "minimal epistemological model", composed of five epistemological assumptions – Minimal Realism, Fallibilism and Critical Attitude, Objectivity and Intersubjectivity, Minimal Empiricism and Logic in the Wider Sense – are satisfactory guides to the realization of a proper scientific attitude.

In the end, I further specified what kind of scientific knowledge clinical research in psychoanalysis could produce in order to defend that it is worth conducting. First, I claimed that clinical psychoanalysis is a context distinctively adequate to investigate unconscious motives in any human dimension, which is the same as claiming it is adequate to investigate unusual or surprising human behaviour that have little or not fully to do with congenital factors or with immediate environmental constraints and pressures. Second, I defended the usefulness of adopting the idea that clinical psychoanalysis deals with three kinds of hypotheses – core, categoric and particular hypotheses – and I defined these kinds from a scopic, from a historical and from a logical standpoint. Third, I argued how we should understand the idea of mental causation in clinical-psychoanalytic research: it is a causation composed by a primary fantasy, a resultant schema and the environmental features that trigger this schema; it includes the idea that old and new representations are constantly interacting, in other words, that mental causation is always bidirectional; last, we should understand the triad of primary fantasy, schema and trigger as an INUS condition for a certain behaviour. I also presented other three kinds of causation (NS, INNS and SU) and the patterns of evidence for all four kinds.

The time has come for us to struggle with the epistemology of clinical-psychoanalytic inferences.

5 THE PROBLEM OF CAUSAL INFERENCE *EX POST FACTO*

5.1 INTRODUCTION: THE BLACK BOX OF OUR INFERENCES

In order to understand how a specific science works, we cannot be content with its surface, no matter how well we are able to detail it. Central bits of what scientists actually do cannot be found in what, along papers and lectures, they avow they do, nor in the overt behaviour of the lab or the couch. These central bits should be pursued throughout, behind the scene, between the lines; in philosophy of science, as in science itself, the unobservables that are producing the observables are important. The term "synthetical" of the approach I have devised and chosen for this study is condensing such working principle.

In agreement with it, Michael (2015) criticizes Grünbaum for having characterized as hypothetico-deductivist the inferential process that led Freud to the Repression Aetiology, since Grünbaum's premise was only *what Freud himself avowed about the process*. From the very fact that there exist philosophers of science, Michael (2015) sees such premise as illegitimate. He explains:

> The very fact that there exists a group of people – philosophers of science – whose research aims include, as a core and yet-to-be-fully-resolved problem, describing exactly what scientific reasoning is, suggests that the evaluation of hypotheses is not a transparent process. If it were, then why not simply ask scientists what they are doing? The raison d'être of philosophy of science is that, though scientists are proficient at doing science, they are not necessarily proficient at describing what it is they are doing. They are much like the expert cyclist, who though proficient in cycling, is not *qua* cyclist proficient at explaining the process of cycling. […] Given this, it seems strange to think that somehow Freud should be an authority on the inferential processes he was using. Why should he be? He is no epistemologist (Michael, 2015, p. 131).

The cycling-cyclist example is also found in a similar context more than a decade before, in Lipton (2004): "It is easy to ride a bicycle, but hard to describe how it is done […]" (p. 1). He stresses the point with another example: "[…] it is easy to distinguish between grammatical and ungrammatical strings of words in one's native tongue, but hard to describe the principles that underlie those judgments" (p. 1). With this point, Lipton (2004) is opening his book on Inference to the Best Explanation, a model that would account for how we make inductive inferences in everyday life and in science. He is reminding us of the trivial fact that, although we may be good at some task, we may never be able to describe how we do it or why we do it so well – and of the uncomfortable fact that the same is applicable to inductive inference and to explanation.

Lipton (2004) remarks how incompetent we are at describing, in general, the principles of whatever we do and, in particular, the principles of our inferential habits. By "we" he means not only ordinary reasoners, but also those, like him, who have burned a lot of mental fuel in the search for such principles. Indeed, he claims, the main accounts proposed by prominent epistemologists of the gears that would move our inductive inferences lack many of the latter's actual features. The epistemologist's struggle is a strange one, for we humans use these principles all the time; we possess them, but we cannot see them; they are close to us, but out of reach.

The astonishment around Chomsky's work on the principles involved in the blossoming of language and around Kuhn's work on the ones involved in the stream of science would, for Lipton (2004), consist in proofs that the principles of such common deeds as talking and researching cannot be accessed by introspection nor observation. Their results would not be so important and controversial were we already aware of how we distinguish grammatical from ungrammatical sentences or of how scientists prefer one hypothesis over another. Speakers and scientists are as conscious of the principles they use as anyone supposing that a friend is having problems at home or that the street is wet because it just rained. Lipton (2004) claims that, "although we may partially articulate some of our inferences, if, for example, we are called upon to defend them, we are not conscious of the diverse principles of inductive inference we constantly use" (p. 12).

Since these principles cannot be plainly accessed by introspection nor observation, we must reconstruct them from what can be thus accessed. The solid things we have are the evidence we considered and the inference we made, and therefrom we can try to solve the mystery that happens between them. The philosopher invokes the concept of "black box" here. We try to reconstruct the mechanism governing the "box" (in this case, our inferring brain) from the patterns between its inputs and its outputs. Thus, he argues, a harsh matter of underdetermination arises. As long as these patterns are compatible with more than one mechanism (which is expected), the question of which mechanism is actually operating remains. Our evidences and rules of deduction underdetermine our inferences, which incites us to search for principles and provides hints to find them; but that imbalance is also not enough to enable us to determine the principles we miss. What is worse: it is common that epistemologists are unable to bring forth *any* description that would sustain the patterns observed; a tendency, by the way, suffered also by scientific theorizing. The epistemologist's struggle mentioned above, then, should not appear as strange at all:

> Why should we suppose that the project of describing our inductive principles is going to be easier than it would be, say, to give a detailed account of the working of a computer on the basis of the correlations between keys pressed and images on the screen? (p. 13)

The epistemologist's struggle, though, is not only over descriptions, but also over justifications. It is one thing to know how the mechanism works, another to know if it is successful or not in the pursuit of truth and why so. The way we infer may bear fundamental vices which, although innocuous in most cases, may lead to error in some.

One might think that one cannot begin to justify inductive principles before concluding the task of their description; or else that one should not try to detail the description of inductive principles unless one can trust them at some level. Lipton (2004) remarks, though, that the problem of description did not wait for a solution to the problem of justification to be vigorously tackled; no wonder, since the description of how our inferences actually work is obviously valuable even if we grant that they lack justification. The philosopher also remarks that in the history of epistemology the problem of justification preceded the problem of description; among other classical sceptical arguments, the Humean argument that the justification of induction is circular is so radical that it remains effective no matter how we specify our processes of induction. Despite the ravages caused (?) by Hume, the task of justification is still active among epistemologists:

> The peculiar difficulty of meeting Hume's skeptical argument against induction arises because he casts doubt on our inductive principles as a whole, and so any recourse to induction to justify induction appears hopeless. But one can also ask for the justification of particular inductive principles and […] this leaves open the possibility of appeal to other principles without begging the question. For example, among our principles of inference is one that makes us more likely to infer a theory if it is supported by a variety of evidence than if it is supported by a similar amount of homogenous data. This is the sort of principle that might be justified in terms of a more basic inductive principle, say that we have better reason to infer a theory when all the reasonable competitors have been refuted, or that a theory is only worth inferring when each of its major components has been separately tested (Lipton, 2004, p. 11).

Although complementing each other, then, the tasks of description and justification are relatively independent.

Nonetheless, and in spite of all the hazards involved, I shall in the following lines present *both* a plausible theory of what happens in the "black box" of psychoanalysts while inferring about the mental life of a patient in a clinical context *and* some arguments in favour of the reasonableness of the box's performance. At the end, I shall argue that the Grünbaumian charges related to the inferential practices of clinical psychoanalysis are founded in a poor picture of what these practices are and/or can readily become.

Both philosophers Peter Lipton and Michael Michael presented the same cycling-cyclist example to introduce a same point about the distance between knowing and knowing how one got to know. Could we claim from this simple data that Michael has read Lipton and taken the example from him? Or should we consider the chance that the example in Michael came from elsewhere – maybe from the popular saying "we never forget how to ride a bike"? (Could some background data be a game-changer to decide which one of the hypotheses is more probable – for example, the facts that Michael (2015) expressed that the model of Inference to the Best Explanation accounts well for psychoanalytic inferences, that Lipton (2004) had been a prominent substantiator of such model and that, at any rate, Michael (2015) cited the book in which Lipton (2004) had presented the example?) Why are we so tempted to infer causal links from similarities such as these? Is such an impulse justifiable? In which circumstance?

The plausible theory unravelling how psychoanalysts infer in a clinical context that I shall present is all about leaps, like the above paragraph's, from similarities – or, in Grünbaum's jargon, "thematic affinities" – to causal links. I shall argue that Inference to the Best Explanation (IBE) is the best explanation for the course that seemingly strong inferences in clinical psychoanalysis tend to take, and that this model leans on Millian and Bayesian principles. We shall see that the kind of leap from thematic affinity to causal link in the mould of IBE found in clinical psychoanalysis can also be found in the historical sciences, and that such an analogy allows us to extend the problems and solutions of justification found in the historical sciences and the IBE model to the domain of clinical psychoanalysis.

5.2 THE CAUSAL INFERENCE IN CLINICAL PSYCHOANALYSIS: A MODEL

5.2.1 Introductory Remarks

With the model I am about to present, I should be able to demonstrate several points in favour of the worth of psychoanalytic inferences. I ask the readers to have such points in mind in order to anchor themselves to what is vital along the stream of paragraphs. The points to be demonstrated are:

1) That clinical psychoanalysts can make rich and surprising inferences, in spite of some theoretical bias, which is in part reflected in some bias of material-selection;

2) That clinical psychoanalysts can employ Mill's Method of Difference in their investigations, and that they probably have always employed it at least implicitly and/or unrigorously;

3) That clinical psychoanalysts can discriminate between thematic affinities that comprise a close causal network and those that do not, in fact that probably they have always reasoned thus, at least implicitly and/or unrigorously;
4) That the *ex post facto* inferences in clinical psychoanalysis can be strong;
5) That inferences from the singular level to the general level, such as the ones made in clinical psychoanalysis, can be strong;
6) That clinical psychoanalysts infer in a manner similar to historians, and therefore that the former can be at least as reasonable as the latter;
7) That inferences of psychoanalysts are special cases of inferences to the best explanation, or abductions, and that, as such, they can include, and perhaps enrich, both Millian and Bayesian rationales.

The task I have chosen to fulfil, thus, is not merely to describe the gears of the black boxes of analysts, but to describe solely such gears that are integer and polished. The model below is not devised to encompass every possible inference every analyst has ever made, including the weak ones. My intention with it is humbler: to *hypothesize* over what is the *strong parcel* of the inferential process of *most* psychoanalysts that have *critical-realist principles*[41]. This is almost as herculean as haphazard description; but the word "hypothesize" could not be overemphasized. What I intend is to propose a *theory* about the inferential process of analysts, a theory that is clear enough to be questioned and revised by philosophers and analysts, clear enough to bear the risk of falsehood. As long as such theory can contribute to keep similar attempts alive, it should not matter if it comes to be considered an unfaithful description of what transpires in the mind of an analyst.

But, of course, the model is intended to be faithful. Though my empirical sources to construe it were not systematically consulted (my research is not empirical, after all), they *are* the psychoanalysis "out there": every contact I have had with psychoanalytic literature, especially the literature of clinical cases, the psychoanalytic circles of which I have taken part and the psychoanalytic treatments I have traversed[42]. So, when I say "analysts" in "the black box of analysts", I am referring neither to every real analyst nor to an ideal, divine analyst; I am referring to a compromise of both, to the ideal part of what is real, to a synthesis of the rational behaviour of real analysts. Even if the model presents concepts that are not found in

[41]Or, from another angle, to reveal the *potential* of *any* clinical psychoanalyst *in reaching (approximately) true theses* about their subjects (in the two senses of "subject").
[42]I should remark that, although I am a psychologist and I have done an internship in clinical psychoanalysis, I have never been a clinical psychoanalyst since I got my degree.

psychoanalytic literature, of course I want an analyst to look at it and recognize something very psychoanalytic indeed. Needless to say, I leave for the critical reader the decision of whether or not I am successful in this conciliating endeavour.

But the chance that the model to be presented reflect the real practice of analysts cannot be low, given it is highly inspired by the writings of many philosophers who have attempted, since Freud's time, to come up with models of the kind, or at least with concepts that could facilitate the elaboration of models of the kind. It would be dishonest not to grant that my empirical sources could have only been absorbed via the theories and concepts these philosophers have provided. The model below could be seen, thus, as just a deepening of previous efforts. I should now mention some of these efforts.

The model below is inspired, first, by pre-Grünbaumian arguments. In a timid but clear way, Freud himself suggests that analysts should not reject the idea of estimating probability-values in their work. By the end of the discussion on AJ's forgetting of the word "*aliquis*", Freud (1960) tells that AJ asks "but mayn't all this just be a matter of chance?", to which the analyst responds:

> I must leave it to your own judgement to decide whether you can explain all these connections by the assumption that they are matters of chance. I can however tell you that every case like this that you care to analyse will lead you to "matters of chance" that are just as striking (p. 11).

And in a footnote to this last phrase added in 1924, he cites a 1919 book by Eugene Bleuler[43] where the latter had demonstrated that psychoanalytic interpretations are amenable to statistical analysis:

> This short analysis has received much attention in the literature of the subject and has provoked lively discussion. Basing himself directly on it, Bleuler (1919) has attempted to determine mathematically the credibility of psycho-analytic interpretations, and has come to the conclusion that it has a higher probability value than thousands of medical "truths" which have gone unchallenged, and that it owes its exceptional position only to the fact that we are not yet accustomed to take psychological probabilities into consideration in science (Freud, 1960, pp. 11-12).

Would this be the seed of a Bayesian psychoanalysis still to bear palatable fruits?

In the classic "Psychoanalytical Method and the Doctrine of Freud", originally published in 1936, Roland Dalbiez (1941) discusses the criteria to generate and assess psychoanalytic interpretations that would "enable us completely to dismiss the objection of arbitrariness" (p. 113) that many have raised against such causal inferences – criteria among which are how *similar*, *frequent* and *convergent* the patient's associations are. As far as I know,

[43]The book's title is "*Das autistisch-undiziplinierte Denken in der Medizin und seine Oberwindung*" – "The autistic-undisciplined thinking in medicine and its overcoming".

he was one of the first post-Freudian philosophers[44] to discuss the principle that "similarity implies a relation of causality, either direct or indirect[45]" (p. 111) in the context of clinical psychoanalysis, and to comment that "the criterion of similarity is used in a number of various investigations besides psycho-analysis, from geology to scientific police-work, by way of philology, which involve the relating of a clue to its cause" (p. 111). Like Bleuler, he discussed probability in relation to interpretations:

> If instead of a single symptom we have at our disposal a series of symptoms exhibiting a common element, the probability of a cause resembling this common element will be more considerable. In that case the criterion of convergence comes to the aid of that of similarity (p. 116).

In his famous 1974 paper "Freud, Kepler and the clinical evidence", Clark Glymour shows that testing a psychoanalytic hypothesis through another psychoanalytic hypothesis can be an uncircular and legitimate strategy; that clinical data can be probative, after all. Glymour (1982) ascribed a so-called bootstrapping strategy of hypothesis testing to Freud's study of the Rat Man. One is bootstrap-testing a theory when one: 1) is using bits of the theory to compute non-theoretical data, drawing from this data theoretical and non-theoretical inferences; 2) is checking whether these results cohere with some of the other bits of the theory and other non-theoretical data. It is "bootstrapping" because part of the evidence used for testing a theory is computed by that same theory. According to him, it is a strategy frequently used in the non-experimental sciences, but also "whenever an experimenter attempts to test a theory containing quantities which the experimenter does not know how to determine save by measuring other quantities and computing using some of the very hypotheses to be tested" (Glymour, 1982, p. 18). Freud's study of the Rat Man would embody this strategy with states of affairs rather than quantities.

When in contact with Rubinstein's discussion on the self-atypical and group-atypical behaviour tracked by analysts in their quests for *explananda*, I have settled the idea, to be presented below, that clinical analysts are able to devise control-instances. "Once we have learned to know a person", he disserts, "deviations from his self-typical behavior are not difficult to recognize"; for example, "if a person has, to the best of our knowledge, all his life

---

[44] I would write "the first philosopher" had not Dalbiez himself cited the following excerpt by Raymond de Saussure: "Associations often lead to recent events which are so analogous to the dream-image that the most probable hypothesis is that the two phenomena are causally related" (de Saussure, 1925, p. 49, as cited in Dalbiez, 1941, p. 111).

[45] "Given two similar entities E_1 and E_2, we have a choice between three hypotheses to explain their resemblance: (i) E_1 is the cause of E_2 (direct causality); (ii) E_2 is the cause of E_1 (direct causality); (iii) E_1 and E_2 have a common cause E_0 (indirect causality)" (Dalbiez, 1941, p. 111).

been shy and withdrawn, then his suddenly becoming quite outgoing is bound to catch our attention" (Rubinstein, 1975, p. 8). As to knowledge about group-atypical behaviour, it would be acquired

> by each of us in going about the business of living, which includes reading, dealing with people of various sorts and in various ways (including living with some of them), hearsay, travel, etc. This knowledge (which hardly extends to all the groups and subgroups of the species) takes the form of *rough generalizations* about the *ranges* of "normal" behaviors. It seems safe to assume that each form of behavior has its own "norm" and each "norm" its own range, and that both may vary from group to group (p. 8).

Like Glymour did within his particular framework, Rubinstein (1975) showed me that particular hypotheses are inferred "in accordance with one or several general clinical hypotheses in conjunction with the original observation (i.e., the event to be explained) and whatever other relevant data we may have available" (p. 11).

In the years immediately following Grünbaum's critique, two authors have responded to it by launching new understandings of how analysts reason, or should reason. James Hopkins and Wilma Bucci, already appreciated in Chapter 2 (see pp. 46-48, 51-52), are also central sources to the model below. Hopkins (1988) was one of the firsts, perhaps the most famous, to have responded to Grünbaum by not letting go the idea that thematic affinities are competent guides to strong inferences, Grünbaum's argument notwithstanding, and to bring the late-20th-century account of Inference to the Best Explanation closer to clinical psychoanalysis; it shall become clear that I am part of this tradition.

With her observable version of the Tally Argument and her introduction to the dual-code model of the mind, Bucci (1989) made me feel confident with the concepts of specificity and repetitiveness, by which I had already been seduced in reading other epistemologists of psychoanalysis. She states that a fundamental premise of the dual-code model is that

> concrete and specific words are most likely to connect to imagery. Thus referential connections joining the two systems[, i.e., joining the verbal system to the perceptual-motoric-emotional one,] would be most direct for concrete descriptions of specific entities – e.g., "a fuzzy, longhaired orange cat" – progressively less direct for higher order category terms, e.g., "a cat", and supercategory terms, e.g., "an animal", for which there are no specific exemplars; still less direct for abstract terms such as "truth", or "goodness" (Bucci, 1989, p. 264).

Thus, the specificity of the words used by patients would indicate that they are "touching" their own nonverbal schemas; in other words, it would be an index to the patients' Referential Activity. Patients would manifest these nonverbal, emotional schemas many times in many ways, and Bucci (1989) tells she and her colleagues have developed procedures to identify such repetitive and specific structures: "we condense what the person says and does to its common elements across specific situations, people, and places by a successive process of increasing generalization" (p. 274). Best of all, Bucci (1989) provides experimental evidence favouring this dual-code model.

Finally, I recognize my model and concepts also stands over the shoulders of models, concepts and hints provided recently by the philosophers Michael (2008, 2015) and Lacewing (2012a), both appreciated in Chapter 2, as well as by Kevin Lynch (2014). The list of factors making a psychoanalytic interpretation more valid devised by Michael (2008) and deepened by Lynch (2014) were certainly the first instigators of the ideas found in this chapter. Michael (2008) argues that a dream interpretation is more probable the more "striking" the connections it establishes, the greater the number of dream elements it unifies, and the shorter the association chain it explores. Bleuler's and Dalbiez's intellectual descendants, both he and Lynch (2014) demonstrate that clinical psychoanalysis is indeed an arena for probability calculations. I am thankful for Lynch's (2014) observation that Michael's notion of a "striking connection" might have something to do with the *rarity* of the akin element that sustains the connection. Also helpful was Michael's (2015) notion of "distinctive similarity", which seems to be his own guess of what a "striking connection" should mean: "distinctive similarity occurs when an associated thought is similar in uncommon detail and to a high degree of precision to a dream element" (p. 90). As seen in Chapter 2, above all he has advanced Hopkins' suggestion and has rigorously put clinical psychoanalysis inside the frame of Inference to the Best Explanation. Last but not least, I found in Lacewing (2012a) an endorsement that psychoanalytic inferences could be seen as essentially Millian (see Chapter 2), contrary to what Hopkins (1988) had argued.

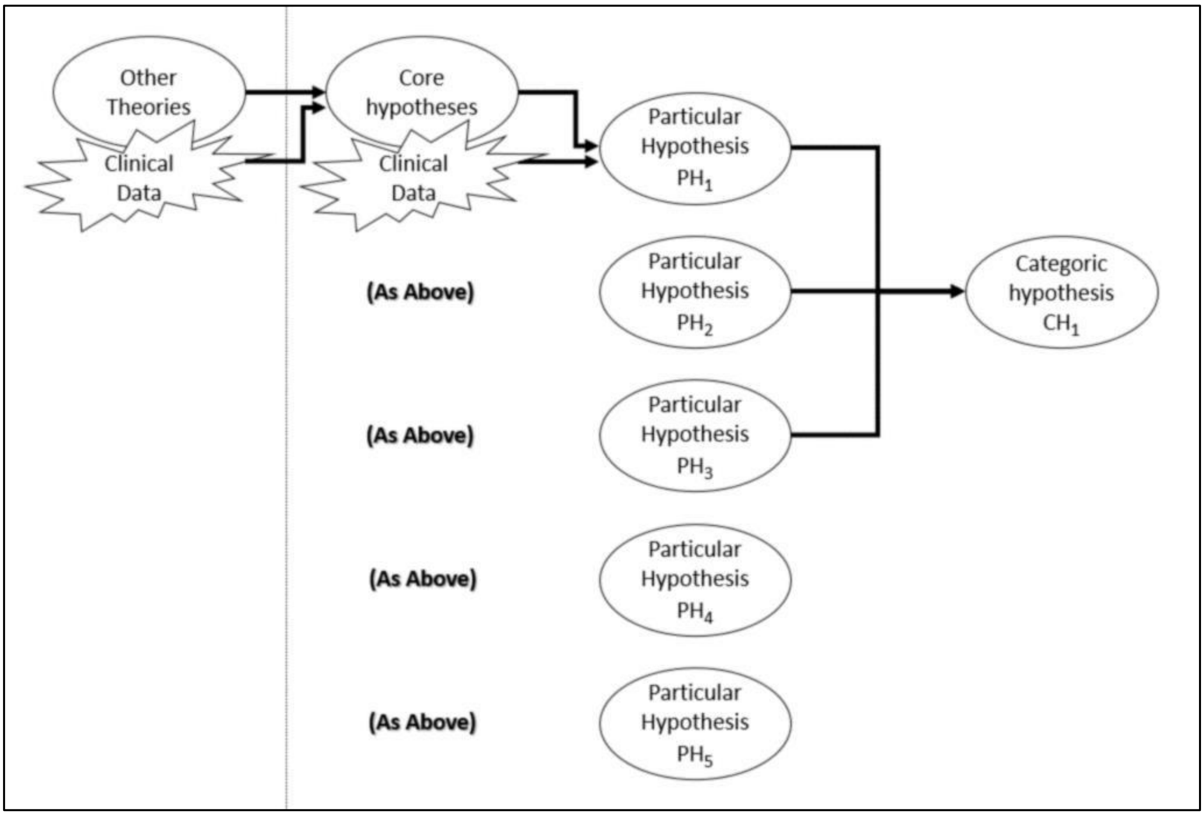
All of these authors inspire or endorse the model below. It shall become clear that it was also inspired or endorsed by others, such as Edwin R. Wallace, Carol Cleland, Elliot Sober, Aviezer Tucker, Peter Lipton – and by Grünbaum himself, to which ironically the model serves as a counterargument.

5.2.2 The Dynamics Between Core, Categoric and Particular Hypotheses in Clinical Psychoanalysis

Let us start with the widest picture of the model so we can then explore its details with a proper sense of direction. As we have defined in Chapter 3, there are three kinds of hypotheses in clinical psychoanalysis. Core hypotheses are the ones valid for all human minds, categoric hypotheses are the ones valid for the minds of a segment of humankind and particular hypotheses are the ones valid for just one person. The core are the hypotheses that devotees have kept despite all of psychoanalysis' transformations and the ones without which the clinical research in psychoanalysis would not be able to explain human behaviour, while the categoric and particular are the hypotheses that as a rule have been historically questioned and which are not indispensable to explain some behaviour in psychoanalytic terms. Something important was not defined in Chapter 3, though: how the three hypotheses relate to one another.

The particular hypotheses are born and raised out of the interaction between core hypotheses and the data provided by the patient in question. The categoric hypotheses, out of an assemblance of particular hypotheses across many patients who share the variables under study. Finally, the core hypotheses rose and rise out of non-psychoanalytic theories and unprecedented, or hitherto unnoticed, data. Thence, categoric hypotheses are based on particular hypotheses, which are based on data and core hypotheses, which are based on other theories and data. We can also understand the dynamics with the figure below:

Figure 7 - The Dynamics Between Core, Categoric and Particular Hypotheses in Clinical Psychoanalysis



Source: Elaborated by the author (2021).

We have already discussed the risk of theoretical bias in clinical psychoanalysis (see p. 33). If truthfulness in clinical interpretations is seen as a function of how much they cohere with psychoanalytic theory, if clinical data are always seen through the lens of the theory and considered as much as their confirmatory character goes, the method can be accused of ignoring refuting data (of showing a confirmation bias) and/or of assuming ahead what is to be proved (there are at least three ways to name this problem: to be circular, to beg the question, to commit the fallacy of *petitio principii*). We need, then, to find a *non-circular, somehow theoretically independent path* to test the theoretical material of clinical psychoanalysis, be it an interpretation, be it a more general hypothesis.

Well, in the figure just presented – intended to represent a rational program of clinical research in psychoanalysis – we see no circle at all. Categoric hypotheses, the most relevant product of such a program, can be refuted by cases and their particular hypotheses. Particular hypotheses, in their turn, can be independently refuted by the raw data of the case in question and by the low reliability of the core hypotheses used to interact with such data. Finally, core hypotheses can be refuted either by the raw data of a significant sample of clinical or experimental instances or by the low reliability of the non-psychoanalytic theories on which they are based. By "raw data" I do not mean a theory-free data, an accomplishment that many agree is impossible, but simply a non-psychoanalytic or a commonsensical account of phenomena (see Edelson, 1984, and Schurz, 2014, for more on this discussion).

Now let us see what the diagram is precluding and why. Why could particular hypotheses not come from categoric hypotheses? Once the truthfulness of categoric hypotheses becomes relatively accepted, why can they not be used for inferring particular hypotheses? One reason of a logical kind is that, if we accept this, then categoric hypotheses could be based on particular hypotheses and other categoric hypotheses (along with the other particular hypotheses founding the latter), and so on; this would make the independent assessment of all of such thick hypotheses very difficult, not to say impossible.

Another reason, not a logical but an ontological one, is that the human kinds whose unconscious leanings categoric hypotheses try to grasp and understand – human kinds such as "men", "women", "artists", "philosophers", "scientists", "bisexuals", "immigrants", "psychotics", "depressives", "suicidal", "antisocial", "religious", "nihilists", "incels", "women with OCD", "men with fibromyalgia", "gay serial-killers", etc. – are contingent to a certain culture as it is manifested in a certain historical interval. Such human kinds are contingent in the sense that a great deal of the legitimacy and consistency they enjoy as concepts is given by this time-bound culture; therefore, they are also contingent in the sense that their general features and behaviours are always changing.

Let us take the categoric hypothesis of "penis envy", for example. From treating a bunch of female patients from Central Europe in the midway between the 19th and the 20th century, Freud came up with the hypothesis that many features of women's minds are determined by an unconscious envy for the symbolical and physical advantages of the male organ. Let us suppose Freud could rigorously support the hypothesis of penis envy with the data he had in hand; it would still be irrational to suppose that the hypothesis could best-explain the behaviour of, say, late 20th century North-American women. Grünbaum is right in disapproving the analyst's conduct in Blanck and Blanck's (1974) case report (see pp. 19-20), although I would add that the analyst's indoctrinating moves constitute her most superficial mistake. The deep mistake there is what led her to make the moves: inferring something about the motive of a particular person from a hypothesis over the class of persons she would belong to, a kind of hypothesis which by its very nature must regularly be amended and updated. In our days, most reasonable analysts are alert enough to never universalize what is obviously contingent, maybe because of the significant number of prominent works pointing to this kind of mistake in the Freudian legacy.

This takes us to why categoric hypotheses must come from particular hypotheses. The only way analysts can divert from dogmatically stating categoric hypotheses or from fallaciously deducing the latter from core hypotheses is by having direct contact with the clinical phenomenon and by allowing the clinical phenomenon to "speak" and surprise them, by allowing it to remind them of how complex we humans are. In sum, particular hypotheses must have the role of keeping categoric hypothesis close to the *reality* of the categories the latter are about. Analysts must delve into the extremely singular before risking hypotheses about human collectives also because the unconscious mind can only manifest itself and be seen and heard inside an intimate and controlled relationship. (It is instructive to remind of the perks and perils that follow from the clinical method's fit to investigate suffering persons, presented in Chapter 3.) A miserable mind is not a "closed universe", it stands in a crossing of cultural influences; actually, it is usually more "porous" than the mind with the ordinary amount of suffering. However, analysts would only begin to touch the truth about the regular features and behaviours of the classes a patient may be placed into if they investigate such patient *as if* he or she were a "closed universe". The analysts who wish to put forward theories about a class of human beings before *individually* and *deeply* investigating a sample of the human beings who compose the class are misleading themselves.

But why could categoric hypotheses not come directly from core hypotheses and the raw material the patient presents, like particular hypotheses do? Well, because then the analyst-researcher would be involved in two steps of hard labour at once: to hypothesize over particular dynamics and to hypothesize over the particular dynamics, among these, that are common to certain categories they may belong to. I have just affirmed that, if classes of analysands are to be investigated, it is *sine qua non* that every analysand is treated as a "closed universe", as a singular and isolated system, since the unconscious mind in general can only manifest itself and be apprehended inside a close and scrupulous relationship. But there is another reason for holding on to such nutshell-principle. A "loose" clinical data, a piece of information with no regard for the personality and history of the individual subject of which it once came out, is worthless: out of its context, clinical data are easily distorted and misunderstood. Thus, the step of generating and testing particular hypotheses cannot be skipped. The step of dealing with particular hypotheses cannot be skipped and, we shall see, it is, alone, a very complex step. It would be burdensome, therefore, to derive categoric hypotheses directly from raw data and core hypotheses.

Finally, why could new core hypotheses not come from categoric ones?[46] For the same reasons, the logical and the ontological, that particular hypotheses could not come from categoric ones. If such a move were allowed, research in clinical psychoanalysis would have to assess hypotheses that would most likely be a mishmash of countless others, hindering a scientific stance. Besides, core hypotheses are supposed to be true in each and every human mind and, as we know now, categoric hypotheses are confined to a more or less extended moment in a certain culture.

In sum, a categoric hypothesis must be like a wagon that is coupled last in a train so that it can be easily replaced if its wheels get too old and rusty. It must always be "loose" – never hidden or implicit, never defining the terms of any other hypothesis, never evading criticism and revision.

Now we shall explore the process of generating and testing each one of these three kinds of hypotheses in clinical psychoanalysis, starting with the particular hypotheses.

---

[46]I have suggested in Chapter 3 that from illness theorems a psychoanalyst infers general psychological or anthropological theorems because he or she assumes that a psychopathology is a hyperbole of non-morbid, or "normal", features of the human mind; for short, that illnesses are natural experiments. Even though, according to my definitions, Freudian psychopathology is usually found in a categoric mode – for example, "hysteria is manifested as…", "a schizophrenic outbreak is triggered by…" –, what I have affirmed in Chapter 3 *does not imply that core hypotheses can come from categoric ones, for both illness theorems and psychological theorems can be either "core" or "categoric"*. An illness theorem in psychoanalysis can be "core" – can be about how and why *any human*, in any time or culture, develops a mental ailment. And a psychological theorem can be "categoric" – for example, a theorem about the mental life of *"normal" masculinities* (in a certain time and culture) is about general psychology, not psychopathology.

5.2.3 Generating and Testing Particular Hypotheses

In order to generate and test particular hypotheses, analysts must regard a human individual as a whole domain. In experimental research, the domain is constituted by numerous *individuals*. In clinical research in psychoanalysis, on the other hand, one must keep in mind at first that the domain is a pool of *mental products of one individual*. Just as in experimental trials each individual is an instance that is struck by different factors, so in clinical cases each mental product of one individual is an instance that is struck by different factors. If experimenters devise groups of individuals exposed and unexposed to the factor under study – respectively, experimental-groups and control-groups –, analysts can devise something similar: a group of instances of the mental product of interest and the circumstances in which they have arisen – which we can call target-instances – and the rest of the patient's mental products and their circumstances – which we can call control-instances.

Nonetheless, analysts can also infer particular hypotheses as if they had another kind of control-instances in mind: the range of mental products that are probable for a subject in a certain circumstance *given the subject's culture*. Thus, analysts not only compare some of the patient's behaviours with the rest of the patient's behaviours; they also tend to compare some of the patient's behaviours with the range of behaviours expected in the same circumstance, in the same culture.

Let us dissipate the fog that may have appeared here with simple examples. A patient reports that her episodes of sleep terrors started around 2012 and happen only in houses that are not hers – her analyst will investigate the circumstances, both of 2012 and of the houses in which the episodes occurred, *which, at the same time, were not in her life before 2012 and are absent at her home*. Another patient tells his analyst he drinks 5 litres of water a day – his analyst will investigate why he does that if no widespread practice of the culture he is part of, religious, sanitary, or what have you, would explain this habit – why he does that *while his fellow citizens do not*, or why he does that if that is not something we would expect a man of his kind to do.

The range of normal behaviour within a culture – "normal" here bearing a quantitative or probabilistic, not a qualitative or moralistic, sense – is unconsciously estimated by analysts from all of their experiences as members of that culture or, in case the patient has been nurtured in an atmosphere very dissimilar to the analysts', from their minimal understanding of the patient's specific heritage. The process of conceiving this kind of control-instances can at times be a bit hazy and idiosyncratic but at other times it can be highly plausible. When analysts recognize a slip of speech, they are conceiving a control-instance "from Normaland", for a slip of speech is, by definition, an error relative to the manner in which a group of people normally uses the language; when witnessing a speech error, analysts ask themselves why the patient spoke this instead of something usual and what is making the patient act differently from the other speakers. Analogously, if an adult patient has been sexually chaste for many years and he is young, healthy, and not religious, the analyst will unconsciously compare him to the "normal men of his kind" to try to find a disparity between the psychosocial features of the two "groups", one composed solely by the patient and the other by the "normal men of his kind". We can also imagine a patient whose family members are all choleric except for her, and her analyst wondering which unconscious motives contributed to make her the deviating member; here the control is a very restricted culture (the family's), but a culture nonetheless.

The examples and principles just presented suggest that the target-instances of clinical psychoanalysis are never routine behaviours, never behaviours we could witness in the whole life of one person or in the present life of many. We have seen in Chapter 3 that the objects of clinical psychoanalysis are unconscious motives, motives that have become or have always been unconscious because they disturb a subject's self-conservation, self-image and/or sociability. The analysts' special esteem for behaviours that are incongruous with the patients' life-course as a whole or with the habits and customs of the patients' fellows becomes quite understandable therefrom. Besides, a dissonant behaviour asks for an explanation and, at the same time, gives analysts the resources to find it. (A preference to explain what "goes awry", however, is found in every science; see Lipton, 2004).

The two kinds of control-instances just discussed are involved in the process of inferring particular hypotheses. Let us characterize this process in detail now. It is composed of two general steps:

1) Inferring the *mere existence* of a causal linkage between mental products – otherwise stated, that a set of mental products share a causal network, that the mental products at hand influence each other in the patient's psyche, etc..

2) Inferring the *exact nature* of this causal linkage.

This last step is composed of two interlaced tasks:

2a) inferring whether or not there must be a *common cause* to the mental products that are causally interrelated – in other words, whether these mental products explain each other well enough or are better explained by the fact that a yet-to-be-known mental product is causing both of them.

2b) inferring the thrust of the *causal process*, in our case, the unconscious reasoning that works in-between the mental products at hand (the complex or schema, as seen in Chapter 3).

Let us begin with *1*. In order to infer the existence of *some* causal linkage between mental products, the pivots of analysts are, yes, *thematic affinities*.

Perhaps a stubborn claim considering Grünbaum's assertive reservation regarding their epistemological value (see pp. 27-29), but I hope to restore this value with the arguments to be presented hereafter. Let us rescue his reservation. We cannot infer that there exists a causal relation between instances with similar themes from the mere existence of this similarity because such instances can be the result of an *independent* unfolding of two or more causal networks, the similarity between them being thus a random and meaningless state-of-affairs, a mere accident. He presents a case where such a prudent reminder is absent: the 16th century medical advice that, in order to cure liver disorders, we should ingest herbs that look like a liver (see p. 28). Grünbaum (1989) claims that the logical problems of this case "are no worse than those of causal inferences from mere thematic connections, which abound nowadays in some theorizing about human behaviour" (p. 489).

There are two points to be made here. First, considering Grünbaum's purpose with the example, it is no trivial matter that it concerns a relationship between plants and a non-psychological malady – that it concerns a system of beliefs over a *biophysical* domain. In dissimilar manner, psychoanalysis is a system of beliefs specifically about the domain of *human psychology*. Well, Paracelsus had a mind and was a man. If Paracelsus made such an inference, what is more, if this kind of inference turns out to show up in many human minds – and in fact it is so common we have a name for it, "magical thinking" –, then perhaps it reflects the workings of *the* human unconscious mind, which is after all the object of psychoanalysis. Well, experimental research shows this kind of inference *does* reflect the workings of the human unconscious mind (Brakel, Kleinsorge, Snodgrass, Shevrin & Arbor, 2000). Grünbaum's example, therefore, *backfires*. Regarding thematic affinities as evidence for causal linkages is not in psychoanalysis as dubious as in 16th century alchemy. His reservation is not for human sciences as eloquent and striking as it is for biophysical sciences.

While I write these words, the humankind suffers with the spread of Coronavirus. Inasmuch as people became aware of the pandemic, there have emerged all sorts of wild and strange theories attempting to draw some sense out of its unexpected and morbid character, but one of them is particularly interesting for our current discussion: that pandemics have occurred every hundred years, in the 20th year of every century, for at least 4 centuries. The Plague occurred in 1720, the Cholera Outbreak in 1820, the Spanish flu in 1920, the Chinese Coronavirus in 2020. The idea that the pattern is inlaid in some causal network seems, at first sight, a bit mystical. What would the advent of powerful microbes have to do with a time-span of 100 complete orbits around the Sun? Our intuition is confirmed by the facts that "the plague cited wasn't a pandemic; two of the other examples didn't occur neatly in the years cited; and, most importantly, numerous other pandemics have occurred without such synchronicity" (Kertscher, 2020). From a psychological perspective, however, what is most important in this case is that human minds noticed a pattern and ascribed significance to it.

The reader should note that this is not a defence that human sciences such as psychology have an exceptional, radically unnatural object, hence that they must use exceptional methods. I am simply recognizing human sciences' *reflexivity*: the object of psychologists, the mind, is also one of their instruments of work. Reflexivity does not apply to today's pharmacology nor to 16th century alchemy: plants and livers are not concomitantly instruments to study plants and livers; plants and livers are not brains or minds. The reflexive condition of psychology and, by extension, of psychoanalysis makes their "magical thinking", the inference of causal links from thematic affinities, rational enough.

But not in all cases, which brings us to a second point regarding Grünbaum's scepticism. Not every thematic affinity between mental products is an index of causality. Fortuitous similarities are always possible in any domain of reality… The philosopher is accurate in claiming that, although mental causality may sometimes generate thematic affinities, there must be something more to authorize an inference of mental causality than just thematic affinities. Let us see him as a sphinx who poses the following riddles for us. How do clinical psychoanalysts identify (or should identify) insignificant coincidences in the patient's productions? How to tell when a thematic affinity in the patient's productions is reflecting a single causal process, is a part of one and the same causal network?

Inspired by Grünbaum's example of a person dreaming of a general specimen of house (see p. 27), let us imagine a patient who, on one session, tells his analyst of his dream about a dark room and, on another session, providing context to a certain episode, says to his analyst: "I was arriving at my friend's house, when…". Would a wise analyst infer, only from the patient's dream with a room (which is, the analyst would remind, part of a *house*) and from his report of arriving at his friend's *house*, that there must be some unconscious motive around his representation of "house"? Let us also imagine a patient who says "I love the wild" and, later that day, "my mom has green eyes". Would it be evidence that, given the wild has trees and trees are green, there is a causal connection between she loving the wild and her mother having green eyes? A lay person may feel like answering: "Of course not! What's more, if inferences like these are usual in clinical psychoanalysis, I understand why it is so discredited!". We could agree with the lay person but, if we do, we should seek to know what exactly would be wrong with these two inferences and their kinds.

Grünbaum (1989) brings to the fore an obscure "background knowledge" to be the something-more that authorizes an inference of causality when all we have is a mere thematic kinship. What is the nature of the background knowledge in such integer cases? It seems to be a reliable – although implicit, diffuse, intuitive, unconscious – awareness of how a certain empirical domain behaves mixed with a fundamental sense of probability. In the case of the student's plagiarism, the professor's background knowledge would be of the immense number of words in the English language and of the immense number of grammatical sentences it is possible to compose with these words, and of the extremely low probability that a text that similar to the encyclopaedia's were a result of the student's autochthonous critical effort. In the case of the mysterious attendance on the desolate beach, the tourist's background knowledge would be of the shapes that kind of wind possibly and usually sculptures by revolving that kind of sand, and of the extremely low probability that a cavity with the exact shape of a human foot were a result of a usual or even possible interaction between wind and sand. In the case of the dream with the Frank Lloyd Wright's house because of a visit to it the day before, the background knowledge would be of the immense number of shapes human houses can have and of the extremely low probability that the image of a house exactly like the one designed by Frank Lloyd Wright was fired at random in anyone's brain in a dream-state.

Hence, Grünbaum himself gives us some hints to find out when causal inferences from thematic affinities between mental products could be strong. Could his hints assist analysts, helping to improve their inferences?

But would analysts not already have in their boxes the tools estimate-of-how-certain-empirical-domains-behave and basic-sense-of-probability, using them constantly? Would the contrast Grünbaum erects between the three examples and, generally, the inferences in clinical psychoanalysis not be, in the end, a false contrast? Would, say, the inference that the Rat Man's rumination of the rat torture was causally related with the memory of his father punishing him for having bitten his nurse like a rat when the patient was a kid not be based on the *characteristics* of the emerging affinities, just like in the three examples – rather than merely on the affinities as such? Would such characteristics have something to do with what we have hypothesized about the control-instances conceived by analysts? After all, control-instances serve to check if some occurrences that seem extraordinary are really so, to compare probabilities…

Grünbaum's examples seem to backfire once more. I defend that this is the case indeed, that the background knowledge analysts activate when inferring about patients, and the Rat Man's Freud is among them, is of the same sort Grünbaum speaks of in his exemplary trinity. The activity of this background knowledge can be deduced from the fact that analysts stick to thematic kinships with a high degree of certain traits. What would these traits be?

In revolving around the insights of Dalbiez (1941), Bucci (1988) and Michael (2008, 2015), among others, I have arrived at this pair of features an analyst should be able to identify along a plethora of mental products in order to decide whether they are sharing a relatively constricted[47] causal network:

a) Theme repetitiveness and/or;

b) Theme specificity.

Even though we intuit what specificity means, it is hard to define the concept without breaking it into other, maybe plainer, concepts. Thus, I understand "theme specificity" as:

b.1.) theme rarity and/or;

b.2.) theme distinctiveness and/or;

b.3.) theme complexity.

Here is the formula that rational analysts would use to infer the *mere existence* of a causal linkage between mental products: *the more repetitive and specific a theme across some of the patient's mental products is, the higher the probability that these mental products are part of one and the same causal network or, in other words, that they are closely influencing each other in the patient's psyche*.

Before discussing this formula, I should minimally define its concepts. A *theme* is understood here as a certain group of elements, qualities and structures present within a mental product. A *mental product* would be the conjunction of something the subject did – to move one's body, to have an idea, emotion, dream, sensation, to form a perception, an intuition, an understanding, a memory, a fantasy, to speak, to make a joke, to commit a slip, etc. – with the circumstance or context of this action or response – that is, everything that does not depend on the subject to occur and is concomitant to some of his or her actions or responses, which

[47]From a deterministic perspective, everything in the universe has some causal connection with everything else in the universe. "At the very least, any correlated or uncorrelated collection of events shares the Big Bang as its common cause" (Tucker, 2004, p. 104). If we take the mind of a particular person, every one of his/her mental products is causally related to each other. Talk about mental causality implies a relatively small "causal area", delimited according to the explanatory demands at issue.

includes times, places, persons, intersubjective performances, etc.. Inside a psychological domain, a *thematic affinity* could be defined as an emergence, across two or more mental products, of a same group of elements, qualities and/or structures; the elements, qualities and/or structures deemed identical across mental products must include at least bits of both the response and circumstance poles of which mental products are composed.

A theme in a patient's mental products is *repetitive* if it emerges more than once in the patient's mental life. Thus, the concept of thematic affinity is intimately related to the concept of repetitiveness. The latter concept is essential, though, because it makes explicit that the degree of repetitiveness matters in a thematic affinity. Even though a theme recurring two and a theme recurring seven times are both constituting cases of thematic affinity, one is very different from the other as regards degree of recurrence.

The concept of specificity, I repeat, is simpler to ascribe than to define and, to escape its intangibility, we should see it as a composite concept. A theme in the patient's mental products is *specific* if it is rare, distinctive and/or complex.

A theme is *rare* if the probability of its appearance in the mental life of a patient is low given the culture the patient belongs to. Let us devise three dreams dreamt by hypothetical 21st century, Western patients. A patient *x* has a dream in which the word "serendipity" comes up. A patient *y* dreams that, of seven wolves, six were white and one was black. A patient *z* dreams about a house designed by Frank Lloyd Wright, describing the house's rich amalgam of features. All of these themes would call an analyst's attention for being considerably "specific"; and, be the analyst conscious of this or not, they would be seen as "specific" due to the fact that they are rare in "the culture". Conversely, it would be relatively commonplace, thus unspecific, if patient *x* dreamt about breakfast food, patient *y* about a black smartphone, and patient *z* about a house with a mainstream design. An analyst would hardly elect as a behaviour calling for an explanation (unless it has been highly repetitive) a casual complaint about the traffic coming from a patient who lives in the city of São Paulo, for such a theme would lack a high level of specificity in the sense of cultural rarity.

A theme is *distinctive* if it is enacted in one circumstance but not (or rather than) in circumstances otherwise similar to that one. In other words, a theme is distinctive when it results from a discrimination or differential treatment between apparently equal circumstances. Distinctiveness could be expressed by the following formula: all things being equal, a subject *s* should have dealt with the object/situation *o* by acting/performing *a*, but, standing before *o*, *s* performs *b* instead (so, not all things are equal).

A patient names all the members of her family by their kinship relations to her, except her mother, whom she calls by the mother's proper name; in this case, the theme mother-name would have a considerable level of distinctiveness, thus of specificity. A patient has panic attacks in certain public places, but not in others; an analyst would want to investigate both the pathological and the non-pathological mental products and would ask what is the difference between those places and these. A patient has outbreaks of violence every Christmas holiday; what would the holiday have of special in his inner life? A notion of distinctiveness is also operating in the very mundane episode of having somebody we sexually desire treating us kindly: in order to track a glorious reciprocity, we ask ourselves if he/she treats this kindly *everybody* or *just us*. Besides being relative to a subject's overall mental life, distinctiveness is also culture-relative – for example, a subject not greeting anyone when showing up at social events, even though it is a common habit in his culture.

Every distinctive theme, be it subject-relative or culture-relative, is also rare (in the cultural sense defined above). Not every rare theme is distinctive, though. The serendipity-dream of patient *x* is an example of a case where the theme is rare but not distinctive: it is not expected that patient *x* dream about another well-defined stuff rather than about "serendipity" (however, it *is* expected that a certain subject greet someone rather than no one in social events). The two notions may be confused in extension (domain composition) but not in intension (meaning); rarity is a quantitative notion while distinctiveness is a qualitative one.

Finally, a theme is *complex* if it is a numerous combination of diverse elements, qualities and structures. Such a combination can only be numerous relatively to the individual mind or the culture in which it arises; like distinctiveness, the property of complexity is identified out of a comparison between the subject's mental products and the usual mental products of his/her culture, or between the subject's mental products themselves. A patient theme is complex if it is *more complex than* the usual themes of his/her mental life or of his/her culture. Since childhood, a patient often dreams she is "falling into a void", but in the last year the dream changes: while falling into a void, she suddenly grows angel's wings and flies over a macabre city, a "city of death". The dream is now more complex than before. In a culture where elephants are usually depicted as they are in reality, grey and on the ground, a patient hallucinates a flying, pink elephant, which he claims to be "the light of the world". Because such an elephant is a relatively complex figure, a noteworthy profusion of features, it would be a mental product with considerable specificity.

The presence of complexity in a theme, as will be argued for in a moment, is a strong index of the theme's motivational charge. The Brazilian comedy group *Porta dos Fundos* produced a sketch whose humour is provoked by this very idea (Porta dos Fundos, 2012). In the fictional video, a pregnant heterosexual couple discusses possible names for the upcoming baby. The mother starts to suggest male names; the father, then, starts to tell her what each male name evokes in him, depicting in detail the personality, the looks and other attributes an adult man with that name would, for him, possess. The depictions are so detailed, so specific, that they become erotic; the future mother expresses unease with the secret homoerotic proclivity of her husband and invites him to discuss female names instead. The female names, however, evoke in the husband only dull impressions like “a nice person” or “cool”...

If a theme is complex, it is inevitably distinctive, thus also inevitably rare. With patient *z* above, who dreamt about a house designed by Frank Lloyd Wright and, by telling it, described the house’s features, we have a theme that is complex (a numerous combination of diverse elements, qualities and structures), distinctive (the dreamt house is more complex than the houses once dreamt by patient *z* or by other persons of his culture), and, of course, culturally rare. However, not every distinctive theme is complex. The patient having panic attacks in certain public places rather than in others is an example of a distinctive but (as it is) uncomplex theme. Once again, these notions are partially sharing extension, but not intension: if one lists every distinctive theme, one lists also every complex theme, but understanding the concept of distinctiveness is not the same as understanding the concept of complexity.

A thematic affinity may be specific but unrepetitive, and vice-versa. The angel-shape of a cloud in the sky may be stunningly precise, but one knows it will vanish soon to never be seen again. Someone born on the 1st of February may find the whole of his traits matching again and again what is supposedly emblematic about Aquarius people, but, as many sceptics have noted, astrological discourse is always articulating concepts that are too ambiguous and/or too universal – in other words, too unspecific. Repetitive but unspecific behaviours are, for example, thinking in the morning that one must get up and get some tasks done, blinking (unless one blinks in the rhythm of, say, the Summer’s Third Movement of Vivaldi’s Four Seasons) and brushing teeth (one’s own, of course). A specific but unrepetitive behaviour is, for example, any thought that came once – say, when one was 8, at school, on an April morning – and appeared no more. A set of instances where a theme is repetitive but unspecific constitute a dull similarity and tend to be epistemically insignificant. A set of instances where a theme is peculiar but isolated tend to be scientifically unimportant. On the other hand, instances where a theme is *both specific and repetitive* are telling and captivating. The inferential adventure of the analyst starts with them.

The cogent analyst does not infer that there is a single causal chain within a thematic affinity only from the existence of the thematic affinity, as Grünbaum claims. In the black box of such an analyst, whether a thematic affinity constitutes the trace of a single causal mechanism at work is a matter of probability. The probability that a thematic affinity be the result of a single causal mechanism grows along with the degree of specificity of the theme in question and the number of this theme's recurrences. From another angle, the probability that a thematic affinity be just a trivial accident involving many independent causal chains would be *lower* inasmuch as the theme in question is more complex, distinct or rare and inasmuch as its recurrence is higher[48]. In other terms: the best explanation for a highly repetitive and specific thematic affinity is that it comprises a single causal mechanism.

As non-delirious persons, we are sure that an actual angel is not causally connected to the angel-shape of a cloud – but we would not be so sure of it were there many clouds in the sky with the same stunningly precise angel-shape (as far as we know, an extremely improbable event). If the astrological charts we find so omniscient were full of unusual and rich details – if their forecasts were not something like "[according to your astrological signs,] you might suffer from conflicts between your professional and your love life" but rather something like "every day of this week it will take you a long time to fall asleep at night because you will think there are still many ideas you want to include in your doctoral thesis" – we could perhaps start to see astrology as a science, but that is not the case.

One little cavity shaped like a human foot is something specific enough: with its bunch of features, it looks like nothing, unless like a human foot. On the same beach, many cavities shaped like a human foot multiply such specificity. In such a case, we do not even have to think clearly about the dances the wind is able to perform with the sand to infer that there have been a "wingless, two-footed, flat-fingernailed animal"[49] walking there[50]. The probability of any other explanation is too low to even be considered. Likewise, sexual-satisfaction-produces-rat-related-death-of-father is not exactly a thought a young man is expected to have while walking

[48]According to our formulas, a repetitive specificity is likely indicating a single causal chain, but it does not follow from this that a single causal chain is always likely to manifest itself as a repetitive specificity. A single causal chain in a human mind can manifest itself as an unspecific and/or isolated behaviour. An unconscious motivation of the kind relevant to clinical psychoanalysis, however, never manifests itself as such. The *non sequitur*, therefore, should not be a matter of concern for us here.

[49]A reference to Plato's definition of a human (Plato, 1997, p. 1684).

[50]Would uncanny "footprints" of what appears to be a two-footed, *pointed*-fingernailed animal make us speculate further over its cause or would it just make us wish Jesus was carrying us away from the beach?

in the *Tiergarten* on a Sunday morning; if there are many versions of such a complex and painful idea in a patient's life, any clinical psychologist is surely warranted in concluding that all the versions are causally connected to each other.

Following Grünbaum, I have been presenting examples that are both psychological and non-psychological in order to discuss when it is legitimate to infer a causal mechanism from a thematic kinship. And, indeed, the background-knowledge and the repetitive-specificity formulas stated above, which complement each other, are valid for any object of reality – they have a logical force that is independent of any physical or psychological theory. We should recognize, however, that the repetitive-specificity formula is even more forceful when we deal with mental entities – of course, given a certain theory of motivation.

The body of a baby signalizes a need of nourishment by producing a tension, which is poorly discharged in a cry. Whenever the baby bursts out crying, her mother comes and breastfeeds her while looking at her face. The baby associates the mother's *specific* visage and smell with the pleasure of having hunger alleviated – with having the tension dissipated. The sight and the smell of the mother starts to be enough to relax the baby at any moment. The baby's motive for crying starts to be the faint pleasure-representation attached to the faint mother-representation and the longing to make both representations vigorous, sharp and stable through the nearness of the real mother. For Freud, motives are constituted when representations with instinctual relevance are associated to each other through their *specific* features. The child knows it is not any person that has been able to save her from feelings of death: it is *that* person, her mother. Freud has also theorized that an association is an opening of a neural route and that, once an association is established, the mental energy tends to flow through it, thus activating it. The motive becomes, therefore, *repetitive*. Psychoanalytic theory maintains that human motives are idiosyncratic and insistent *by nature*, as we can confirm through Mackay's (1989) competent summary of the Father's Theory of Motivation:

> [For Freud,] The basic form of instinctual activity is as described: somatic needs give rise to mental tension and a pressure to discharge. […]
> Behaviour is precipitated […] when external stimuli impinge on the system and are matched with the wishes current in the system. Where a genuine match is achieved, specific actions can take place. This means that certain internal and external stimulus conditions have to obtain for motivated action to take place. In this sense, endogenous and exogenous stimuli precipitate or "cue" behaviour, and these factors, which precipitate activity, are dependent on learning.
> The basic direction of all behaviour is determined by the tendency to discharge tension. […] this means that the pleasure principle is the primary director of behaviour. A slight specificity of direction comes about from the source of the particular energy pressing for discharge. However, the objects of specific instincts (for example, sex) have to be learned. The major modification to the primary direction comes from the secondary process, which is responsible for the vicissitudes the instinctual energy can undergo. The secondary process is the inhibition of the immediate discharge of instinctual energy. The redirection of the upheld energy takes place in

> accordance with learned ways of achieving experiences of satisfaction (lasting discharges of tension). In turn, the learned ways depend on environmental conditions. The structures involved in this directing are complex. However, they include the network of associated memory traces. The patterns of this network aids in giving direction to energy passing through it (pp. 75-76).

Besides corroborating the repetitive-specificity formula, the psychoanalytic theory – its core – has yet another role in the first step of the process of inferring particular hypotheses in clinical psychoanalysis: the role of establishing which points in the flux of data that patients display to analysts are the most telling, are the ones most likely to reveal repetitive specificities. If we follow the formula, we conclude that all repetitive specificities should be epistemically significant. In practice, though, analysts tend to note only some of them, the ones showing themselves in certain dimensions of mental life. When observing a patient, analysts tend to pay more attention to general themes such as family, sexuality, gender, childhood, anxiety, work, power, anger, etc.. This is because, *according to the theory*, such general themes have more chances than others of being charged with unconscious motives – thus, of bearing idiosyncratic and insistent themes.

This discrimination of general themes is a bias, but a harmless one. We must recognize that thorough free-floating attention does not happen in clinical psychoanalysis, but this does not mean that core hypotheses tell analysts *which* of the general themes (family, gender, work, etc.) is central to a patient's unconscious, much less specifically *how* and *why* one of the general themes would be central to a patient's unconscious. By the way, this is similar to what happens in any scientific research. Any scientist conducts research having in hand a theory that indicates a list of possibly relevant variables relative to the object of interest, a list of variables to be registered and analysed; we do not say, in such cases, that scientists are begging the question, for the theoretical constraint at stake does not prevent data from surprising them. Analysts indeed "raise ears and eyebrow" when a patient speaks about mother, but this does not mean that they know beforehand that the patient's mother is one of her central sources of conflicts and fragile compromises, much less how and why her mother would be one of such sources.

We have seen that the first step in the process of inferring particular hypotheses is inferring that mental products with a certain repetitive and specific theme are *somehow* causally linked to each other. We are able, now, to list the sub-steps inlaid in this first step. Initially, analysts observe the patients thoroughly and listen to everything the latter have to say; then, analysts focus their attention on the patients' behaviours that express the general themes the psychoanalytic theory considers potentially dramatic for anyone's mental life (sexuality, childhood, power, etc.); third, among such behaviours, or mental products, analysts note a specific theme that has recurred, in other words, a minimally repetitive and specific thematic affinity; then, with more data coming from the patients, analysts amplify and specify further this same thematic affinity; finally, analysts realize that the repetitive and specific thematic affinity is a very improbable one, and conclude that the mental products bearing it are somehow causally connected to each other.

Now let us finally discuss the second step: the inference of *how and why* such mental products are causally linked to one another. It is composed of two interlaced tasks: the inference of whether or not a yet unknown mental product should be the common cause of the thematic affinity and the inference of the mental schema causally linking one mental product to the other.

In this step, core hypotheses have an even more prominent role than before. In the first step, core hypotheses tell the analysts which of the patients' themes are likely to show clinical relevance. In the second, core hypotheses instruct analysts when there must be a yet unknown common cause at play and how to come up with a pool of hypotheses over the mental schema in question. Every schema-hypothesis of the pool should not contradict the data that has become available, and should be able to explain the highly repetitive and specific thematic affinity that has been revealed; moreover, the contents of such explanatory hypotheses should also cohere with the core of psychoanalytic theory. In sum, particular hypotheses in clinical psychoanalysis are both fostered and limited by both data and core hypotheses.

This role of core hypotheses in the inference of particular ones is fortunately no novelty in the psychoanalytic literature; but it is usually expressed in a less technical and detailed mode. See, for instance, Mackay (1989) commenting on a passage of the Wolf Man case – where Freud (1955b) affirms that "there are in any case narrow limits to what a psycho-analysis is called upon to explain" (p. 105) – and discussing the explanatory role of the "theoretical account" in the genesis of the "clinical account":

> By "a psycho-analysis", in contrast to simply "psycho-analysis", Freud means the "overall clinical explanation". Within this he is identifying two elements. One is what I have termed the "clinical account" […]. It simply states the phobia data and unconscious wishes that lie behind it. The second I have termed the "theoretical account". This consists of more general psychoanalytic propositions. It is not derived from the clinical data of this particular case, and there is no attempt to justify it within the case study. Rather it is brought from without the case study and its descriptions of mental workings serve as explanations of the clinical account.
> […] the theoretical account […] *explains* the […] clinical […] account by *describing* the "psychical mechanisms and instinctual processes" that underlie the unconscious wishes mentioned in the clinical account […] (p. 201).

Some pages later, he writes:

> Freud considers it natural that the clinical should be subsumed by the theoretical account, and indeed that the two comprise a single explanation. […]
> […] I have been pointing out the heuristic role that the metapsychology performs in clinical accounts.

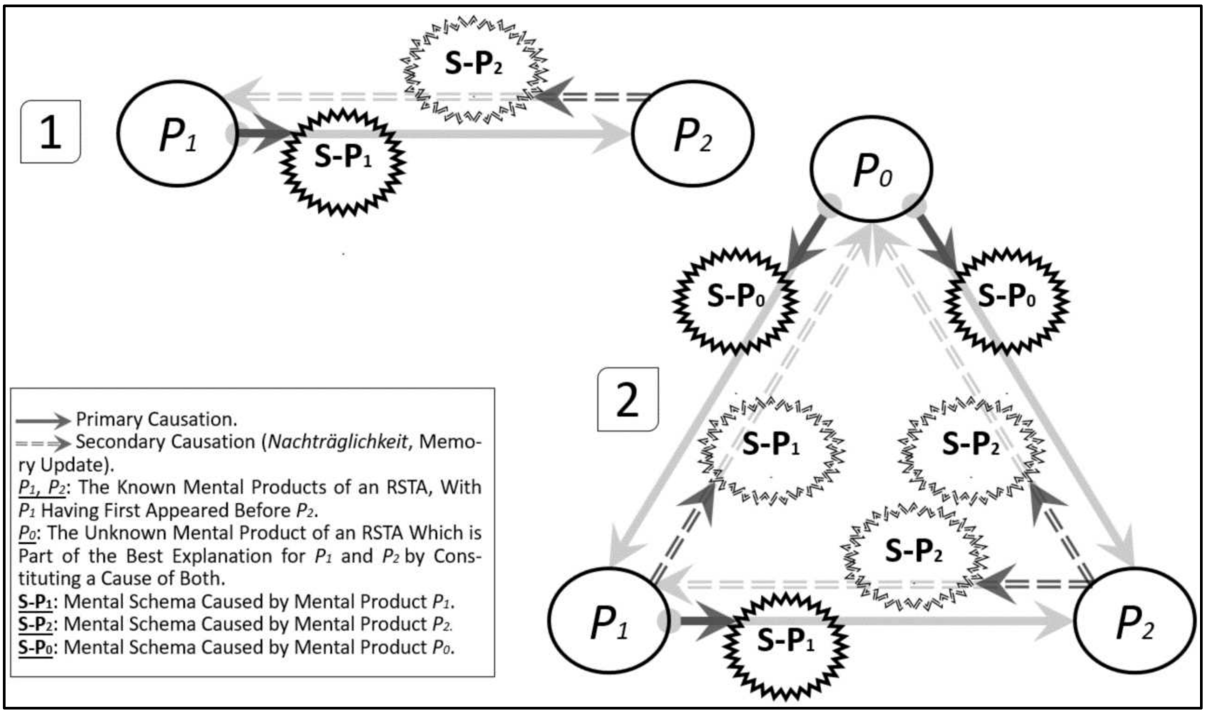
> The view that the clinical principles are in fact independent of the metapsychology is then wrong on at least two counts. One is that many of the clinical principles are conditioned by metapsychological principles that predate them and the stage is set for individual case explanations by these metapsychological assumptions. The second is that as a matter of fact Freud depends on his metapsychological model of mind when he forms his clinical accounts in the case studies (p. 205).

Let us go on discussing the tasks of the second step.

Analysts recur to core hypotheses to decide if the mental products of a highly unlikely thematic affinity are enough to explain it or if it is best explained by the influence of a yet-to-be-known mental product – of a common factor that has primarily caused and keeps causing the mental products that are known. Something would be missing if, among the mental products of the highly unlikely thematic affinity about which a patient allowed an analyst to learn, none of them were an early-childhood affectively-charged more-or-less-prolonged event.

Let us suppose there is a same highly specific theme in mental products P_1 and P_2 – a Repetitive and Specific Thematic Affinity (an RSTA, from now on) – given that P_1 has first emerged before P_2; and let us remember that any kind of mental product at some point becomes a memory, and that every memory is subject to constant and involuntary updates along life (see pp. 99-101). Either P_1 has caused P_2 (and the memory of P_1 keeps causing the memory of P_2), and P_2 has updated (caused) the memory of P_1 (and the memory of P_2 keeps updating/causing the memory of P_1) – or a yet-to-be-known mental product that had emerged before P_1 and P_2, P_0, has caused and keeps causing both (and also P_1 and P_2 have been influencing each other and P_0). In other words, either the hypothesis that the mental products in question have been causing each other best explains (along with the pertinent mental schema and triggering circumstances) the RSTA, or what best explains it is the hypothesis that there exists a not-yet-revealed mental product that has primarily caused, and keeps causing, the ones that were revealed – a mental product that is their common cause. The figure below summarizes the two hypotheses.

Figure 8 - The Two Possible Explanatory Hypotheses Concerning the Existence of a Yet-to-be-known Common Cause of a Thematic Affinity P_1-P_2



(1) There is no withheld common cause: P_2 is well explained by P_1, **S-P₁** and the circumstances in which P_2 occurred, and P_1 primarily caused P_2;
(2) There is a withheld common cause: P_2 is not well explained by P_1 alone, P_1 calls for an explanation as well and both are well-explained by the hypothesis of them having been primarily caused by a prior mental product, P_0.

Source: Elaborated by the author (2021).

In order to decide which alternative is the more probable, analysts revisit a simple idea of psychoanalytic theory: that a great deal of the contours of a subject's personality, which includes the manner the subject may pathologically suffer along its development, are fixed in early childhood. So, the "math" is also simple: if in the RSTA a memory attributable to early childhood is included, we should accept hypothesis 1; if not, we should accept hypothesis 2. Of course, if a psychoanalyst (or a neuropsychoanalyst) said that that simple idea must be refined, I would agree; but my aim here is just to demonstrate how core hypothesis and data work together to determine this inferential task of our second step.

Finally, in order to infer the *mental schema* that is conditioning the patient's RSTA, in parallel to consulting the core hypotheses of psychoanalytic theory, analysts apply, at least implicitly and/or unrigorously, the Method of Difference canonized by the work of John Stuart Mill. In truth, the Method has already been applied in the first step. The features of repetitiveness, rarity, distinctiveness and complexity are not absolute; they are relative to the mental life of the patient as a whole (internal comparison) and/or to the normal behaviour, in the statistical sense, of the individuals that are presumably part of the patient's culture (external comparison). If a theme in a cluster of mental products of a patient is repetitive and specific in comparison to the themes of all the other (reported) mental products of the patient, the analyst can conclude that the factor making the former cluster different from the rest is the activeness of a single causal mechanism. It is still an ill-defined difference, but not a trivial one.

To go forward, analysts have to be attentive to the differences, not only between the acts present in the two groupings (the RSTA and the group of control-instances), but also between the circumstances of such acts. Then, analysts ask themselves *contrastive questions*. Why has the patient repeatedly done *this in this circumstance* rather than (something in the interval of) *what is expected in this circumstance*, according to common-sense? Why has the patient repeatedly done this *in this context* rather than *in the other contexts of which he/she has participated*? The idea that inferences are framed by contrastive questions such as these comes from Lipton (2004) and, by the end of this chapter, we shall see which place they occupy in his account of inference and explanation.

Through high recurrence, analysts are presented with various instances or versions of the specific theme. These different versions, if seen through the lenses of core hypotheses, inspire analysts to generate a pool of probable answers to the rather-than questions above, that is, a pool of good explanations for the RSTA. These answers or explanations are descriptions of mental schemas.

The patient presents more data and the RSTA is thereby more amplified and more specified. Each time it happens, each explanatory hypothesis (on the mental schema) that was devised becomes either more or less probable in the mind of analysts – that is, the degree to which each explanatory hypothesis is probable grows or decreases in the mind of analysts as new data keeps emerging. There may come one version of the theme, one bringing to light details of the theme that could not be noted before, which allows analysts to think of the likeliness of just one of the hypotheses as extremely increased and the likeliness of all the other hypotheses as extremely decreased. We shall call it "the smoking gun", following the soon-to-be-discussed work of Carol Cleland (2001, 2002, 2011). The smoking gun is *usually* a primary cause, a childhood memory related to intense emotions.

Hence, by keeping an eye on what the patient tells and does and the other on the explanatory hypotheses that have been devised, the mental schema and the primary cause that best explains the RSTA are found. The patient's unconscious motivations are no longer in the shadows. Or at least *part of them* is no longer in the shadows – patients are always proposing new questions to be solved by their analytic dyad. Thus, the whole process starts again, but more informed than before – in the manner of a spiral.

The process of the second step, I suppose, shall be better understood through the extended example below.

Let us reasonably state that one of the core ideas of psychoanalytic theory is that the diverse concepts the Ancient Greeks had for what we today call "love" – Eros, Storge, Agape, Philia, etc. –, a diversity that we presume in our current use of the word, would have the same psychological root, being thus continuous concepts. The idea would be present in the famous psychoanalytic hypothesis that there is a sexual dynamic in all of our loving impulses – that the latter are but a transformation of primary sexual impulses. To be rigorous, the hypothesis states that sexuality must be understood in a broader sense, that its concept cannot be reduced to the domain of genitals and impregnation, but to the domain of general bodily pleasure (Freud, 1953b, 1963b).

An akin core hypothesis of psychoanalytic theory – an idea stated with variable formality throughout Freud's texts but undeniably present in the mind of every clinical psychoanalyst – is that the relationship modes we have had with our parents during our childhood determine the ones we come to have in maturity, especially the ones we come to have with spouses and liaisons (Freud, 1961a, 1961b, 1961c).

This "Oedipal" hypothesis can be related to the preceding one on sexuality with the aid of some Darwinian ideas[51]. If along childhood we humans[52] can only survive by being nurtured by an adult of our species, then the fact that we feel pleasure with the experience of having our needs met and correlate such pleasure with the features of the adult figures nurturing us enhances the chances of us surviving until reproductive age (which enables us to transmit such

---

[51]As we learned from many authors (e. g.: Sulloway, 1979; Wallace, 1985; Ritvo, 1990; Simanke, 2009; Ferretti, 2014), the Freudian theory has Darwinian (as well as Lamarckian) premises. In the words of Wallace (1985):

> Darwin's influence on Freud is by now widely known and was frequently acknowledged by Freud himself. By the end of medical school Freud has acquired Darwin's major books; indeed, Freud (1925, p. 8) had counted Darwin's work as one of his reasons for choosing a medical career. In *The Expression of the Emotions* Darwin applied the principle of atavism to the explanation of human behaviours, group and individual. Emotional expressions were conceptualized as ontogenetic and phylogenetic survivals – explained, in other words, by reference to the history of the individual and the race. Freud too would come to view neurosis (ontogenetic *and* phylogenetic) atavism and he (Breuer & Freud, 1895, p. 181) explicitly cited Darwin in support of his own journey into the patient's past to discern the meaning of the apparently senseless current neurotic symptom (pp. 139-140).

[52]And all mammals in general.

attachment-predisposition to our progeny, just like, in virtue of the same process, it was transmitted to us). We are destined, then, to have the movement we feel from unpleasure to pleasure – in Freudian terms, from the charging to the discharging of bodily energy – interwoven with the tender representations we usually have of our caregivers. That movement being the very prototype of orgasm and erotic love, a tendency to look for sexual partners having features similar to one's parents' features should not come as a surprise, nor should it come as a surprise that a dynamic of nurturing and providing is easily found in erotic relationships[53].

Almost in the end of this theoretical diversion – an important one as we will see in a moment, – we extend the Darwinian tone to the relation between the child jealously present in the hypothesis of the Oedipus Complex and the widely recognized phenomenon of erotic jealously in adulthood. One of the consequences of the Darwinian theory is that any one of us living beings should (at least frantically and temporarily *try to*) act in order to prevent us from losing contact with the "objects" that keep us alive and healthy, which includes harming the "objects" that disturb such contact. This should happen from our babyhood to our infirmity – and, as we saw, at any point of our lifetime all of those nurturing "objects" are destined to connect to each other in the mental realm. Adult jealously can regularly emerge, thus, with a childish colour.

Let us now imagine a patient that communicates two of her mental products to her analyst. They would be:

P_x) A "jealously"[54] episode with her husband, whose trigger was: her husband complimented a female friend of them, who turns out to be "successful" in her job.

P_{x+1}) A "jealously" experience with her boss, whose trigger was: her boss told her she should be trained by a female co-worker because the latter would have more expertise than her in a certain area.

At this point, the core hypotheses we have just stated (among others) have already had a role in turning the analyst's attention to some general themes (sexuality, work, aggression) than to others (say, the weather). They have helped the analyst to identify the promising

[53] For a long time in the history of the human species, the quantity and quality of individuals' erotic relationships defined their chances of keeping their integrity and health, and satisfactory erotic relationships would still today have a (retroactively ascribed) protective function were it not the advantages of our post-industrial civilization. In harsher times, an adult without partners could be as vulnerable as a child without parents. That is why we may still have a vestigial tendency to look for partners who *seem* able to take care of our bodily needs. The theoretical path of this footnote may be too specifically Darwinian to be present in the inferential process of *most* psychoanalysts, but *some* psychoanalysts might consider it in their interpretations.

[54] The quotation marks are meant to indicate that the word was chosen by the patient, being thus part of her lay or commonsensical interpretation of the events.

thematic affinity above. Here is the reasonable bias we have discussed: it does not imply a *petitio principii*, as the analyst still has to wait for further data to be able to know which specific forms the general themes will take, how specific the forms will reveal themselves, and even which of them will reveal themselves indeed recurrent in the patient's mind.

Such core hypotheses, besides, would still lead the analyst along the inference of the exact nature of the causal linkage the thematic affinity above – let us suppose – likely indicates.

Our reasonable analyst would speculate about the existence of a yet-to-be-known mental product that is the common cause of P_x and P_{x+1} by privately asking himself: what is the best explanation for the thematic affinity, the picture where P_x and P_{x+1} influence each other or the picture where P_x and P_{x+1} are both primarily influenced by a third mental product, P_0? Consulting P_x and P_{x+1} and the core hypotheses on sexuality and the Oedipus Complex, the analyst would seriously entertain the last alternative. He would thus be presuming that *some* memory related to the early affective development of the patient, a memory regarding her caregivers, is the common cause of the mental products P_x and P_{x+1}. The analyst would not yet know the specific aspects of this memory or even if the patient would someday have it conscious and let him know about it – but, as long as the core hypotheses in focus are not refuted (and below we shall see how they can be), the analyst would anyway be presuming that such memory exists.

On the way to unravel the mental schema that has been connecting P_0, P_x and P_{x+1}, the analyst would privately ask himself some "rather-than questions". Why "jealously" of *husband and boss* and not of other similar persons in the patient's life? Why *"jealously"* of boss rather than fear, respect, hate, or indifference? Why "jealously" *triggered by successful women* rather than by joyful, hilarious or clever women (or men)? Why *"jealously"* triggered by successful women rather than admiration, envy, or indifference? And so on.

While expecting the common-cause memory to be revealed by the patient, the analyst would elaborate in his mind an incipient list of candidate hypotheses concerning the mental schema at work in the patient. In the case we are imagining, the analyst could elaborate the following list[55]:

S-P_0-1) If feeling *helpless*, then becomes "jealous".

S-P_0-2) If representing *herself* as *"inferior"*, then becomes "jealous".

[55]Two important points regarding this list: it is not intended to be exhaustive and it theoretically admits the possibility that more than one schema is operating at the same time (many combinations between the hypotheses are theoretically admissible).

S-P_0-3) If having as *caregiver* a *man*, then “jealously”.
S-P_0-4) If detecting a possible *competitor* in *any “successful woman”*, then becomes “jealous”.
S-P_0-5) If “male” figure can respond to “jealously” with “understanding”, then it becomes a “worthy” caregiver.
S-P_0-6) If “female” figure is “successful”, then it is an “unworthy” person.
S-P_0-7) If “male” figure is “promiscuous”, then it becomes desirable.

The analyst “cooks” this variety of hypotheses by mixing three ingredients: relevant clinical data, a good deal of psychoanalytic core hypotheses and a pinch of non-psychoanalytic common-sense. His cooking instruments would be rather-than questions like the ones already exemplified as well as a rough combinatorial analysis. Core hypotheses and common-sense constitute background information that allows the analyst to formulate candidate hypotheses which are minimally probable and senseful when they pop in the analyst’s mind. In other words, not every imaginable hypothesis deriving from the data is considered by an analyst – only the minimally probable and senseful ones.

However, the truthfulness of such candidate hypotheses would still be under scrutiny. The rest of the inference on the patient’s unconscious motivation would only be possible with forthcoming clinical data: the information the patient keeps presenting in the treatment would force the analyst’s reason to discredit part of these hypotheses and to identify the one that constitutes the best explanation for the patient’s mental products. With the patient firing more and more into the analyst’s ear, there may come a “smoking gun”. Let us see how this process could materialize in our example.

While expecting a childhood memory or fantasy involving caregivers, our analyst could go silently wild by thinking: “I can envision ‘daddy issues’ here, for sure, and the boyfriend and boss as father figures, and sexual interest in the boss”, etc.. If the patient fortuitously told the analyst that, during her childhood, her father was the most caring, that her father and mother were both successful engineers, that she remembers herself weeping when they went out for work, and that this was probably her being jealous of her father, then the analyst could claim to have that private elaboration corroborated – an elaboration presuming and enriching at least S-P_0-3. At the same time, the analyst would have probably found in this memory a common cause to P_x and P_{x+1} and would entertain S-P_0-4 with care. By identifying both a mental product as the primary cause of the patient’s relevant clinical phenomena and a generalization (a schema) the patient unconsciously fabricated from this mental product, the analyst has a good explanation in hand.

But let us imagine that, instead, the patient says to the analyst that she was raised by a single mother (a phrase she has probably forged a long time ago to describe herself). The analyst would have to be wary with the wild inferences, then; her boss and boyfriend may well have taken the place of her father in her mind, but, if so, this place would be, as far as the analyst knows, largely phantasmatic; the picture then would become very complex, thus suspicious, insofar as a virtuous explanation is the simplest possible. The patient could also further say to the analyst that, in her adolescence, near the end of high school, she has had several episodes of jealously with her female friends. So, we have:

P_{x-1}) A memory: "In my adolescence, near the end of high school, I've had several episodes of jealously with my female friends".

P_{x-y}) An identity statement: "I was raised by a single mother".

The analyst could, then, stick to S-P_0-1 alone, thinking that maybe the patient was suffering from a kind of undiscriminated separation anxiety. This would be somewhat wider than "I will lose my male caregiver to a successful woman". But what would S-P_0-1 have to do with the single mother? Where would the common cause be? Finally, there come smoking guns:

P_{x+3}) "Yesterday, when I was going to the supermarket, I saw a little boy being almost hit by a car because his mother was typing [on a smartphone]. Can you imagine this? I got so angry!".

P_{x-3}) "My mother was very busy with her work when I was a kid; she didn't have any time for me. I don't want to be like her when I become a mother myself".

Although the existence of S-P_0-1 would still be highly probable, the analyst could come with a better partial explanation than such schema: S-P_0-6. What seemed to the analyst fear of losing male providers to nice women becomes a moral criticism of successful women (and of men who admire them); what seemed pure jealously is revealed as a disparaging, and maybe envious, attitude. P_{x-3} reveals itself as P_0. S-P_0-6 could be expanded: the patient connects successful women to lack of motherly care, to a wound of her childhood. What the patient was naming "jealously" toward those two men was not a love-letter or a memo stating "I cannot lose you for her", but rather "The fact that you admire her is incomprehensible and outrageous". The analyst may predict that more mental products consistent with S-P_0-6 will emerge and have his predictions confirmed. He may come to know from the patient that the women of the instances P_x and P_{x+1} have kids and, in her fantasy at least, little time to look after them, etc..

The reader may have noted that the "smoking guns", the crucial evidence enabling the analyst to reach the best explanation for all the relevant evidence he had in hand (P_{x+3} and P_{x-3}, in other words, P_6 and P_0), is at bottom just another recurrence, but one that specifies the theme in a way that no other instance did until then; it "cuts out" what is essential in the other instances, showing what the actual recurrent theme was all along.

Perhaps this example is as artificial as other invented instrumental examples in philosophy: perhaps it evoke events that are hard to imagine happening in real life[56] and that cohere too neatly with the previous theorizations. The reader may otherwise charge the example with an opposite misstep: that, in it, the order and the manner in which the analyst considers and concludes things are not that clear, that the model of inference it displays presents too many indeterminacies. I would understand both accusations. With the example, I indeed intend to show how complex and sometimes chaotic a clinical-psychoanalytic inference, an *ex post facto* inference in general, can be, and, at the same time, to show that this kind of inference is not random or whimsical, or at least that it has the potential not to be so.

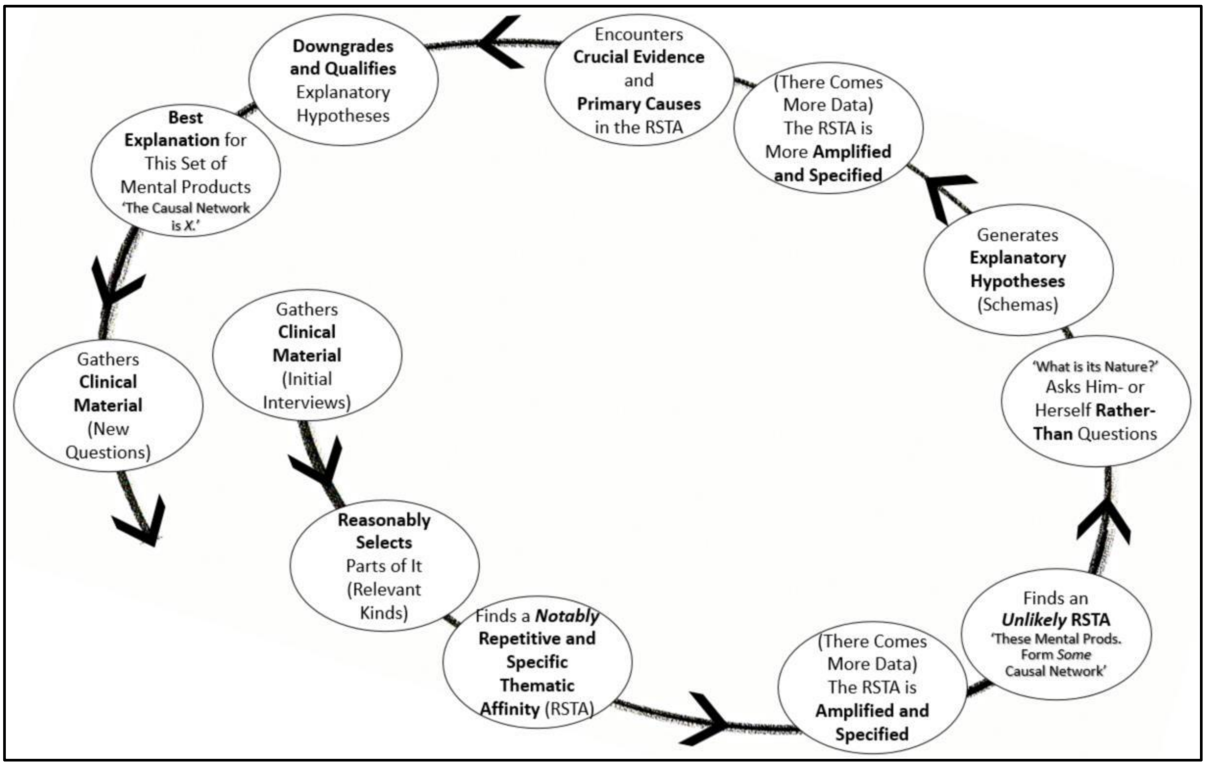
Analysts can perfectly embrace indeterminacy and refuse to project theorizations onto their patients. Even if analysts are primed by a very sophisticated theorization, like the Freudian-Darwinian on jealousy presented above, they can leave it for later in the face of the reality's insistence to therefrom divert. In entertaining the particular hypothesis that that patient is being resentful or envious rather than jealous, the analyst would have to resort to another theorization, to other core hypotheses. In the chaotic and complex reality of our minds, a very nice story usually has plot-twists demanding that we reinterpret its characters. It follows that clinical psychoanalysis cannot be reduced to a mere application of previous knowledge, nor be charged with inescapably committing *petitio principii* or confirmation bias (even if some practicing analysts may justly be charged with such "crimes").

At the same time, there is no way to deny that psychoanalytic inferences are constricted. The set of candidate hypotheses S-P_0-1–S-P_0-7, for example, could be all poor, improbable explanations for that patient's behaviours after all; however, without the background knowledge that constricted their generation, the only thing left for analysts to do would be a phenomenology, that is, an accurate report of the patients' behaviours. This constriction is good news, actually, for it means there is some manner and some order in the analysts' process of inferring particular hypotheses.

In this long section, I have attempted to describe and justify this manner and this order, which are summarized below.

[56]An optimal demonstrative example of my model would have to come, for sure, from transcriptions of real sessions and the reports of the real analysts who conducted them; let us keep working.

Figure 9 - The Inference of Particular Hypotheses in Clinical Psychoanalysis



Source: Elaborated by the author (2021).

5.2.4 Generating and Testing Categoric Hypotheses

The endeavours of generating and testing particular hypotheses in clinical psychoanalysis are thus intertwined with one another and involve a movement backwards, from effects to causes. As long as analysts, when treating their patients, group and inspect target-instances and control-instances through fixing mental products of interest, that is, as analysts compare the features of the instances where the mental product is present with the features of the instances where the mental product is absent, they depart from effects and are guided by effects. And, as we have just seen, this process frames a search for explanatory hypotheses for the mental products of interest, leading analysts to a pool of explanatory hypotheses which already bear a significant degree of probability; this is why we can say that, as regards psychoanalytic hypotheses that we are here naming "particular", the generating phase is at the same time the beginning of the testing phase.

When analysts are dealing with categoric hypotheses, though, the exact opposite should happen. The categoric hypotheses that analysts *generate* do not have to be probable in birth (though one might say they have to be at least *plausible*), and should be tested having the cause, not the effect, as a reference – that is, having the target-group as the group where the hypothesized cause is present, and the control-group as the group where the hypothesized cause is absent. This seems to bring the efforts to generate and test categoric hypotheses in clinical psychoanalysis closer to a scientific agenda with which we are most familiar, the one represented by experiments and quasi-experiments. But how exactly could this inferential picture work? And why should analysts infer categoric hypotheses so differently from the manner they infer particular hypotheses?

As we have seen in Chapter 3, we can best sustain an idea of mental cause in clinical psychoanalysis if we take it to be an INUS mental cause, that is, a mental event that is an *Insufficient but Necessary part of an Unnecessary but Sufficient* cause (Mackie, 1965) of a mental product, as this is the safest way to harmonize a premise of determinism with the complex *ontos* of the mental domain, as well as to account for the metaphysical and ethical issues surrounding this premise. With the idea of an INUS cause, when we say that a type of mental event is a cause of a type of mental product, we are *not* automatically claiming that the former will invariably cause the latter to develop, or that the former is invariably a part of what is causing the latter. We have seen that, this way, we are respecting two pressing questions that haunt the principle of mental determinism (while also evading a thorough analysis of such questions): the existence of agency and free-will and the existence of non-mental (e.g., purely biological) causes of mental phenomena. Categoric hypotheses, for this reason, must be about INUS causes: childhood episodes where one is sexually abused can be said to cause one's depression in adulthood, even though not everyone who was sexually abused in childhood will develop depression in adulthood and even though depression in adulthood can be caused by many other types of events.

But we have *not* seen in Chapter 3 that, *in the ambit of particular hypotheses*, the possibility of more than one sufficient condition for a mental phenomenon to occur (possibility that is expressed by the "US" in "INUS") is just absurd, if not self-contradictory. Particular hypotheses, as we have conceptualized them, are about things that do not happen to anyone's psyche except to the psyche of a particular person; they are about the variables and the processes that moulded a person's particular psyche, and they are only valid for this one person. So, when talking about what is particular to a patient, we are logically unallowed to claim that all the singleness of the patient – his or her unrepeatable symptoms, tastes, thoughts, etc. – would have transpired just the same had he or she been stricken by other factors, that is, had he or she lived another life. An unrepeatable effect could not have had a sufficient cause different from its actual cause – unless we assumed, which seems absurd, that it is impossible to divert from our destiny, we becoming the same person no matter what we had done or lived, like in Greek mythology, or assumed that there is a parallel universe or a twin-Earth where a *Doppelgänger* of ours exhibits our very personality in spite of having been exposed to different factors. Therefore, as regards particular hypotheses, analysts do not talk about "INUS causes", but about "INNS causes": about events that are Insufficient but Necessary factors of a Necessary and Sufficient set of factors.

Now we are able to understand why inferences about a patient can go from effects to causes while inferences about a category of patients can only go from causes to effects. If we assemble on one side instances where an effect of a INNS cause is present and, on another side, instances where the same effect is absent, it is extremely probable that we behold a difference related to the INNS cause; for, although in the group where the effect is absent, *some* instances *may* bear the cause, in the group where the effect is present, *every* instance *must* bear the cause (see p. 105). If we did the same with INUS causes, however, anything would go: between a group where the effect of an INUS cause is present and a group where this effect is absent, there are no logical or probabilistic constraints for the incidence of the INUS cause. Between a group of people with lung cancer and a group of people without lung cancer, we may or may not see a difference in smoking habits; not every person with lung cancer has smoked, and not every smoker will develop lung cancer[57]. On the other hand, the habit of smoking, as an INUS-cause, enhances the probability of developing lung cancer; so, if smoking really is a(n INUS) cause of lung cancer, the incidence of the disease will be more probable in the group of smokers relatively to the group of non-smokers. Similarly, in order to test a hypothesis that childhood episodes where one is sexually abused causes one's depression in adulthood, it is useless to compare depressed with non-depressed patients and check whether the group of depressed patients presents a significantly greater number of episodes of sexual abuse in childhood – or to look for any other causal difference between such groups, for that matter. To test such hypothesis, what must be checked instead is whether there is a difference in the incidence of depressive phenomena between the group of patients where the presumed cause – the episodes of sexual abuse in childhood – is present and the group of patients where this presumed cause is absent. Categoric hypotheses in clinical psychoanalysis are about INUS (mental) causes for

[57]Of course, in this case and in other cases of INUS-causation, we could probabilistically expect a significant difference between the groups if, in the considered domain, other sufficient causes were uncommon (or neutralized by being more or less equally present in both groups) and/or the other factors that, united to the INUS-cause in question, would be sufficiently effective were common. As these pictures are contingent and arduous to be detected (unless there is a large and pertinent background knowledge allowing us to detect them), I am here sticking to the simpler idea that significant differences in these cases may or may not exist.

(mental) phenomena, and an INUS cause should make the phenomena it causes more probable; but the effect of an INUS cause should not make more probable that it was this INUS cause that (partially) caused it.

Let us suppose there is a database containing details of thousands of rigorously conducted psychoanalytic treatments and assessments; in each treatment or assessment, such details would be composed of the reasons the analyst in question gave to his/her inferences about the psyche of the patient in question – (a summary of) the evidence and the logic behind his/her particular hypotheses about the patient –, as well as of the transcripts of each of the treatment's sessions or assessment's interviews. Let us also suppose that a community of analyst-researchers has access to this database and wants to know whether the murderous fantasies and actions of males are (partially and contingently) rooted in an unconscious need to "take revenge" upon their rejecting mothers[58]. How should the community use the database to test this categoric hypothesis? It should collect from the database all cases where a male patient shows unconscious motives of revenge upon his rejecting mother, compose a control-group of cases of males with similar features (age, social class, sexual orientation, etc.) but who do not show such revengeful motives, and verify whether there is a significant difference between the groups concerning murderous fantasies and actions. In abstract terms, the community should "fixate" the cause and verify the effect, not the other way around. If the significant difference is seen (and given that the hypothesis is explanatorily virtuous), then the community can claim the hypothesis has been confirmed – that its truth is highly probable.

This would be very far from what Freud used to do to concerning categoric hypotheses, which makes us admit that his inductive arguments toward the latter were indeed very weak. Let us take, for example, Freud's hypothesis that castration anxiety and the disavowal (*Verleugnung*) of the perception of the female genitals is the cause of male fetishism (Freud, 1975; see Simanke, 2013). Or his hypotheses, extensively discussed by Grünbaum (1983, 1984), that a premature sexual incident plays an important part in obsessional neurosis, and that paranoiac symptoms are defences against homosexual impulses. What did Freud use to do to conclude such things? He used to refer to a group of patients with a certain mental ailment (paranoia, for example) that he or another analyst had treated, and to observe that the reports and behaviours of all the patients in this group presented a certain element (in the cases of paranoia, homosexual impulses and an abhorrence before them). He for certain has appealed to a background knowledge and to explanatory concerns in every step of his reasonings, from the

[58]This is a hypothesis presented in Schachter (2003), who takes it from Abrahamsen (1973).

dough to the icing. Nonetheless, with this conduct, we can now see, he was committing two errors: he did not devise a control group to make sure his background knowledge and explanatory concerns were leading him to an actual cause; and he looked for causes before effects, not for effects after causes – in other words, he composed groups according to psychopathological categories, not according to their hypothesized causes.

For the sake of justice, the conceptual framework we are here using is not found in Freud, especially the one related to INUS causation. It seems that Freud treated what we are here calling categoric hypotheses as hypotheses over what we are here calling *INNS causes*. For example, in the paper on fetishism:

> Probably no male human being is spared the fright of castration at the sight of a female genital. Why some people become homosexual as a consequence of that impression, while others fend it off by creating a fetish, and the great majority surmounts it, we are frankly not able to explain. It is possible that, among all the factors at work, we do not yet know those which are decisive for the rare pathological results. We must be content if we can explain what has happened, and may for the present leave on one side the task of explaining why something has not happened. (Freud, 1975, p. 154-5).

In "The aetiology of hysteria", this INNS picture was put forward even more completely:

> In order to establish the bacillus as the specific aetiology it is enough to show that tuberculosis cannot possibly occur without its playing a part. The same doubtless applies to our problem. It does not matter if many people experience infantile sexual scenes without becoming hysterics, provided only that all the people who become hysterics have experienced scenes of that kind. The area of occurrence of an aetiological factor may be freely allowed to be wider than that of its effect, but it must not be narrower. Not everyone who touches or comes near a smallpox patient develops smallpox; nevertheless infection from a smallpox patient is almost the only known aetiology of the disease (Freud, 1962b, p. 209).

If Freud defended this form of causation for the categoric hypotheses of psychoanalysis, it is acceptable that he were always heading from categoric effects to categoric causes; but, then, logic would tell him to expect the conjectured cause in *every* case where he encountered the effect; and, when not encountering the conjectured cause in such cases, he would have to *abandon* the "INNS hypothesis" *altogether*. Therefore, the following passage of "The aetiology of hysteria" demonstrates Freud's conceptual and/or logical inconsistency in this respect:

> No doubt you may raise the objection that the nineteenth or the twentieth analysis will perhaps show that hysterical symptoms are derived from other sources as well, *and thus reduce the universal validity of the sexual aetiology to one of eighty per cent* [emphasis added]. By all means let us wait and see; but, since these eighteen cases are at the same time *all* the cases on which I have been able to carry out the work of analysis and since they were not picked out by anyone for my convenience, you will find it understandable that I do not share such an expectation but am prepared to let my belief run ahead of the evidential force of the observations I have so far made (Freud, 1962b, pp. 199-200).

His bet was that reality would invariably remain showing the same conjunction of hysteria and repressed sexual memory; but he could not have affirmed that, in case it did not, the universal validity of the sexual aetiology would be reduced to eighty percent; for, in case it did not, there could not be any INNS causation and the universal validity of the sexual aetiology would immediately reduce to exactly *zero* percent.

A full account of Freud's logical usage of the concept of cause would be unquestionably interesting and would help us to do him justice but, as this is not our aim here, let us leave it for other authors.

Finally, a word on the process of *generating* categoric hypotheses. Why, with categoric hypotheses, the generating process is not already a context of justification, like it is as regards particular hypotheses? Why should categoric hypotheses in clinical psychoanalysis be formulated first and tested second, instead of being (in a certain sense) formulated and tested at the same time?

Let us suppose a community of analysts wants to know what has changed in the unconscious life of Brazilians from the outbreak of COVID-19 on. The hypothesized cause would be very clear: the outbreak. The target- and control-groups would also be clear: respectively, sessions/interviews[59] from after the outbreak, sessions/interviews from before the outbreak. But what to look for between the groups? Someone could answer: well, for great thematic differences! But where could we start to look for them? Someone could suggest: in dreams, for dreams are the royal road to the unconscious. The investigation would still be too indeterminate, for it can be reasonably expected that great thematic differences between the dreams of the groups would reveal themselves multifarious and abundant – what would then be the crucial ones? It is true that one could be assisted here by core hypotheses, like in the inference of particular hypotheses, and formulate a list of candidates for the effects of the pandemic on dreams. But the complex data, the groups of numerous dreams, would hardly be a material to fabricate such a list. Let us remind that the pandemic would be an INUS cause in this investigation: its effects, whatever they would be, would not be unanimous in the target-group, but only more probable than in the control-group. The moral here is that it is impracticable that categoric hypotheses be *generated* from a close inspection of the very data that would ultimately test them.

[59]As they were systematized by respective case studies; in other words, by respective particular hypotheses.

Therefore, categoric hypotheses should be plausibly generated from core hypotheses, common-sense and coincidences some analysts have noted in everyday practice; the data would only be the material for their sheer *testing*. In the example above, the community could propose the following question: was the pandemic responsible for more dreams about, say, contamination or, say, about the death of loved ones? And the clinical cases would answer it: yes, or not so much.

5.2.5 Remarks on the Justification of Core Hypotheses

We should by now be convinced that it would be epistemologically dishonest to try to confirm core hypotheses through derivatives of clinical data, namely, particular and categoric hypotheses. This core is what leads the perception and the primitive interpretation of clinical data throughout the inferential process of particular hypotheses; and assemblages of the latter can confirm categoric hypotheses, as we have just seen. It follows that to cite psychoanalytic cases as well as psychopathological and cultural theses of clinical origin as evidence for the core of psychoanalytic theory is to engage in a circular business. If we are in a Jungian mood, such attitude is comparable to a snake that eats itself by the tail; if we are instead in a Freudian mood, a way to understand the attitude is by comparing it to an act of autofellatio. Rubinstein (1980) argues that it is

> impossible for us to count the confirmation of the particular clinical hypothesis also as confirmation, in the indicated sense, of the general clinical hypotheses that are involved. The reason is clear. I will refer to the particular clinical hypothesis as h and to the relevant set of general clinical hypotheses as S. The fact that, having assumed the validity of S, we can in part confirm h obviously does not allow us, as it were, to turn around and claim that, since h now is in part confirmed, S is also confirmed in some measure (p. 435).

Thus, the core of psychoanalytic theory must lean on something external to itself – on non-psychoanalytic theories and concepts. We have already made a claim along such lines (see figure of p. 117), but now we shall reflect on it with more care. When we mind about which and how non-psychoanalytic theories and concepts support and are adequate to support the core of psychoanalytic theory, two refinements may move us forward. First, the problem of how such psychoanalytic core was born is different from the problem of how it grew, changed and keeps growing and changing. Second, we do well to bear in mind that the external theories and concepts supporting such core may not be restricted to scientific ones, but may also extend to common-sense.

The scientific theories on which Freud leaned along the delivering of his baby theory by the end of the 19th century had been in their turn delivered through many justificatory regimes and many methods, including methods we could characterize as purely clinical. There exists a plethora of historiographic works showing the origins of Freud's theory, but a recent and competent one makes it especially clear that it was born out of an exchange between pre-psychoanalytic theories and the clinical phenomena Freud had had contact with. Padovan (2018) shows this origin within each of the three models proposed by Freud between 1886 and 1896: his neurological model of the mind, his nosological model and his dynamic model. Padovan (2018) traces the origin of Freud's new dynamic model and the notion of transferential resistance it contains, for example, back to Freud's dialogue between theorizations by Bernheim (the existence of autosuggestions against the idea of a cure) and Salpêtrière doctors Charles Féré and Alfred Binet (the existence of a kind of erotic reaction to the hypnotic relationship) and his own experimentations with patients Mathilde Schleicher and Fanny Moser.

This transformation of external theories into psychoanalysis – this digestion – sure went on happening after psychoanalytic theory was already distinctive and substantial, in the beginning of the 20th century. It also keeps happening today, in our anxious 21st century. Yet, while at the beginning psychoanalytic theory was voracious and had diversified meals, today we see communities imposing restricted diets to it. Because the scientific and political mainstream accept only the clinical practices based on theories developed under the lights of experimental and quasi-experimental methods, theories from the realms of, for example, neuropsychology, social psychology, cognitive psychology, etc., some psychoanalysts have repressed methodological eclecticism. Others have not. There are analysts who keep departing from historical and cultural studies to inform and transform Freudian perspectives on, for example, sexuality and gender.

In other words, the data made available and apprehended in a certain setting where associations, attentions and fantasies can flow (relatively) free, a setting historically called "psychoanalytic", have been filtered and enriched by diverse pre- and non-psychoanalytic theories, and this process is what has generated and keeps generating the core of psychoanalytic theory. There is first a *problem of description* here. How this comes about – how this data has transformed and keeps transforming non-psychoanalytic theories into a psychoanalytic core, in logical terms – is an interesting and complicated question, which would demand more time and space to be appropriately tackled (a substantial start can be found in Michael, 2019). Maybe, similarly to categoric hypotheses, the psychoanalytic core has been plausibly suggested from an untraceable blend of impressions and knowledges; or, similarly to particular hypotheses, it has emerged from a clear list of competing explanations. In any event, if core hypotheses are meant to be valid to every member of humanity, how could we devise here something similar to a control-instance? At the end of this section, I shall suggest an initial answer to this question.

Apart from a problem of description, this picture shows a partial *solution* regarding the *justification* of the psychoanalytic core. The burden of the justification of the psychoanalytic core could well be put on the shoulders of the external theories from which it has been generated. This picture gives this core the right to outsource most of its empirical and logical support… Many Freudians have realized it – most of them influenced by Grünbaum, by the way – and, as commented above, have only dared to tinker with the psychoanalytic core having in hand reliable methods and solid theories.

But is it really impossible to have an "internal" justification of core hypotheses, namely, to have the clinical data of psychoanalysis virtuously confirming (or potentially falsifying) its theoretical core? Well, there is no such thing as a raw, atheoretical, clinical datum. The closest we could get to a raw clinical datum would be one interpreted by common-sense. For example, phenomena of pleasure and unpleasure: anyone can point to them without the backing of some scientific theory. We know that someone is feeling happy without having to know that feeling happy is a *qualia* accompanying the discharge of a certain amount of bodily energy, for example. The justification of core hypotheses over phenomena of pleasure and unpleasure, and over many familiar phenomena, thus, would not need to have a circular fate, nor depend on non-psychoanalytic theories and on the methods that constructed such theories. Of course, it would be fair enough if we took common-sense to consist in some sort of non-psychoanalytic theory constructed by certain methods, and this way our first formula would be preserved. But there is another way through which the clinical data of psychoanalysis could test its own theoretical core without falling into circularity.

Let us suppose that a group of analysts use core hypothesis x to primitively interpret a clinical datum; now, could the result of this, the meaningful clinical datum, be incoherent with *another* core hypothesis, y, namely, could it falsify the conjunction of x and y? Well, of course it could, and in this case the analysts would have three alternatives: to question the integrity of the datum in order to keep x and y; to reject x; or to reject y. Let us suppose, now, that the analysts evaluate y as less solid than x because it has been questioned before, because it has already clashed with other meaningful data, and that they decide to reject the trembling y in order to fix the inconsistency. End of story.

But, if y still remained coherent with – and explanatory of – a *lot* of data, the analysts could stick to this and decide that y must be somehow amended or detailed, and maybe also inserted in a wider structure, in order to accommodate the refuting data without ceasing to accommodate the long-amicable data. This is a case where there would be an *internal* testing of the psychoanalytic core, where part of the latter is falsified and then transformed solely from an exchange with psychoanalytically meaningful data. There is no confirmation *stricto sensu* of this restructuring, that is, of the new hypotheses – for they are tailored to accommodate data – but, as they make the whole of the core hypotheses more explanatory, we could say they represent a sort of "confirmatory generation". (By the way, we could also say this about particular hypotheses.) In this case, the generation is focal, restrained, for it aims to resolve a specific inconsistency and at the same time to preserve all the successful rest; the new core hypotheses, thus, are elaborated so as to show a good chance of mirroring the truth; they are born tested, so to speak.

(Of course, this process may also be triggered when the discrepant clinical datum had been made meaningful by common-sense or by any other scientific theory.)

We can find this happening in Freud's great theoretical revisions. In the beginning of the paper "Beyond the Pleasure Principle", for example, the core hypothesis of the Pleasure Principle trembles before uncanny phenomena. There, Freud attempts to salvage the good-old Principle by questioning the integrity of the phenomena. Were analysts looking properly at the discrepant cases? Were they missing something? But he admitted that the refutation was insisting and insisting. He had to propose, then, a new foundation for psychoanalysis: there must be a psychological principle that goes beyond the tendency to seek pleasure. The tendency to seek pleasure was not exactly considered amiss and discarded, though, but only included in a more encompassing paradigm, organized by the famous concepts of "compulsion to repetition" and "death drive".

In the first three sections of this paper (Freud, 1955a), his reasoning follows the steps below:

1) He restates the Pleasure Principle: it is the assumption that the course of mental events "is invariably set in motion by an unpleasurable tension, and that it takes a direction such that its final outcome coincides with a lowering of that tension – that is, with an avoidance of unpleasure and a production of pleasure" (p. 1). He also tells us where the Principle came from: from "the facts of daily observations in our field of study" – clinical data –, and from "impressions […] so obvious that they can scarcely be overlooked" – common-sense (p. 1). He was also "not indifferent" to see that G. T. Fechner's "view on the subject of pleasure and unpleasure" – that is, a view present in a non-psychoanalytic domain – "coincides in all essentials" with the psychoanalytic view on it (p. 2).

2) He presents the first *data shaking the Principle up*: the sparsity of pleasure in mental life. He points out that, because "universal experience completely contradicts" the idea that most mental processes are connected to pleasure or lead to pleasure, it would be incorrect to say that the Principle is dominant along mental processes (p. 3). He *explains away* the fact by the idea that the path toward pleasure would be just a tendency, and as such liable to be interrupted by certain "forces or circumstances" (p. 3). He cites, from a "rich fund of analytic experience", three of such forces or circumstances that could hamper the outcomes of the Principle (p. 4): when the pursuit of pleasure threatens the organism, which, imbued with the instinct of self-preservation, is thus coerced to temporarily tolerate unpleasure, in other words, to postpone satisfaction (a process condensed in the idea of the Reality Principle); when the primitive instincts that have become incompatible with the aims and demands of the ego as a whole once it has become more complex and organized keep reaching satisfaction (a process related to the concept of repression); at last, most of the unpleasure we feel would be simply due to menacing perceptions from inside the body (unsatisfied instincts) and outside the body (dangerous situations). Such sources of unpleasure would only promote detours from the Pleasure Principle and would not refute its pre-eminence.

3) The second *anti-evidence* for the Principle is the cluster of dreams dreamt by patients diagnosed with "traumatic neurosis" (what we today know as Post-traumatic Stress Disorder), a condition that "occurs after severe mechanical concussions, railway disasters and other accidents involving a risk to life" (p. 6). It would be expected that the dreams of these patients projected images of recovery or of a healthy past, for dreams should be a great medium for the Pleasure Principle; instead, their dreams often recreate the situation of their accident with vividness, causing them to wake up in terror (here, it is Freud's common-sense that tells him this does not look like pleasure). He then admits that, if the universal wish-fulfilment tenor of dreams is to be clung to, *one would explain the phenomenon away* by the hypothesis that "the function of dreaming, like so much else, is upset in this condition and diverted from its purposes" (p. 7). He seems to find this *ad hoc* unconvincing, so he starts to wonder about "the mysterious masochistic trends of the ego" (p. 8).

4) The third of Freud's *refuting instances* comes from a close and long observation of the playing habits of a toddler – his grandson. The grandson invented a solo game, engaging in it again and again with glee, which consisted in throwing a wooden reel away while uttering "o-o-o-o" (from the German "*fort*", "gone") and pulling it back by a string while uttering "a-a-a-a" (from the German "*da*", "there"). Before coming up with such elaborate game, he used to just throw away from himself any small object he could grab (uttering "o-o-o-o") and to leave

it at a distance. Considering the history of the toddler's attachment to his mother and the core hypothesis of the Oedipus Complex, Freud interpreted that the complete game was representing him losing the reach of mother and then getting mother back. Freud stresses that the toddler "cannot possibly have felt his mother's departure as something agreeable or even indifferent", and then questions how would "his repetition of this distressing experience as a game fit in with the Pleasure Principle" (p. 9). This is one of the occasions where a datum interpreted by core hypothesis *x* contradicts a core hypothesis *y*, there appearing, thus, a problem of consistency. Freud then presents *the three hypotheses that could preserve the Principle*. The first hypothesis is that the symbolized departure was an inevitable prelude for the re-enacting of the return, and there was no doubt that the latter was a merry event; but Freud remarks that the departure act "was staged as a game in itself and far more frequently than the episode in its entirety, with its pleasurable ending" (p. 10). Freud's second *ad hoc* hypothesis tells that the boy's game was a way to embody an active part in the dismay and distress he had suffered and thus satisfy an "instinct for mastery that was acting independently of whether the memory was in itself pleasurable or not" (p. 10). His third *ad hoc* points to the chance that he was giving expression to a wish to retaliate against his mother, as if shouting to her: "'All right, then, go away! I don't need you. I'm sending you away myself.'" (p. 10). At a certain point, however, Freud observes that "no certain decision can be reached from the analysis of a single case like this" (p. 10).

5) The cluster of evidence that definitely pushes Freud to the acceptance of a principle in the human mind operating beyond the Pleasure Principle was collected by him both on and off the couch. On the couch, it is the same evidence that made him change his therapeutic aims and techniques: the evidence that patients have not been touched by plain and direct interpretations, because of the ubiquitous process of resistance, nor by suggestive incentives to abandon their resistance, a fact flagrant in their long-run failure to remember the whole of what in them was repressed and pathogenic; and the evidence that, instead, the only way they were able to remember the pathogenic material (attaining thereby therapeutic results) was by *reliving* it in the therapeutic rapport and later recognizing that "what appears to be reality is in fact only a reflection of a forgotten past" (p. 13). Why would this technical evidence refute the Pleasure Principle? First, because it was making patent that the patients *cannot help but* to relive past conflicts inside the analytic relationship: the reproductions "emerge with […] unwished-for exactitude", are "invariably acted cut in the sphere of transference" (p. 12), are "under pressure of a compulsion" (p. 15). A second and more important reason is that there is no imaginable hypothesis through which the Pleasure Principle could make sense of such compulsions. They bring back the conflicts that involved the decay of early sexual life: a parental love that got

extremely conditionalized or was significantly lost, a failure to rescue it, a sexual research that led to no satisfactory conclusion, etc.. Freud observes that

> None of these things can have produced pleasure in the past, and it might be supposed that they would cause less unpleasure to-day if they emerged as memories or dreams instead of taking the form of fresh experiences. They are of course the activities of instincts intended to lead to satisfaction; but no lesson has been learnt from the old experience of these activities having led instead only to unpleasure (p. 15).

To get to this, he told the reader what should be stressed and refined in his metapsychology from the recognition of the patients' compulsion to repeat: resistance does not come from "the repressed", which he would later call "the id", as it forever forces its way to consciousness or to action, but from a higher system of the mind, the ego, a part of which is also unconscious. He then reminds that the unpleasure felt by the ego with the compulsion to repeat does not need to contradict the pleasure principle, for an idea can elicit unpleasure in one mental system while eliciting pleasure in the other, the archaic one. "But we come now to a new and remarkable fact", Freud then concedes, "namely that the compulsion […] also recalls from the past experiences which include no possibility of pleasure, and which can never, even long ago, have brought satisfaction […]" (p. 14). Off the couch, namely, as regards people considered normal or non-neurotic, the evidence adduced by Freud to abandon the simple version of the Pleasure Principle is that many among them – something easily witnessed – develop their intimate relationships in a way that always obtain the same uncomfortable outcomes, giving thus the impression of bearing a "malignant fate".

6) Finally, Freud recognizes his metapsychology is not good enough for reality; it does not cover all of the data that have presented itself to him and his fellow doctors. Having provided the reader with all of the refuting evidence, he announces a need to accommodate it without the previous *ad hoc* hypotheses, enunciating thus a new principle, rooted deeper in the human mind:

> If we take into account observations […] based upon behaviour in the transference and upon the life-histories of men and women, we shall find courage to assume that there really does exist in the mind a compulsion to repeat which overrides the pleasure principle. Now too we shall be inclined to relate to this compulsion the dreams which occur in traumatic neuroses and the impulse which leads children to play.
> […]
> […] Enough is left unexplained to justify the hypothesis of a compulsion to repeat – something that seems more primitive, more elementary, more instinctual than the pleasure principle which it over-rides. (pp. 16-17).

On the other hand, the new principle could not imply the withdrawal of the Pleasure Principle, but could only complement it; for the compulsion to repeat would be easily found convoluted with instinctual satisfaction, as occurs in children's play and in the psychoanalytic process.

Along the rest of the paper, Freud's aim is to argue for a plausible compromise between the principles – with an eye on the uncanny evidence.

To maintain the mood of the momentous paper, we could say Freud's reasoning here resembles a war. There were battles: the adverse evidence hit and the core hypotheses resisted; more adverse evidence hit harder, and the core hypotheses trembled; finally, adverse evidence delivered a *coup de grâce*, and the core hypotheses fell. The old metapsychology was defeated. But it still holds some dignity and proposes a peace treaty, so that some of its borders and cultures are preserved. The Pleasure Principle still fully accounts for the clinical data that was grasped before that moment, so the new principle must accommodate it as well. The new principle is determined on one side by empirical limits, and on the other by theoretical limits.

Kuhn and Lakatos competently described this negotiating habit – this dialectic between the old theory and the new problems, with a tendency to cling to the old theory – and ascribed it to all scientists. It is triggered by a problem of internal consistency, where the data as interpreted by a hypothesis contained in a paradigm or research program (or by common-sense) is inconsistent with another hypothesis of the same paradigm or research program (or with common-sense), and it can be properly understood as a habit of *internal testing*. Around 40 years ago, Clark Glymour proposed a model for this sort of testing, labelling it a "bootstrap strategy of theory testing", and, what stands very much in favour of our argument, ascribed it also to Freud (Glymour, 1980, 1982).

According to the model, one bootstrap-tests a theory if one: 1) uses bits of the theory to compute non-theoretical data, drawing from this data theoretical and non-theoretical inferences; 2) checks if these results cohere with some of the other bits of the theory and other non-theoretical data. It is "bootstrapping" because part of the evidence used to test a theory is computed by that same theory. Glymour (1982) claims it to be a strategy frequently used in the non-experimental sciences, but also "whenever an experimenter attempts to test a theory containing quantities which the experimenter does not know how to determine save by measuring other quantities and computing using some of the very hypotheses to be tested" (p. 18). He carefully demonstrated that Freud's study of the Rat Man would also embody this strategy, but through dealing with states of affairs rather than with quantities.

The bootstrap strategy of theory testing as instantiated by Glymour in the Rat Man's study is in many respects similar to the "internal testing" we have seen occurring in "Beyond the pleasure principle" – which could be taken as a possible, actual, and intelligent way to question and advance, *in general*, the core hypotheses of psychoanalytic theory without having to depend on extrinsic inputs. An important difference between the accounts, though, is that mine is restricted to core hypotheses, while Glymour's does not work with any such conceptual point.

There is a last question about the core hypotheses clamouring for a comment. If the core hypotheses are supposed to be valid for each and every human mind, how could we devise control-instances relative to their factors, how could we apply the Method of Difference – a step that seems *sine qua non* in any of the accounts of "confirmatory generation" of core hypotheses just presented? If a mental factor, mechanism, process, or principle is supposed to be universal in the human domain, where could its negative subdomain be? And, without a negative subdomain, how could we validate the mechanism, process, or principle?

A group of people has sunbathed for 10 minutes a day, for 1 whole month, and none of them got cancer. We know by now we have no grounds for concluding that the sun has had some influence in the group's lack of carcinogenesis. I may propose the hypothesis that all animals are born with a "magic glow" influencing all of their behaviours, *all of them*, *from birth to death*. But, please note, the "magic glow" is not visible, *only its effects in behaviour are so*. Well, put this way, we could not prove nor disprove the existence of the "magic glow"… In both cases, we could never be sure of what are exactly the effects of the factor in question – or even if its existence makes sense – because it would be presumably unanimous in the domain. We know by now we have to compare something to something, to estimate a difference, if we wish to investigate any causal action, and we could not treat the all-pervading ones as an exception to this rule.

But there is in fact no *uroboros* here. Although talking of universal stuff, core hypotheses do *not* ascribe stuff to the whole domain of the science in question *in an equal manner*. Every element in nature presents some degree of radioactivity, but some elements show radioactivity more than others. Every species is under an evolutionary process, but the Galapagos was a nicer place to behold such process. *Every* human mind suffers from an impulse to repeat its traumas, but *some* minds suffer *more* from it and succumb *more* to it than others. According to psychoanalytic theory, the death drive is universal, but it is not equally distributed among us. So much so that Freud took some time in his career to realize (or to recognize) the phenomena ranging beyond the Pleasure Principle; he needed natural experiments, extreme cases, in order to do it. While categoric hypotheses talk of contingent traits or conditions – gender, age, political sympathy, sexual fantasy, physical impairment, etc. – core hypotheses talk of *contingent* – yet *always potential* – *experiences*. So, for example, if a psychoanalytic hypothesis is about the process of falling in love, it is a core, not a categoric, hypothesis, for every human has the potential to fall in love throughout life; nevertheless, not every human is in love right now. So, here are the target and the control-instances for the "confirmatory generation" of core hypotheses: cases where the universal experience is extreme and cases where it is mild, imperceptible, or just potential. Through this contrast, one can better see what is being postulated. It is easy to accord that core hypotheses are falsifiable; now we know how they become supportable.

This idea would perhaps need to be further developed, but I am leaving the task for others or for my future other because I have already gone on too long with this model. Let us now put the figure in ample ground: we shall see that the patterns of inference in clinical psychoanalysis are far from unique.

5.3 CLINICAL PSYCHOANALYSIS IS SIMILAR TO HISTORY: WALLACE

Erected from multiple sources, the model just presented is intended to be a *realistic description* of how analysts infer the causes of patients' mental products and of how a community of clinical analysts could, with the data of the treatments led by them, enrich psychoanalytic theory and add to what we know about the human mind, be it in sickness or in health. Its *reasonableness*, however, is still in question. I shall now demonstrate that the model possesses many of the features current in recognizably strong inferences from the historical sciences.

I shall first present, through Edwin R. Wallace's arguments, the convergences between psychoanalysis and the historical sciences concerning both their topics and their methods. The objects of psychoanalysis – the unconscious motivations – develop over time and the investigation in clinical psychoanalysis transpires mostly *ex post facto*. With Wallace, I contend that such pair of facts tint clinical psychoanalysis with indissoluble historiographic colours and that, accordingly, the epistemological problems and possible resolutions of the historical sciences can be transposed to clinical psychoanalysis.

Then I shall be allowed to compare life phases with geological ages, fantasies with dinosaurs. In the following sections, the model I have proposed above for clinical psychoanalysis shall be compared with a model, proposed by Carol Cleland, of how inferences in historical natural sciences come about. Since Cleland's model is declaredly a good example of Inference to the Best Explanation – an endeavour to understand how we weigh the evidence at hand and come to new beliefs about the way things are (or, in this case, were) (Lipton, 2004) – I shall also examine in what such wider model consists and discuss some of its merits.

In a 1985 book, Edwin R. Wallace presented a strong case for the proximity between psychoanalysis and history. Many authors had already appreciated the fact, he reported to us then:

> In his 1935 introduction to Zilboorg's *Medical Man and the Witch in the Renaissance*, Henry Sigerist, the dean of American medical historians, remarked on the striking similarities in the procedure of academic historians and dynamic psychiatrists. The historian Hughes (1964, p. 47) concurs, asserting that "Psychoanalysis *is* history." Similarly, the former president of the American Historical Association, William Langer (1958), the philosophers of history Meyerhoff (1962) and Walsh (1969), the philosopher of psychoanalysis Ricoeur (1970), and the psychoanalysts Schmidl (1962), Wolman (1971), Novey (1968), Schafer (1976, 1978), and Leavy (1980) have written about affinities between the two disciplines. I believe that psychoanalytic clinicians have long sensed that they are practicing what Loewald (1977) has termed "the history of the individual." (Wallace, 1985, p. 1).

But none of them, he continued, had yet treated the topic with the systematic and comprehensive vigour he intended to realize in what would follow. With an M.A. in History, as well as a "lifelong passion" for it, the dynamic psychiatrist had become more and more adhered to the idea "that history and psychoanalysis share methodological-epistemological issues which it is mutually advantageous for them to examine" (p. ix).

Three clusters of arguments in Wallace's book are valuable for us here: the arguments on the similarities and differences between history and psychoanalysis; the arguments on the role of theory in both historical and psychoanalytic inferences, and on the possible justification of this role; at last, the ones regarding the peculiar idea of cause enacted by both historians and analysts.

History and psychoanalysis would be interested in the same thing, namely, the *symbolically mediated occurrences* of the human domain – be they actual or potential, explicit or implicit, intended or unintended, verbal or non-verbal – and would try to understand such occurrences through a *diachronic perspective*. This thing which both sciences investigate would have an "inside": the communicative occurrences of the human domain would be determined and supplemented by occult *meanings and motives*. So, in the second place, both would have to count on psychological intermediates for recovering the past and, because of this, *their facts would not be facts per se, but only traces of facts* (Wallace, 1985).

This means that *both interpret interpretations, instead of flawless reports*. Historians select, order and interpret a material that chroniclers and archivists have already selected, ordered and interpreted; the same happens between analysts and patients. Wallace (1985) describes the four levels of interpretation presupposed in both sciences: the original interpretation of the event, at the time of the event; its subsequent mental elaboration; the elaboration at the time of the event's retelling; the historian's or analyst's interpretation of this retelling. Hence, and at last, *the data of both are subject to be corrupted*, consciously and unconsciously, through the motives of such informants. Patients, archivists and chroniclers "delete, disguise and distort, to ends both known and unbeknownst to them" (Wallace, 1985, p. 8).

Wallace (1985) suggests an analogy between the psychoanalytic patient and the historical document. Both would have a meaning to be unpacked little by little, a process demanding good eyes, good ears and a good, sharp, mind, operating with synchronic *and* diachronic approaches. He even refers to Boisen's (1936) assertion that the human experience is a living document. However, he adds: "Of course there is a sense in which the analyst's text is more like a palimpsest; it has been continually rewritten and yet it bears the traces of its previous chronicles; through the analysis it will be rewritten once more" (Wallace, 1985, p. 70).

As people transmit and transcribe documents throughout the millennia as continually as patients rework their remembrances, we have to consider that documents and remembrances carry the same risk of continually losing their original forms. Would analysts benefit, like historians, from a "diplomatics", namely, from a study of the nature and authenticity of their "documents"? Maybe not, claims Wallace (1985), for analysts, unlike historians, are not so interested in distinguishing psychical from actual reality (even though there are moments in every treatment where this would be paramount). Anyhow, historians would admit, like analysts, that it is not always disclosable how much myths and illusions are corrupting a precious past (Wallace, 1985).

Wallace (1985) imagines someone objecting: does it make sense to talk of a history of *psychical* reality? And he would reply that "fantasized interactions with others are just as legitimately a part of one's history as actual ones", "actual" here meaning the "publicly observable aspects" of such interactions (p. 45). Thus, "even where it is ascertained that the original interpretation was more a function of a wish or defensive trend than of the child's accurate perception of the situation, this interpretation is still an integral part of the historical event" (p. 88). (Psychical reality would also be equally material, since biological substrates and neurological processes are involved in encoding the memory trace of a fantasy or interpretation as much as in establishing the memory trace of the publicly observable aspects of an event.) Besides, there would be no definite limit between the outer and the inner worlds: the former would determine the latter and vice-versa.

Another plausible objection to demean Wallace's amalgam of the two sciences is that analysts and patients would not deal with the past *stricto sensu*. Indeed, an analyst would never ask a patient to bring an old letter, journal, drawing, filming, etc., to foster the treatment, unless the patient had the wish to bring something of the sort due to a *current* question, conflict, drive, etc.. But he contends that, although the evidence serving the treatment is limited to the present, this is specifically a present referring and drawing itself to the past – and, as any present, condensing and preserving the structure of the past at a considerable degree.

The most phenomenological of psychologists cannot deny that the reach and the interest of any psychotherapy are also in the past. Denial of this would imply an absurd: “To argue that personality structure and memory traces bear no relationship to historical events is to exempt psychological processes from the principle of historical continuity that appears to apply to the rest of the universe” (Wallace, 1985, p. 81). Hence, we can concur that both analysts and historians deal after all with *traces* – with a past made present –, struggling to find rules by which to infer what has gone from what remains. In a tone of conclusion, Wallace (1985) writes:

> Although it is true that the analytic process transpires totally in the present (when else could it occur?), it is only through the interpretation of certain of the patient’s current behaviours as traces of a past reality (i.e. as transference) that they are appreciated in their full significance. In other words, *without a concept of the past, one cannot fully encounter the present* (p. 94).

Withal, Wallace (1985) grants there are pivotal differences between the disciplines. Historians, save among them the biographers, are usually more interested in cultures and periods than in personalities. Analysts dig for the past to help the patient act in the future, by restructuring, and in the present, by alleviating, while historians should go for the past with no end but to reflect it in writing; a historian writing history to make history would be accused of a bias, acknowledged as epistemologically pernicious, called “presentism” (or “whiggism”). Most significantly, the “document” or “relic” of analysts, opposing the ones investigated by historians, *lives*; like a time-travelling amulet, “it” responds to their questions and, through de-repressions and acting-outs, opens a portal to a great number of unknown, fresh “relics”. The time-travelling paradox is trivial in a treatment. Analysts are in the most radical manner a part of the history of the individual about whom a history is under register, their patients. Thus, “the metaphor of the patient as document is only approximate” (Wallace, 1985, p 113). Also, in contrast to historians, analysts must examine the *current* influence of institutions and persons, including, of course, themselves, in the behaviour of their subject matter. Finally, Wallace (1985) claims that, while lawful schemes of universal development are seldom proposed in history outside Marxist circles, such an issue is still alive in psychoanalysis; a controversy over covering laws, however, is alive in both.

In his book, Wallace (1985) also raises the question of the relationship between theory and data in the historical sciences, which, we have learned by now, includes psychoanalysis. This relationship, he affirms, is “the issue that overarches all others” in *any science*, not only in the historical ones (p. 9). He presents two extreme positions on the matter: on one side, there is the understanding that data could assemble themselves spontaneously into coherent and accurate theories; on the other, that the theory determines the data, and that historians and analysts would engage in constructions rather than in reconstructions, inventing rather than discovering. Our psycho-historian opts for a middle-ground, and, perhaps inspired by the famous Kantian aphorism[60], utters: "If psychodynamics without phenomenology is sterile and reductive, phenomenology without psychodynamics if presentistic and chaotic" (p. 26). And, in the same vein: "No point of view and one drowns in one's data; too strenuous a point of view and one never tastes them" (p. 25). For him, historians and analysts have refrained from comprehending the interaction between theorizing and data gathering due to a caricature of the scientific attitude as primordially data-driven and to an anxious will to conform to this heedless but praised picture "long after natural scientists have abandoned it as illusory" (p. 10); however, many among such specialists have been advancing in this respect and recognizing that data can only be assembled to facts through the power of theory.

Theory would have the power to indicate what are the facts and, among the facts, which ones are important for the research. Abraham Melden (1969) and Hans Meyerhoff (1959) have observed this regarding history (Wallace, 1985). For them, a history which is just a chronological catalogue of past events, although one which all could agree upon, would be trivial and would explain nothing. Yet, according to Patrick Gardiner (1961), the theories guiding the historical research are composed of terms *open and vague*, terms indicating a range of unsharp borders and comprising a great number of events; a generalization such as "economic changes in society are accompanied by religious changes" gives to the historian who investigates religious changes hints of what factors to look for, but leaves open to him or her the task of finding out "the specific nature of those factors on that occasion and of the manner in which they are causally connected to one another" (pp. 93-94 as cited in Wallace, 1985, p. 49). Even if the explanation in history start deductive, its end and substance have an inductive character; that is to say, the specifics of the data impose themselves on the theory in the end, even if this theory have been what made historians look in the direction of this data at the start. Wallace (1985) argues that "the same holds true for the psychoanalytic clinician" (p. 49).

Without a theory – a set of core hypotheses – telling analyst-researchers to take a diachronic look at the subjects' behaviours or to interrogate the subjects' sexual and phantasmatic dimensions, for example, they would hardly obtain explanatory data. By the same

[60]"Thoughts without content are empty, intuitions without concepts are blind".

token, historians used to neglect economic questions before Marxist theory. We are thus sent back to our reflections on the selection bias in clinical psychoanalysis: the selectivity seen there is not a problem in and of itself. According to Wallace (1985):

> that historians and psychoanalysts must select from the data at their disposal does not involve them, *ipso facto*, in a subjectivist enterprise; selectivity, as Nagel (1963, p. 78) affirms, is necessary in all science. At the outset it is necessary to distinguish between several levels of selectivity. Many idealists write as if selection occurs at the level of initial data gathering. Certain types of data, they argue, simply will not be looked for or apprehended from the outset – because they are not considered significant to the problem at hand or within the domain of the writer's theoretical orientation.
>
> Indeed, such early selection often occurs in practice and, where it involves patently trivial matters, poses no problem. For example, the military historian will not normally concern himself with Napoleon's dessert preferences or the historian of psychoanalysis with Freud's penchant for mushrooms (p. 97).

The theory would help analysts not only to select the data coming from the patients, but also to organize this data; in return, this data would help them to select and organize the bits of theory to be used. Their process of "absorbing and reconstituting, observing and interpreting, is […] an inextricably interactive, mutually determining one […]. The interpreter imposes himself on the data no more than the data impose themselves on the interpreter" (Wallace, 1985, p. 21). This back-and-forth process would involve the practice, a time-honoured one in the historical sciences, of searching for "themes, patterns, and parallels" and of "drawing causal inferences therefrom" (Wallace, 1985, p. 36):

> Historians and psychoanalysts go over and back over their patient or texts in exquisite detail, saturating themselves in their sources. They encounter the same datum repeatedly – but often in slightly different contexts or connections with data already uncovered. Patterns and themes begin to appear, analogies are drawn, hitherto meaningless mounds of data are ordered; some of this is the fruit of the practitioner's preconscious and unconscious connections between, and elaborations upon, the material. From these arrangements, causal hypotheses will be constructed and tested (Wallace, 1985, p. 21).

The themes that emerge and repeat themselves are wild, free to be what they wish, but the background idea that repetitive, thus causally bound, themes will be found is in truth theory-dependent. It comes from the disputable idea of a human past determining the human future – of historical determinism. But who could say that such idea does not come from the datum of the recurring themes? Without the idea of historical determinism, "one would often overlook the form and content common to otherwise discrete and idiosyncratic events", but without the patterns "one would not be alerted to the necessity for a concept of historical determinism" (p. 36).

Anyway, historians and analysts would not be able to relinquish generalizations, even if what both aim to explain are non-replicable episodes and particular constellations. A problem would arise, on the other hand, in the cases where their theory is permitting them to ignore a great amount of relevant data. Concrete examples of such fault are, cites Wallace (1985), historian Gregory Zilboorg's turning a blind eye to refractory evidence relative to his hypothesis that all witches were either psychotic or hysterical, and Little Hans' father turning a deaf ear to the child's negative Oedipus. Psychoanalysis and history teach us that humans are many a time consistent and predictable, but also that "no two of us are human in precisely the same way" (Wallace, 1985, p. 51).

As the Oedipus was brought up, let us remark, with Wallace (1985), that the hypothesis is opportune to amend the relationship between the general and the particular in clinical psychoanalysis. At its bottom, the Oedipus is no simple hypothesis about every boy's tragic hetero-love and homo-hate, no strong heuristics connecting "man" to lust for mother-imago and to anger for father-imago, or "woman" to the corresponding opposite, but actually the vague hypothesis that any person (of any gender) enters (and shifts within) an unconscious dynamic of love and hate (and blends of both) with caregivers (one, two or more, of any gender) in a search for pleasure (be it direct or mediated by social rules), and that such dynamic determines this person's personality. The patients are the ones who should "tell" whether their Oedipus complex is in the classical, boy-centred version, or in any of the other numerous versions in which it may be manifested.

> […] the dynamic psychiatrist realizes that in even so apparently ubiquitous a phenomenon as the Oedipus complex the content of the unconscious fantasies and the mode in which they are expressed and defended against varies from person to person; moreover, one must reckon with variations such as negative complexes and the effects of cross-cultural differences in family structure and childrearing. The analyst's deduction regarding any particular individual's Oedipus complex do not obviate the need to inductively gather the data. Covering laws such as those relating the Oedipus complex to antecedent conditions and subsequent effects are, like the generalizations of the historian, approximate and open, and must be filled in with phenomenological and historical spadework concerning the concrete individual. The nexus of causes that produce adult personality structure and symptomatology is never identical in any two human beings (Wallace, 1985, p. 52).

From the case of the Oedipus hypothesis and from our previous points, we may jump to a recommendation – again, alongside Wallace (1985) – that analysts should open their minds to categories of themes that are not the most insistent in the theory – for example, to the institutional structure of the patient's work settings, to the patient's religious feelings through time, etc.. The crucial evidence may be found in such unsuspected domains instead of in the very-psychoanalytic domains of "mommy", "daddy", sex and anger, to name just a few. Analysts and historians should gather a "broad initial data base" in spite of embracing theory as a guide; they are "in danger of finding only what […] [they are] looking for" in case their selectivity begins too soon (p. 98).

Responding to the famous objection that (at least some) psychoanalytic inferences use theory to interpret data and then take the same data as support to the same theory, demonstrating thus to be circular and unwarrantable, Wallace (1985) argues that "the construct's *interaction* with the data" (p. 14) gains credence from its "explanatory power" and from "the convergence of multiple lines of evidence" in its favour (p. 15). He makes a similar statement regarding the hypotheses of "applied psychoanalysis": inasmuch as they bind many disparate facts together, the difficulty to devise quasi-experimental or quantitative paradigms to test them would not be that worrisome.

The last arguments we are bringing from Wallace (1985) are the ones regarding the peculiar concept of causality present in both psychoanalysis and history. The similarities between the disciplines are very conspicuous around such concept. Indeed, we find not only a historical causality in psychoanalysis, but also, so to speak, a psychoanalytic causality in history.

Wallace (1985) contends that the concept of causality in the inaugural movements of historiography was by and large a *psychological* concept of causality – "recall Thucydides' emphasis on character and motive as the mainsprings of history", he writes (p. 124) –, and that it has probably been the most prevalent concept of causality in historiography along with the theological one. The causes under the historians' pen would be, therefore, very similar to the ones that analysts find from the couch. Wallace (1985) cites in this context the historian Robin Collingwood (1946), for whom the work of his discipline consisted basically in *looking for the idea behind the past action*. For Collingwood, the historian wishes to know what did Brutus *think* that made him decide to stab Caesar, and not, like the natural scientist, on what kind of occasions would "Brutuses" stab "Caesares"; the historical cause would be, simply put, a *particular idea*. But Wallace (1985) considers Collingwood's concept of motivation "too rational, conscious and affectless" (p. 125) and prefers William Walsh's (1969). The latter explicitly compared the historian with the analyst by writing that the task of both is to reveal "carefully worked out plans behind apparently irrational actions" (p. 66, as cited in Wallace, 1985). A description of how a historian delineate causes, or "colligate", proposed by the same Walsh, is "virtually identical to what the analyst does" (Wallace, 1985, p. 125): in this process, historians would explain an event by contextualizing it and by detecting its intrinsic relationships to other events.

The psychoanalytic investigation of causes is, in its turn, imbued with a historiographic attitude. Wallace (1985) talks of a "historical determinism" active in psychoanalytic investigation, defining it as the postulate that the history of the actual and fantasied interpersonal relations of an individual is causally connected to the individual's current mental products; in other words, the postulate that the constitution of today's personality stemmed from the personality constitutions that have hitherto been established. A behaviour would be determined by "historical", rather than "current", factors, if determined by factors that, although current, have originated in specific conditions of the past; such active factors would be precisely the representations, or memories, of such historical conditions. He adds:

> The psychoanalytic theory of psychopathology as a maladaptive way of dealing with unconscious intrapsychic conflict is of course founded squarely upon the bedrock of historical determinism. Both the conflict and the maladaptive means of handling it arise from the interaction between one's biological endowment and nascent personality structure on the one side and the environment of the other. The symptomatic and characterological modes of coping with conflict are special cases of atavism – the persistence of once adaptive, historically determined modes of behaviour into the present, where they are no longer necessary and useful. Our "task", as Freud (1915-1917, p. 270) said, "is then simply to discover in respect to a senseless idea and a pointless action, the past situation in which the idea was justified and the action served a purpose" (Wallace, 1985, p. 178).

So, the idea of cause operative in historiography would also be operative in psychoanalysis, and vice-versa. We have seen that causes in psychoanalysis should be treated as at least INUS causes. Would the idea of cause be treated in history in a similar manner? Wallace (1985) would agree with us and answer "yes" to this. He discusses with care the place of free-will in the human sciences, especially by the end of his book, and he asserts that the question of whether human events plausibly seen as determined by the past could not have been otherwise, i.e., are *inevitable* events just because of their condition of having been determined, "is forever unanswerable" (p. 128): "any thesis for or against inevitability lies in the *post hoc explanation*, not in the unrecapturable events themselves [...]" (p. 129). Moreover, he compares Edward Carr's distinction between the immediate cause and the remote causes of a historical event, with the immediate cause demonstrating to be just a final, effective, point of a long, previous, chain of causes, to the analysts' aged distinction between the precipitating and the predisposing causes of a symptom. Such points indicate that Wallace thinks the causes stated in psychoanalytic and historical hypotheses to be far from necessary and sufficient for the occurrence of their corresponding effects.

In the same spirit, but heading from the ontological to the epistemological dimension, he then asserts that psychoanalytic and historiographic determinism would not imply predictability:

> It does not imply, whether in principle or practice, that human behaviour is predictable in more than a highly approximate and probabilistic manner. […] While accurate *retro*diction is often possible, absolute *pre*diction is impossible for at least two reasons: (1) we are never aware of all the relevant antecedent conditions – either within the person or in his environment; and (2) because the human capacity for creative synthesis (of the components of his psychological structure and function), while itself determined, can allow for novel and surprising responses (p. 129).

As a consequence of the complex character of the human domain, and considering that the human events to be explained by the scientists at issue have already happened, it is impossible to reproduce all of the relevant factors of such events. Therefore, both the analyst and the historian are unable to experiment in the controlled and prospective manner; in order to infer causes, they have to infer them retrospectively from "patterns, parallels, sequences and correlations with varying degrees of regularity" (Wallace, 1985, p. 201).

Finally, the two sciences could be brought together through the concept of causality not only for how similar this concept would be between them but also for how different the "causality" alive in both would be from the "causality" alive in physical and biological sciences. According to Wallace (1985), while for the latter the cosmos and the living things are moved by blind forces, for the former the human behaviour is moved by ideas, elusive things that represent potential and affect-laden states of affairs – goals and purposes – and that are somehow communicative. Nevertheless, all sciences, human or not, would take causes to precede effects and would see everything in their domain as caused, namely, as not coming to existence by mere accident or chance.

To conclude this interlude, I assent to Wallace's argument that, in spite of some variances between fully psychoanalytic and fully historiographic practices, psychoanalytic investigation is historiographic in essence. It is plain to me that "psychoanalysis is the pre-eminently historical psychology" (Wallace, 1985, p. 113). Now we are able to see that, were I to compare the epistemological problems and resolutions of psychoanalysis with the ones of any historical science, I would not be promoting any nonsense. Let us learn, then, from Carol Cleland's portrayal and justification of the inferential practices of geologists and palaeontologists.

5.4 THE EPISTEMOLOGY OF HISTORY: CLELAND AND OTHERS

5.4.1 The "Smoking Gun"

What do fantasies have to do with dinosaurs? We know that many people, especially during childhood, are very interested in dinosaurs and that a classical analyst could regard this Jurassic obsession as a symbol of many typical primal cravings; for example, a craving to learn about how life was before one was born or about one's "origins". But that shall not be the point of this part of the chapter. Instead, I shall present the inferential principles Carol Cleland (2002, 2011) argues to be ubiquitous in the minds of historical natural scientists with the hope of demonstrating that such principles are similar, if not identical, to the ones of the inferential model presented earlier in this chapter – the ones ubiquitous in the minds of clinical psychoanalysts. The inferential process that has hypothesized the extinction of the dinosaurs as having been the result of an asteroid impact is the main case to inspire Cleland's identification, refinement and justification of such principles. As is the case with an adult patient's particular childhood feeling, witnessing the extinction of the dinosaurs is (at least with today's technology) impossible. Nevertheless, both events are quite warrantable, as we shall now see.

In the book we have just discussed, Wallace (1985) writes that

> embryology and developmental psychology and the genetic sciences – history, psychoanalysis, archaeology, evolutionary biology, palaeontology, and historical geology – provide considerable evidence that psychological, social, biological, and geological structures emerge in an historically continuous manner from precursors whose vestiges they continue to bear (p. 81).

The principle of "historical determinism" he was arguing to be ubiquitous in clinical psychoanalysis, thus, could well come from beyond the *human*, *social* history – the one we knew as just "history" when at school – for there exist historical accounts of non-human or non-social stuff. We often forget there are *historical natural* sciences, or at the least natural sciences with historical branches. Geology, palaeontology and evolutionary biology would be among them – but also, surprisingly for some, planetary science, astronomy and astrophysics. It is not for lack of their products' exposure that most of us refrain from thinking of them as (also) historical sciences. The Pangaea and the Big Bang hypotheses are famous, and everyone is sure no scientist was ever the Pangaea's cartographer or has ever recorded the Big Bang with a camera. Yet, for some reason, the Pangaea and the Big Bang hypotheses are not usually seen as resulting from the same reasoning that gives us hypotheses about the domestic values current in the Roman Empire, for example. Those hypotheses seem to be children of experimentation, just because every knowledge produced by the natural sciences presumably is.

However, Cleland (2011) is forceful in her claim that “some of the most widespread practices of the historical natural sciences do not resemble those of stereotypical experimental science” (p. 552). But, if those practices are this distinctive and, as the celebrated hypotheses prove, recognizably successful, why have philosophers of science not been more interested in them? Maybe because the experimentalist wing of natural sciences sometimes expresses the prejudice that the support allowed by historical evidence is not as strong as the one capitalized by “good” science. A philosopher of science herself, Cleland (2011) defends, though, that “the findings of historical science are just as tentative and subject to revision as those of experimental science” (p. 554). One practice cannot be seen as more rational or objective than the other simply because one is moulded into dealing with one sort of data, and the other with a different sort of data (Cleland, 2002).

According to Cleland (2002, 2011), historical natural scientists work from puzzling traces of the present to their best explanation in terms of a longstanding, multifaceted and singular causation. As they would not be allowed, in both practical and ethical senses, to replay such causation, their work could never be reduced to an experimentalist endeavour. Holding that “the more improbable an association among a collection of traces[,] […] the more psychologically appealing the claim that it is the product of a common cause” (Cleland, 2011, p. 569), they would start their work by looking for, or bumping into, an odd and iterative structure. They would then seek to advance in the explanation of the structure by proposing a group of causal hypotheses, all of which presupposing the structure’s instances to integrate one and the same causal chain. Finally, they would look for, or bump into, some instance which proves to extend or enrich the structure, substantiating thus an inference over which is the most probable of the lot of causal hypotheses – in other words, over which would best explain the whole evidence in hand and include such evidence in a consistent story. Cleland (2002, 2011) calls this kind of illuminating instance a “smoking gun”. The high iridium levels in the K-T boundary, she argues, was a “smoking gun” for the investigation over the end-Cretaceous mass extinction: only the hypotheses of asteroid impact and volcanic eruption could explain those levels. Cleland (2011) also remarkably holds that, in history, explanations could not be potential predictions, as in the Hempelian understanding, for in long intervals of time we are dealing with “high degrees of contingency” (p. 566). Affinities with the inferential framework of clinical psychoanalysis could not be more explicit. Let us now detail each one of Cleland’s points.

Basically, historical natural scientists infer through the following interrelated steps: (1) they promote a copious generation of "competing hypotheses" that could explain a puzzling body of traces they have found in fieldwork; (2) they search for a smoking gun "to discriminate among them" (Cleland, 2011, p. 554). It should be stressed that by "competing hypotheses" she means "mutually exclusive hypotheses" (Cleland, 2002, p. 480). Let us remind this is not an epistemological demand for the clinical psychoanalyst, for some or all competing hypotheses about a patient devised by an analyst (particular hypotheses) are usually consistent with each other. Clinical evidence entitles the analyst to change each hypothesis' degree of likeliness (according to how much each hypothesis explains the whole evidence), but hardly to attribute likeliness close to zero to every hypothesis but one.

Once the intriguing traces are found, and the competing explanations to them are elaborated, the aim of such scientists becomes to seek for a "smoking gun" (Cleland, 2011). A smoking gun is a yet-to-be-found trace, or subcollection of traces, that would show to be congruent more with a fair number of the competing explanations that scientists have elaborated for the traces already assessed than with most of these explanations – in the best scenario, it would show to be significantly and relatively congruent with just one. It "cinches the case for a particular causal story" (Cleland, 2011, p. 554). Although the quest for a smoking-gun trace is the hardest part of historical research, it would be always possible to discover one for a past event, even when this past is a very remote one and the traces of the event have become very attenuated (Cleland, 2002).

Cleland (2002) stresses that a piece of evidence can only fulfil the function of a smoking gun relatively to a group of specific competing explanations. It could never contradict an isolated hypothesis, and, for this, "there is little in the practice of historical science that resembles what is prescribed by falsificationism" (Cleland, 2002, p. 483). Moreover, a trace could cease to be regarded as a smoking gun with the advent of new evidence and new explanatory hypotheses (Cleland, 2002).

Because the consensus around the Alvarez hypothesis became strong and cross-disciplinary, Cleland (2011) takes it as "a particularly compelling case study for illustrating [...] [an] account of the methodology of prototypical historical natural science" (pp. 557-558). Before 1980, explanations for the extinction of the dinosaurs abounded; among them were the scenarios of contagion, climate change, volcanism and asteroid impact. Then, father and son Luis and Walter Alvarez, in analysing the K-T (Cretaceous-Tertiary) boundary, discovered that it contained an unusually high level of the element iridium (more than 33 times higher than background levels). It was a smoking gun because that much iridium only exists in the Earth's mantle and in asteroids: global volcanic activity and asteroid impact became, then, the mechanisms most probable to bring the kingdom of the terrible lizards to an end.

Three different and complementary smoking guns followed. Bohor, Foord, Modreski and Triplehorn (1984) discovered, in the same K-T boundary, extensive deposits of shocked quartz with specific qualities, a kind of quartz which on Earth is only found, as Alexopoulos, Grieve and Robertson (1988) established some years later, in asteroid craters and places where there was a nuclear explosion (Cleland, 2002). Looking for a crater of the right size and age, geologists found one: the Chicxulub Crater, in Mexico, with 65 million years old and 200-300 km across (Cleland, 2002). The philosopher reminds, though, of the many geologists admitting then "that failure to find the crater wouldn't count heavily against the hypothesis", for "the active geology of Earth might have obliterated all traces, particularly if the impact had occurred in deep ocean" (Cleland, 2002, pp. 482-483). But, while the evidence could establish that the impact had occurred, a causal link between the impact and the extinction remained uncertain. Finally, Ward's (1990) meticulous study of the fossil record of the ammonites showed that the Cretaceous extinctions were rapid and global, which greatly enhanced the likelihood of the hypothesis that an asteroid impact was their primary cause. In sum:

> The asteroid-impact hypothesis became the widely accepted explanation for the extinction of the dinosaurs. For of the available hypotheses, it provided the greatest causal unity to the diverse and puzzling body of traces (fossil record of the dinosaurs, fossil record of the ammonites, etc., and iridium anomaly, shocked quartz, Chicxulub Crater, etc.) (Cleland, 2002, pp. 482-483).

This is the main case inspiring Cleland (2002) to delineate the basic itinerary of historical natural scientists. Therefrom she realizes that such itinerary contrasts in significant ways with the itinerary of experimental scientists. For starters, historical sciences would turn their heads toward long-past, token events, while experimental sciences would be all about regularities among event-types. Further, while experimental researchers would artificially reproduce an event-type in a laboratory to obtain crucial evidence, historical researchers would obviously not be able to do the same with their particular past events; in searching for a smoking gun, they are "stuck with what nature has already provided" and "must be clever enough to ferret out incriminating evidence that is often well hidden in the messy, uncontrollable world of nature" (Cleland, 2002, p. 486). Finally, experimentation heads from causes to effects, and its task and challenge is to artificially prevent false positives and false negatives, whereas *ex post facto* investigation walks in the opposite direction and focuses just on ruling out alternative explanations (Cleland, 2002).

5.4.2 Predictions and Laws in the Historical Sciences

Explanations, indeed, are the protagonists throughout the process of acceptance and rejection of historical hypotheses, according to Cleland's model. She argues, moreover, that prediction and retrodiction play no role in such process, for historical explanations express causal, not logical, relations, and the Hempelian model is wrong in assuming that every explanation is a lawful calculus and a prediction in reverse (Cleland, 2011).

Only in rare cases would hypotheses about bygone token events rely on successful predictions in order to be accepted. No one had demonstrated that the iridium anomaly in the K-T boundary is a logical consequence of the meteorite hypothesis before the Alvarezes discovered it: they were not expecting it and simply "stumbled upon it while exploring a different question" (p. 558). Around the time Cleland published her paper, in 2011, we still had no knowledge to predict the anomaly from the hypothesis:

> Our current understanding of earth and planetary science informs us that there are just too many highly plausible, extenuating circumstances capable of defeating an inference to an iridium anomaly from a gigantic meteorite impact, e.g., an iridium-poor meteorite, dispersal of an initial iridium anomaly by geological processes, and unrepresentative samples of the K–T boundary (exposed outcrops of which are rare) (Cleland, 2011, pp. 558-559).

The motive for Ward's ammonite studies, too, was not a prediction, for the Alvarez hypothesis and the knowledge available for him before he discovered about the ammonite extinction did not logically entail that the ammonites must have gone extinct with the dinosaurs by the end of the Cretaceous (Cleland, 2011).

By the same token (tongue in cheek), we could devise circumstances in which a rival hypothesis over the cause of the K-T extinction, say, the contagion hypothesis, would be compatible with the current existence of an iridium anomaly in the K-T boundary, and so conclude that this evidence has no power to refute, strictly speaking, that hypothesis:

> as the scientists involved would readily admit, the presence of iridium in the context of their background understanding of Earth history does not provide evidence that the dinosaurs did not go extinct as a result of an epidemic shortly before or after the [meteorite] impact (Cleland, 2011, p. 562).

The evidence uncovered by the Alvarezes was *just* a strong index that, in the Cretaceous, there occurred a meteorite impact, or volcanism in a global scale, or both, and with more background knowledge we are able to know that, *under the right circumstances*, one of the two could alone have produced a mass extinction.

It is not surprising that historical scientists never definitely rule out a hypothesis when faced with predictive failure. Besides having to deal with "an enormous number of worrisome auxiliary assumptions given the length and complexity of the time spans involved" (Cleland, 2011, p. 562), they cannot, like experimental scientists, control for them in a lab. Many factors interfere between a historical cause and a historical effect, and no time-machine has yet been invented. As historical scientists cannot experimentally specify conditions for testing and evaluating target hypotheses, all they have are prognostications serving only

> as tentative guides – educated guesses, based informally upon both theoretical and empirical background knowledge – about where additional evidence (ideally, a smoking gun!) *might* be found for a hypothesis and perhaps even what form it *might* take. […] the Alvarezes' discovery of an iridium anomaly may be interpreted as guided by an extremely vague (tacit) prediction. Walter Alvarez took samples from the K–T boundary because, like many geologists, he believed that crucial evidence for what caused the end-Cretaceous extinctions might be found there even though no one at the time had any idea what form it might take. Walter Alvarez took samples from the K–T boundary because, like many geologists, he believed that crucial evidence for what caused the end-Cretaceous extinctions might be found there even though no one at the time had any idea what form it might take (2011, p. 563).

Historiography, thus, would not be able to provide precise predictions, such as what will cause the next mass extinction on Earth, or which kind of political situation will host the next communist revolution. A similar statement about an asymmetry between explanation and prediction can be serendipitously seen in one of Freud's clinical cases:

> So long as we trace the development from its final outcome backwards, the chain of events appears continuous, and we feel we have gained an insight which is completely satisfactory or even exhaustive. But if we proceed the reverse way, if we start from the premises inferred from the analysis and try to follow these up to the final result, then we no longer get the impression of an inevitable sequence of events which could not have been otherwise determined. We notice at once that there might have been another result, and that we might have been just as well able to understand and explain the latter. The synthesis is thus not so satisfactory as the analysis; in other words, from a knowledge of the premises we could not have foretold the nature of the results (Freud, 1955a, p. 167).

Indeed, now and again scientists obtain a significant comprehension of the causes of an event without gaining with it the power to predict its occurrence: past earthquakes, for example, can be explained "in remarkable detail" but future ones cannot be successfully predicted (Cleland, 2011, p. 567).

According to Cleland (2011), the meteorite hypothesis is, *not the only, logically imposing, explanation* for the evidence of that mass extinction – which would logically authorize predictions related to this kind of evidence and to similar mass extinctions – but just *the best of a set of available explanations, given the available evidence*. She claims that "the reasoning involved [here] is that of inference to the best explanation, which does not have the same logical structure as retrodiction or prediction" (Cleland, 2011, p. 559). The scientific practice she is describing rejects a hypothesis, not because it is inconsistent with some evidence, but because “another hypothesis does a much better job of explaining the total body of evidence available” (Cleland, 2011, p. 562).

Such arguments lead Cleland (2011) to criticize the universality of Hempel’s deductive-nomological model of explanation. According to the model, to explain a phenomenon is to find the general law and the initial conditions which, in conjunction, would deductively entail the occurrence of the phenomenon; thus, an explanation would simply consist in a deductively valid argument, and it would be possible to predict the occurrence of the phenomenon in case the same conditions were identified. However, argues the philosopher, a scientific law should not be contravened by any instance of the domain it applies to, and “explanations in the historical sciences rarely invoke even rough generalizations of this sort” (Cleland, 2011, p. 564). This happens because, as we have seen, the chain of factors between a past event and its contemporary vestiges is too long, complex and unpliable to be entrapped by a law: “scientifically compelling statistical or probabilistic laws require reliable information about frequencies, which is rarely available, particularly in cases involving uncommon events such as mass extinctions” (Cleland, 2011, p. 564).

5.4.3 Common Causes: Cleland, Sober and Tucker

As expected, many epistemologists had preceded Cleland in the effort to describe how historians infer hypotheses from traces. Roughly, epistemologists would recognize two manners through which historians do it: on one side, through a narrative heuristic, and on the other, through common cause explanations. Historians could accept a hypothesis because it makes a *good story* out of the evidence available, placing it by omissions and additions into a *coherent* and *continuous* sequence of events, or they could accept a hypothesis because it was generated through reliable methods for establishing when a whole of present traces is, by statistical and theoretical forces, the result of one and the same factor in the past. Cleland (2011) characterizes a narrative explanation with no concern for empirical warrant detrimentally as something “quasi-fictional”, a “just-so” story; this would be the reason why “narrative explanations and common cause explanations frequently go hand-in-hand in the historical natural sciences” (p. 567). On the same line, Tucker (2004) contends that historiography is no mere story-telling and that “what turns hypotheses into knowledge is not their position in a narrative, but their relation to the evidence” (p. 139).

The expression “common cause” comes from Hans Reichenbach’s “Principle of the common cause”, which states that “if events X and Y are correlated, then either X caused Y, Y caused X, or X and Y are joint effects of a common cause" (Sober, 2001, p. 331). It is this principle plus some evidence of two or more correlated events which would justify common cause explanations.

Cleland (2011) claims that the principle is what sustains the close relationship between explanation and confirmation found in the reasoning of historical natural scientists. We should note, though, that she takes the correlation (or similarity, or association) mentioned by the principle to be a seemingly or ostensibly *improbable* one, and we should also note that this condition is not necessarily implied in the original principle. If we have this addendum in mind, we understand why she claims that the principle is a powerful explanatory tool. Ascribing an ostensibly improbable association among traces to chance would explain nothing about it, while taking it as a segment of a single causal network would allow us to grant that the association is not improbable after all (Cleland, 2011). Therefore, it is expected that historical natural scientists tend to channel their inquiries to whatever "seems in light of their background beliefs to be the most unlikely (and hence puzzling) correlations or similarities among contemporary phenomena", for "the more improbable an association […] seems the more psychologically appealing the claim that it is the product of a common cause" (Cleland, 2011, p. 569).

In order to convince us that such scientists are right in acting thus – not only logically but also *metaphysically right* – the philosopher convokes a "thesis about the nature of the world for which there exists overwhelming empirical evidence" (Cleland, 2011, p. 570). She brings from David Lewis' (1979) "Counterfactual Dependence and Time's Arrow" the idea that our universe displays an asymmetry of determination along causes and effects, which would provide "a nonlogical, objective foundation for the epistemic principle of the common cause" (p. 571). The thesis would state that "most localized cause and effect relations in our universe form many-pronged forks opening in the direction from past to future" (pp. 571-572). Because of this, causes would epistemically underdetermine effects, while effects would epistemically overdetermine causes; that is, while the identification of a few factors is rarely enough for us to be certain that a specific phenomenon shall develop from it, the identification of a few vestiges can be perfectly enough for us to know what *must* have happened in the past for these *specific* vestigial components to be found today. Thus, counterfactually:

> If the temporal structure of causal relations in our universe were different – if most causal forks opened in the opposite direction (from future to past), or most cause and effect relations were linear (one-to-one) instead of fork-like, or most events were chance (uncaused) occurrences – one would not be justified in inferring the likelihood of a common cause from an ostensibly improbable association among traces (2011, pp. 571-572).

She asks us to consider a volcanic eruption, whose effects include "extensive deposits of ash, pyroclastic debris, masses of andesitic or rhyolitic magma, and a large crater", and observes that "only a small fraction of this material is required to infer the occurrence of the eruption"; in contrast, to infer future eruptions would be "much more difficult" (Cleland, 2011, pp. 570-571). If a smoking gun is nothing but a decisive evidence of a specific common cause, the fact that specific events ordinarily spread a large range of specific effects out to the future would advise us that smoking guns are somewhere around us, that the sights and sounds of present times are likely to contain a good amount of them (Cleland, 2011). The dicta "there is no perfect crime" and "a lie has short legs" are popular expressions of such metaphysical assumption.

The philosopher grants that separate, independent causal processes sometimes produce puzzling, improbable correlations, and that the then-clause in Reichenbach's principle should also gain a probabilistic tone ("if events X and Y are *abnormally* correlated, then *probably*…"). Her argument, however, is that Lewis' asymmetry authorizes historical scientists to divert from this evil demon:

> given the global reach of the asymmetry of overdetermination, it is rational for historical scientists to opt for common cause hypotheses over separate causes hypotheses *in the absence of theoretical or empirical reasons for believing that a specific seemingly improbable association among traces is the product of separate causes* [emphasis added]. That is, one would expect common cause explanation to be the *default mode* [emphasis added] of evidential reasoning in historical natural science (2011, p. 574-575).

This corollary leads Cleland to see herself in disagreement with the corollary about common causes defended by the philosophers of historical sciences Elliot Sober – the same Elliot Sober used in Grünbaum's thematic-affinity argument (see p. 28) – and Aviezer Tucker. The philosophers would defend that the default position before an improbable association should be one of *incertitude* unless there are – *positively* – strong reasons to favour a common cause hypothesis over a separate causes hypothesis (Cleland, 2011).

Sober (1987) famously devised a counterexample to the Principle of the Common Cause: if the rise of sea levels in Venice as well as of the bread cost in Britain had been monotonic in the past two centuries, it would not be plausible, given the rest of the beliefs we all usually hold, to explain the observed correlation between the variables with the hypothesis that some yet unknown variable is causing both rises; we would regard the sea levels and the bread prices, rather, as increasing by local and independent conditions. The philosopher discusses an easy reply to his point (Sober, 2001). The example would refute Reichenbach's principle, but not a slightly altered version of it, which would read as follows: "if events X and

Y are correlated *and so are changes in X and changes in Y* [emphasis added], then either X caused Y, Y caused X, or X and Y are joint effects of a common cause" (Sober, 2001, p. 335). With another counterexample, not an invented one this time, he diagnoses this weaker version as equally liable to error.

Developmental stages and their timings strongly correlate with one another between any pair of members of the human species. This is because humans inherited the same "developmental program" from a common ancestor; the correlation is thus a *homology*, like the correlation of wing-shapes between robins and sparrows. But what if we found similar developmental sequences, Sober (2001) asks, between organisms of different species?

> It is possible that this, too, is a homology, but must it be? The answer is *no*; it is perfectly possible that similar developmental sequences evolved independently in the two lineages.
> […]
> […] If 'common cause' means common ancestry and 'separate cause' means separate ancestry, then […] [the altered version of the Principle] will be false (pp. 336-337).

It is not rare that similar features between species not be due to a common ancestry; scientists call this conjunction a "homoplasy". The wings of a robin share anatomical properties (location, shape, function) with the wings of a bat, but the two flying animals do not share a history of mutations and natural selections; bats evolved from wingless mammals, robins, from other winged birds. There we find a homoplasy. The fins of dolphins and fish are also a homoplasy.

Although Sober (2001) concludes that both the original and the altered versions of the principle of common cause are not reliable shortcuts for causal inferences, he grants that, if two changes are introduced into the original version, "we can identify the kernel of truth that […] [it] contains" (p. 343). First, faced with an observed association, we should put its common cause explanation beside its separate causes explanation: "[…] one cannot decide whether a common cause explanation is well supported by a correlation unless one knows what the competing hypotheses are" and "this contrastive element is absent in the […]principles I have considered" (p. 343). Second, we should compare the probability of the observed association given a common cause explanation – Pr(O|H1) – with the probability of the observed association given a separate causes explanation – Pr(O|H2). Finally, if the former probability is greater than the latter – Pr(O|H1)>Pr(O|H1) – "the observed association favors, but does not conclusively prove, the common cause hypothesis" (p. 343). A very sober conclusion, indeed.

Cleland (2011) disapproves of Sober's core argument because, according to her, "it is important to distinguish purely numerical correlations (such as rising British bread prices and Venetian sea levels), from truly puzzling associations" (p. 577). Perhaps the deflection between the two philosophers comes down to their interpretation of the principle. Cleland is charitable with the principle originated under Reichenbach's pen in adding the "puzzling" element to it, and Sober fails to see an element like this in there – indeed, the classical principle is as crude as Sober sees it. But, details aside, we could say that Sober's "repairing" of the principle is very much consistent with Cleland's point, for, according to her, "scientifically puzzling" could mean "seemingly improbable" (Cleland, 2011, p. 578). We could say something on the same lines about her alleged disagreement with Tucker (Cleland, 2011).

In his book "Our Knowledge of the Past: A Philosophy of Historiography", Tucker (2004) argues that the way historians infer how much a hypothesis can be true in relation to emergent evidence was already established long ago by Bayes' Theorem. Tucker (2004) contends that "an interpretation of Bayesian logic is the best explanation of the actual[61] practices of historians [of society, nature and language]" (p. 96). The famous theorem developed by the reverend Thomas Bayes and published posthumously in 1763 was an initiative to track the degree of probability that evidence can confer on a hypothesis. Tucker (2004) cites Laudan's comment that, because the problem of all scientists, including historians, is at bottom that addressed by Bayesian logic – the problem of warranting claims from partial evidence – investigations into the past would not be epistemically unique (Laudan, 1992, p. 65, as cited in Tucker, 2004, p. 96).

The full version of Bayes' Theorem is this:

$$Pr(H|E\&B) = \frac{Pr(E|H\&B)\ x\ Pr(H|B)}{Pr(E|B)}$$

"Pr(H|E&B)" means "the probability (Pr) of a hypothesis (H) given (|) some evidence (E) and background knowledge prior to the evidence (B)", or "the posterior probability of the hypothesis"; "Pr(E|H&B)" means "the probability of the evidence given the hypothesis and background knowledge prior to the evidence"; "Pr(H|B)" is the prior probability of the hypothesis; finally, Pr(E|B) expresses the probability of the evidence given background knowledge, or the "expectancy". In the context of historiography, "E" usually indicates the "similarities between two or more independent sources such as documents, material remains, species, and language", and "B", "established theories and historiography, prior to knowledge of the evidence […]" (Tucker, 2004, pp. 96-97).

[61]Tucker's approach toward a model of historiographic inference seems "propositional-synthetical": "We should note that though Bayesian analysis can explain most of what historians do and how they reach an uncoerced, heterogeneous, and large consensus […], there are occasional deviations from the Bayesian ideal"; over such deviations, however, "[…] there is never a consensus" (Tucker, 2004, p. 139).

The Bayesian formula describes that, if we find evidence for a hypothesis we hold, the probability that our hypothesis is true given all we know then – Pr(H|E&B) – is equal to: how much our evidence is consistent with our hypothesis plus the knowledge we had before we encountered the evidence – Pr(E|H&B) –, times how much the hypothesis was already supported before, or independently of, the evidence – Pr(H|B) –, divided by how frequent would be the evidence independently of our hypothesis – Pr(E|B). The formula tells us that the posterior probability of the hypothesis will be higher than the prior probability of the hypothesis given some evidence *only if* Pr(E|H&B) is higher than Pr(E|B). It also tells us that the probability that our hypothesis is true increases with the value of Pr(E|H&B) *only if* the prior probability of the hypothesis is higher than Pr(E|B).

This last element relates to the discussion made in this chapter over the specificity of thematic affinities in mental products: the more specific (rare) the thematic affinity in normal circumstances – the lower the value of Pr(E|B) – the more we have reasons to conclude that it is part of a motivational structure – the higher the value of Pr(H|E&B).

Tucker (2004) helps readers suffering from matemaphobia to understand the formula by telling them to

> imagine all the possible occurrences of a piece of evidence based on all that we know of the world, and then ask in what percentage of these cases a hypothesis H is the case, and then compute the ratio of such cases (p. 98).

The critics of Bayesianism, in particular of its pertinence for understanding the gears of scientific inference, argue that "it is difficult to quantify and compute the expectancy of the evidence" (Tucker, 2004, p. 99). We should add: let alone to quantify and compute the other values of the formula[62]. Another charge against classical Bayesianism, he tells us, comes from the historiography of science: as scientists would "rarely consider the merits of a hypothesis in isolation" (Tucker, 2004, p. 99), the reverend's formula would not reflect their actual reasonings. Some epistemologists suggested, then, that the formula could be adapted to serve an appraisal of the *relative* probabilities of a set of competing hypotheses. Tucker (2004) cites Sober as one who interpreted the comparison of likelihoods of a set of rival hypotheses as the election of which one among them *best explains* the evidence, an interpretation of Bayesianism "in the tradition of Peirce (1957) who proposed that scientists do not evaluate explanations according to ideal independent criteria but compare them to each other in what he called

[62]We have seen that the notion of specificity implicitly used clinical psychoanalysis is an intuitive one, but this fact is not exactly good nor immutable. On the other hand, it is not exactly a bad sign that other sciences are still stuck with intuition around important matters as much as clinical psychoanalysis.

'abduction.'" (p. 99).

In his book, Tucker (2004) shows that historians evaluate their common cause hypotheses through this peculiar Bayesianism. In commenting on Reichenbach's principle, and again referring to Sober's arguments, he calls attention to the fact that the principle does not distinguish

> the inference of *the properties of a particular common cause*, from the inference that *there must have been some common cause* without characterizing it further. […] It is one thing to say, as Darwin did, that man and apes had some common ancestor. It is quite another to infer descriptions of some of the properties of that ancestor, as contemporary evolutionary biologists do (pp. 102-103).

After *the hypothesis* that there must have been *some* common cause makes the evidence probable more than the hypothesis of separate causes makes it probable – after *some* common cause hypothesis explains the evidence best –, historians generate *a host of hypotheses* about the *properties* of that common cause. And, again, the one among them that makes the evidence more probable is the one that explains it best. Tucker (2004) describes two stages inside this second stage: a first where the historian defines the general structure of the causal net in question (whether it comprises ancestral common causes, mutual influences between the segments of the evidence, or a combination of both cases) and a second where the historian defines its particular properties. Such stages are almost identical to the steps of psychoanalytic inference described earlier in this chapter.

It is worthwhile to go through some of the examples Tucker (2004) provides to show that historians evaluate common cause hypotheses in a Bayesian-Explanationist fashion. Before that, let us be aware that Tucker uses the term "variational group" in place of "thematic affinities":

> [Walter Wilson] Greg called sets of texts that are considered similar or correlated according to at least one property "variational groups". I adopt this useful term. Variational groups are founded on temporally "horizontal" correlations between members of a group that share certain variables that others outside the group do not (p. 105).

One variational group that historians have thought through is the one of the great pyramids that both Egyptians and Aztecs have erected once upon a time. Would this similarity have a (token) common cause? Had there been any cultural contact between Egyptians and Aztecs, or between their cultural precursors, that could be the cause of the coincidence that pyramids were constructed on both sides of the Atlantic? The likelihood of cultural contact given pyramids, Pr(H|E&B), is extremely low, for two complementary reasons: Pr(H|B) in this case is extremely low, as there had never been another kind of evidence in favour of this contact;

Pr(E|B) in this case is much higher than Pr(E|H&B), as the hypothesis of separate cultural developments would explain this variational group much better than the hypothesis of contact:

> given a level of building technology that does not include columns and arches, and the laws of physics, the only likely shape of a big and tall building is that of a pyramid, a solid wide base that can support a lighter structure above it. This is the only type of solution that ancient Egyptian and Aztec architects could have devised for this problem (p. 112).

The existence of a variational group whose shared property has "no conceivable functional value, or even confer a disadvantage on their bearers" (Tucker, 2004, p. 114) would radically decrease, in general, the likelihood that it is due to separate causes. Tucker's examples here are grammatical mistakes in texts and species' nonadaptive traits (which Darwin called "rudiments"): "the likelihood of any single grammatical mistake or nonadaptive trait is low; the likelihood of identical ones given separate causes is very small indeed" (p. 114). In these cases, the probability that a grammatical mistake recurs along different versions of a text because these versions were copied from the same source (or because one of the versions served as a source for the others), or that certain nonadaptive trait recurs in different species because they share common ancestry (or because one evolved from the other), thus, is demonstrably high. "For this reason,", he observes, "geneticists and textual critics alike cherish them" (p. 114). No problem in transposing this to clinical psychoanalysis: as we saw, analysts depart exactly from a variational group of (statistically) abnormal – in other words, highly specific – behaviours, behaviours that are usually painful for the patient, to claim that that group is causally fastened.

Tucker (2004) also discusses common-cause-*versus*-separate-cause hypothesizing when words in different languages sound and mean the same. This similarity could be the effect of the influence of one language on the others, of a lost language on all of them, or of the mutual influence of all of them on one another – of a common causation – or it could be the effect of independent developments – of a separate causation. It is not unusual to find phonetic similarities across languages no matter what their origins are. On behalf of this claim, Tucker (2004) cites Cristopher Hitchcock, who notes that, although the English "mama" is found with the same sound and meaning in Mandarin Chinese, it is "among the simplest sounds for infants to make", so this repetition "would be due not to the common origin of the languages in question, but rather to the similar circumstances in which the words enter the language" (Hitchcock, 1998, p. 444). Tucker (2004) concludes that, "unless there is evidence for historical contact between languages, phonetic similarities between a few words are more likely given the separate causes hypothesis than given some common cause hypothesis" (pp. 116-117).

According to Cleland (2011), both Sober and Tucker think that the principle of the common cause is "highly dubious", that "it is a mistake for historical scientists to favor common cause explanations over separate causes explanations unless they have information specific to the case at hand favoring one over the other" (p. 576); herself, on the other hand, thinks that when faced with an improbable coincidence such scientists "exhibit a general preference all other things being equal (in the absence of empirical or theoretical information suggesting otherwise) for common cause explanations over separate causes explanations" (p. 553), and that they are right in having this preference. She sees a significant difference there. However, as we have seen, all three philosophers would agree that an improbable correlation indicates in general that a common cause hypothesis is more probable. Given that, moreover, the three philosophers praise the IBE model and ascribe it to the investigative practices of historical scientists, we can conclude that their arguments are not that different from one another. I would say the three are equally sceptic, and that the only difference between them is one of approach: Cleland wishes to describe a heuristic of the investigative practices in question, while Sober and Tucker submit these practices to Bayesian rigours.

5.5 INFERENCE TO THE BEST EXPLANATION: LIPTON

Whatever the differences between Cleland, Sober and Tucker, all three agree that, in the research of natural and social historians, the inference of hypotheses about the available evidence comes down to generating a list of explanations for it and then picking the best one among such explanations; in the first part of this chapter, we have argued that a procedure such as this is also found in clinical psychoanalysis. This model of inference is called Inference to the Best Explanation ("IBE"). In this section, hand-in-hand with the philosopher Peter Lipton, we shall define it with care and show it to be consistent with Bayesian and Millian canons.

In the classical model of how humans make their inferences, inferences come before explanations. According to the model, we would first make a bunch of inferences; afterwards, if we needed an explanation for some phenomenon, we would search for one in that pool of inferences; and, if the pool did not contain the needed explanation, we would aspire to make new observations and new inferences. This seems plausible, for an information should be correct before serving as an explanation. But, according to Lipton (2004), this picture earnestly undervalues the role of explanations inside our day-to-day inferences.

Explanatory considerations would tell us what to look for and whether we have found it. Lipton (2004) invites us to bring to mind cases of "self-evidencing explanations", as when "tracks in the snow are the evidence for what explains them, that a person passed by on snowshoes" (p. 56):

> In these cases, it is not simply that the phenomena to be explained provide reasons for inferring the explanations: we infer the explanations precisely because they would, if true, explain the phenomena. Of course, there is always more than one possible explanation for any phenomenon – the tracks might have instead been caused by a trained monkey on snowshoes, or by the elaborate etchings of an environmental artist – so we cannot infer something simply because it is a possible explanation. It must somehow be the best of competing explanations (p. 56).

This outstanding role of explanatory considerations in inference, thus, calls for "a new model of induction", one that binds the two terms "in an intimate and exciting way" (Lipton, 2004, p. 56). The IBE model establishes that *first we try to explain* the evidence with the help of background knowledge; *then we infer* something as true or correct about the evidence. But what do we infer? We claim that the best candidate to explain the evidence is exactly the case. In sum: we are guided by the idea that, if a proposition explains the evidence best, then it must be true or correct. "Far from explanation only coming on the scene after the inferential work is done, the core idea of Inference to the Best Explanation is that explanatory considerations are a guide to inference" (Lipton, 2004, p. 56).

Lipton (2004) argues that this idea of explanatory considerations serving as a medium to the knowledge of what is the case only makes sense if IBEs are implicitly taken to be ILPEs – Inferences to the Loveliest Potential Explanation. With this, he purports to establish what the words "best" and "explanation" should mean when we say that a particular inference is an IBE. Explanations can be *actual* or *potential*, but IBE cannot be Inference to the Best of the *Actual* Explanations, because then the model would fail to account that: some of our inferences can reasonably be false; the explanations for a phenomenon compete with each other and are normally incompatible (so they cannot all be true); we should have a way to get to actual explanations, not presume we have them already. "Telling someone to infer actual explanations is like a dessert recipe that says start with a soufflé" (Lipton, 2004, p. 58). Thus, in IBE, the inference is "that the best of the available *potential* [emphasis added] explanations is an actual explanation" (p. 58).

Likewise, the "Best" in "IBE", according to Lipton (2004), should mean "loveliest" rather than "likeliest". The likeliest explanation would be the one that is best supported by the evidence and the loveliest, the one which would, if correct, provide the most understanding[63]. Again, the IBE model should discriminate the principles that lead us to the likeliest explanation,

[63]Lipton (2004) comments that "the criteria of likeliness and loveliness may well pick out the same explanation in a particular competition, but they are clearly different sorts of standard" (p. 59). He provides an example of a likely but unlovely explanation – "smoking opium puts people to sleep because of its dormative powers" (p. 59) – and another of a lovely but unlikely explanation – any conspiracy theory, where "many apparently unrelated events flow from a single source" (p. 60).

not presume we have already reached it. So, in order for IBE to be a genuine "attempt to account for epistemic value in terms of explanatory virtue" (p. 61), it should be a synonym of Inference to the *Loveliest* Potential Explanation.

> This version claims that the explanation that would, if true, provide the deepest understanding is the explanation that is likeliest to be true. Such an account suggests a really lovely explanation of our inferential practice itself, one that links the search for truth and the search for understanding in a fundamental way (p. 61).

In considering five of the main models settled in philosophy of what providing an explanation consists in – providing reasons for belief, turning what is strange or surprising into something familiar, making a deduction from a law, showing how diverse things fit together, and showing that things had to happen the way they did – Lipton (2004) argues that "none of […] [them] gives an adequate general description of our explanatory practices" (p. 29). Recognizable explanations may provide none of such things and a statement may provide all of them and not be recognizable as an explanation. The philosopher, thus, pursues a sixth model to attach to IBE – the causal model.

> According to the causal model of explanation, to explain a phenomenon is simply to give information about its causal history (Lewis 1986) or, where the phenomenon is itself a causal regularity, to explain it is to give information about the mechanism linking cause and effect (p. 30).

It would be "natural and plausible" (Lipton, 2004, p. 30) and it would avert many of the problems ingrained in the other models. For example, it would accommodate self-evidencing and non-lawful explanations, which, respectively, the reason model and the deductive-nomological model would not (Lipton, 2004).

An objection raised against the causal model is that, as "causal histories are long and wide" (Lipton, 2004, p. 32), most of the information about an event's causal history is not considered a good explanation for it. For example, "the big bang is part of the causal history of every event, but explains only a few" (Lipton, 2004, p. 32). Therefore, the model would be too permissive, underdetermining our explanatory practices. Lipton (2004) sees the objection as a challenge and offers "a modest development of the model" (p. 29) to make it a smoother mirror of our explanatory practices and, consequently, to make a smoother mirror of (an enriched) IBE.

The problem of which causes of a phenomenon we should select to explain it is solved by Lipton (2004) with a reference to the *interest-relativity* of explanation and the *contrastive questions* that express its existence. Our particular interests help us to select from among causes because they mobilize particular questions which highlight a particular structure in the phenomenon. Whether one of the causes of the phenomenon will actually explain it will depend on which aspect of the phenomenon we wish to explain.

> As Carl Hempel has observed […], we do not explain events, only aspects of events (1965: 421–3). We do not explain the eclipse tout court, but only why it lasted as long as it did, or why it was partial, or why it was not visible from a certain place. Which aspect we ask about depends on our interests, and reduces the number of causal factors we need consider for any particular phenomenon, since there will be many causes of the eclipse that are not, for example, causes of its duration (p. 33).

The kind of intrigued question that can highlight a particular structure or aspect has a contrastive character, and it makes us see the phenomenon under a contrastive light. The interest-relativity of explanation determines that we are never merely asking "why this?". Rather, our questions would always have, explicitly or implicitly, a longer form: "why this *rather than that*?". A phenomenon is always composed of a fact and a foil, and the same fact may have different foils; moreover, the causes that explain a fact relative to one foil will perhaps not explain it relative to another. The following of Lipton's examples are precious in this context:

> We may not explain why the leaves turn yellow in November simpliciter, but only for example why they turn yellow in November rather than in January, or why they turn yellow in November rather than turn blue.
> […] When I asked my, then, 3-year old son why he threw his food on the floor, he told me that he was full. This may explain why he threw it on the floor rather than eating it, but I wanted to know why he threw it rather than leaving it on his plate (p. 33).

Citing Garfinkel (1981), Lipton (2004) shows that explaining from contrasts is strikingly similar to applying Mill's Method of Difference. When we wish to know why the leaves turn yellow rather than blue, we are looking for what yellow leaves have that blue leaves have not. It is as if we were comparing two groups, one containing instances of yellow-leaves(-in-November-in-place-x), and another containing instances of blue-leaves(-in-November-in-place-x), and were looking for the difference between them – which, according to the Method of Mill (and to our interests), is one important cause of the yellowness of certain leaves. We would discover another cause of their yellowness were we to pick a different foil, or control-group: leaves are yellow rather than blue in November because of a pigment, while leaves are yellow in November rather than in January because of a temperature. Given the resemblance, Lipton (2004) proposes that to explain from contrasts is to select causes by means of what he called "the Difference Condition": "*To explain why P rather than Q, we must cite a causal difference between P and not-Q, consisting of a cause of P and the absence of a corresponding event in the case of not-Q*" (p. 42).

But what if there were no blue leaves in the world? Would we then still be able to answer why certain leaves are yellow rather than blue? How would we apply the Difference Condition in cases where the contrastive question does not supply two distinct instances?

> For example, we may explain why a particle was deflected upward rather than moving in a straight line by observing that the particle passed through a particular field: this field is a causal

> presence that explains the contrast, but it is not clear in such a case what the corresponding absence might be. (I owe this point and example to Jonathan Vogel.) This sort of case is not unusual, since *we often ask why a change occurred rather than the status quo* [emphasis added] […] in the particle case, we seem […] to have only one instance, and that seems at first to block the application of the notion of a corresponding event.

This may remind us of psychoanalytic particular hypotheses. If we think of a patient as an instance, particular hypotheses are also about single instances with no corresponding events. A particle deflected upwards rather than moving in a straight line is like a patient not drinking water from a glass rather drinking from it. And we have seen that indeed what guides analysts in the inferences of particular hypotheses is more or less the question of "why a change occurred rather than the status quo" – the status quo related to the patients themselves or to the culture they are in. In the end, Lipton (2004) declares that the Difference Condition does apply in cases of single instance, after all, including the case of the wayward particle, "though this is […] difficult to see at first because what the Condition requires in that […] case is the absence of an absence" (p. 43):

> we have the presence of the field, where in this case the corresponding event would be the absence of any field, and this (absence) is absent, there being a field present. Although the application of the Difference Condition is easiest to see in cases like those where the Method of Difference applies, with two quite distinct instances in one of which the effect occurs and in the other of which it does not, the Condition applies to single instance contrasts as well (p. 44).

If we associate the recognition that satisfactory explanations are causal and come from contrastive questions to a model of Inference to the Loveliest Potential Explanation, we are walking toward a faithful description of our more or less successful inferential practices (Lipton, 2004). But here a contrastive question would also be welcome: why prefer this account rather than classical ones, like the original by John Stuart Mill, or the Bayesian account? Why would Lipton's version of IBE be the best description of the three?

We have just seen that the Method of Difference is compatible with IBE. On the other hand, IBE addresses some points of our inferential practices about which Mill was practically silent. For starters, every philosopher agrees that his four "experimental methods" taken together (Difference, Agreement, Residues and Concomitant Variation) do not fully account for our inferential practices; even Mill himself agrees with this, for he complements such methods with the Method of Hypothesis, "which looks a great deal like a method of explanatory inference" (Lipton, 2004, p. 126). That said, the Method of Difference should still be charged with two serious shortcomings that IBE cannot be charged with: the Method alone does not show us how to infer unobservable or unobserved differences (inferred differences) or how to select from among multiple differences (Lipton, 2004).

It tells us we are entitled to attribute a causal role to a difference, but is silent about how the latter is discovered; thus, it does not account for cases where the difference itself must be inferred. The Method does not advise us to look for any *relation* between the antecedents and the evidence of the two cases under comparison – while, in IBE, as inference is determined by the potential *explanatory connections* between a difference and the contrastive evidence, we could, in principle, conclude that a difference we are not observing is the cause we are seeking (Lipton, 2004). This shortcoming has a grave consequence:

> if the Method of Difference does not account for inferred differences, it is unclear how it can account for any contrastive inferences at all. The trouble lies in the requirement that we know that there is only one difference. Even if only one difference is observed, the method also requires that we judge that there are no unobserved differences, but it gives no account of the way this judgment is made (pp. 127-128).

The second shortcoming of the Method of Difference is that it does not describe the manner we would reasonably select from among the multiple differences emerging between the fact and the foil. According to Mill's strict statement, an inference is only warranted when we know there is a sole difference between the histories of fact and foil – but even Mill would recognize this is a rare, if not impossible, condition (Lipton, 2004). Although we could rely on background knowledge and on ever more controlled experiments to select the crucial difference, the Method "does not itself tell us where this knowledge comes from, and even a careful experiment will seldom if ever leave us with only one possible cause, once we allow for the possibility of unobserved and indeed unobservable causes" (Lipton, 2004, p. 128).

Philosophers sympathetic to Bayesianism such as Bas van Fraasen and Wesley Salmon see it as a threat to the worth of IBE; their argument is simply that, if Bayesianism is right, then Explanationism must be wrong. Although Lipton (2004) agrees with their premise that Bayesianism is right, he sees their argument as a *non sequitur*, for the reason that the two models/approaches are "not only [...] compatible, but [...] [also] complementary" (p.). According to our philosopher, explanatory considerations are ancillary to the actual procedure through which the Bayesian formula is realized. Even if this formula grasps how the likelihood of a hypothesis changes when evidence emerges, the process that brings about this change would be explanationist in essence. So,

> arguing that Inference to the Best Explanation is wrong because Bayesianism is right is like arguing that thinking about technique cannot help my squash game because the motion of the ball is governed by the laws of mechanics. Even if Bayesianism gave the mechanics of belief revision, Inference to the Best Explanation might yet illuminate its psychology (pp. 107-108).

Leading to the same result as that formula would, but with a touch of lifelikeness, Explanationism would be more complete than bare Bayesianism. Bayesianism can be aided by it: Explanationism can account for the *generation of hypotheses* and for the determination of *what counts as relevant evidence* where Bayesianism cannot, as well as help in the harsh matter of *attributing some value to the components* of the famous formulaic theorem (Lipton, 2004). Lipton (2004) reminds us we are not very good at directly calculating probabilities, so without a heuristic of explanatory considerations we would barely be able to follow the process of conditionalization the theorem comprehends.

We would not be able to estimate the likelihood of an evidence from a hypothetical state-of-affairs – a value for Pr(E|H&B) – if we did not cogitate *how well the state-of-affairs would explain the evidence*. Actually, as in real life H usually does not entail E, it is not clear how else we could estimate how probable E becomes with H. If Lipton (2004), the historians and the psychoanalysts are right in that loveliness is reasonably well-correlated with likelihood, and in that the former is easier to assess than the latter, then Bayesianism does well to rely on mediating explanations:

> when we consider the loveliness of a potential causal explanation, we may consider how the mechanism linking cause and effect might run, and in so doing we are helped in forming a judgment of how likely the cause would make the effect and how unlikely the effect would be without the cause (p. 114).

Analogously to the Bayesian model, in the IBE model the inferences are influenced by the prior likelihoods of each competing explanation. Unsurprisingly, in the IBE model such priors would be guessed with the help of explanatory considerations. But Bayesians themselves would also be "happy to acknowledge" (Lipton, 2004, p. 115) that explanatory virtues such as "unification, simplicity and their ilk" (p. 115) may play a role in fixing a value for Pr(H|B).

Finally, Lipton (2004) argues that explanatory considerations may also play a role in the determination of the evidence's prior – Pr(E|B). The evidence prior to observation is in many respects just another prior hypothesis and, as such, its likelihood is "influenced by prior conditionalization and by considerations of simplicity *et al*" (p. 115). "Pr(E|B)" can be translated to "how surprising it would be to observe E" (p. 116), which is judged from how well our prior beliefs would explain the existence of E. For example, had an analyst the hypothesis that his patient's repetitive comment on the weather is the manifestation of a crucial motive, she would have to estimate the "standard" likelihood that a person with approximately the same features, in the same context and culture, would comment on the weather (Pr(E|B)); if she comes to estimate that it is a high likelihood, namely, that this is not a surprising event, it will be because she thinks that many "standard" factors, not a special motive, *explain* the repetitive behaviour well.

Explanationism would allow the application of Bayes' formula not only in establishing the values of its items, but also in establishing its more radical stuff: it tells, in the first place, how to generate a hypothesis to conditionalize in the formula and which evidence is proper to conditionalize it. The theorem "says nothing about where H comes from" (Lipton, 2004, p. 116), while the IBE model gives a good, perhaps the best, explanation for our successful habits of hypothesis construction, for our "contexts of discovery". Without explanationist resources, Bayes' theorem would also have a hard time to judge which evidence, in the chaos of existence, is relevant to judge a specific hypothesis.

In sum: "generally speaking the explanationist heuristic does a good job of enabling us effectively to perform the Bayesian calculation, or at least to end up in pretty much the same cognitive place" (p. 119). Lipton's measure before Van Fraasen and Salmon's instigation is, to use his word, irenic: "the Bayesian and the explanationist should be friends" (p. 120).

My measure before Grünbaum's instigation is also irenic: the Freudians and the Millians should be friends, because the Freudians are arguably explanationists, and the latter are very close to the Millian folk. We could also follow a longer line: Freudians are friends of Bayesian historiographers, Bayesians can be (and are better served in being) explanationists, and the Millian Method of Difference is essential in the process of inferring to the best explanation. Such labels are all friends having fun in an energetic but restricted party. The separations between Freudians and Millians such as Grünbaum are not that numerous – if "legend has it that there are 'six degrees of separation' between us all" (Lynch, 2014, p. 792), here we have one degree, two at most.

5.6 CONCLUSION: A RESPONSE TO GRÜNBAUM

Authors such as Hopkins (1988), Michael (2008, 2015, 2019) and Lacewing (2012a) have maintained that psychoanalytic inferences reflect the IBE model of inference. Throughout this chapter, it has become clear that I concur with them. But Hopkins has also maintained that, *because psychoanalytic inferences are Explanationist, they cannot be Millian*; and Michael, in his turn, has maintained that psychoanalytic inferences reflect a *non-Millian Explanationism*. Michael thus implies, in distinction with Hopkins, that a Millian version of Explanationism exists, but both of them clearly think that Grünbaum's strategy of checking whether Freud applies the Method of Difference to his "science" is misplaced. *I disagree with them on this point – but I also disagree with Grünbaum in that analysts do not apply the Method*. Lipton's book compels us to accept that IBE depends on the application of the Method of Difference, even when inferences around single cases are at stake. More precisely, he convinces us that IBE depends on the Method and goes beyond it – but this is still proof that the former model is consistent with the latter. In this chapter, I aimed to show that psychoanalytic inferences contain this synergy. I have argued, for example, that Explanationist inferences of particular, categoric and core psychoanalytic hypotheses depend on the determination of control-instances and that rather-than questions guide the generation of a pool of particular hypotheses on the mental schema at play. I believe that many analysts are Millian but are not aware of it – and that if they became more explicit and rigorous in their application of the Method of Difference, the spirit of their practice would not only remain the same but also gain stronger foundations.

With Wallace, we have seen that clinical-psychoanalytic objects and inferences are much like historiographic objects and inferences. With Cleland, Sober and Tucker, we have seen that historiographic inferences are abductive (that is, inferences to the best explanation); Sober and Tucker, in particular, argue that historiographic inferences are also Bayesian, and we have just seen how the Bayesian and the Explanationist models work in synergy as much as the Millian and the Explanationist. That synergy is explicit in our model: repetitiveness and specificity are at bottom statistical notions; they are properties that confer values of likelihood for psychoanalytic evidence, thus also for psychoanalytic hypotheses. Well, if clinical-psychoanalytic inferences are considered weak only because they are *ex post facto* and sometimes based on a single "object" (in the case of the inference of particular hypotheses), then the most respected historiographic inferences should also be considered weak. Are normativist epistemologists willing to discredit the hypotheses about the origin of the universe, the extinction of the dinosaurs, and the conditions leading to the Great Navigations, for example, just because they are not subject to experimental control? If such hypotheses are recognized as scientific, psychoanalytic hypotheses should be recognized as scientific as well.

The reader may raise an objection to the last paragraph: that there is a fallacy here, namely, an Appeal to Consequences (Van Vleet, 2021). The argument above would imply something like: "if Grünbaum is right about clinical psychoanalysis, then we could not trust historians, not even the natural ones, and this would be an epistemic, social and spiritual *disaster*; therefore, Grünbaum must be wrong". Well, we know that the undesirable consequences of a claim do not imply its falseness. Another reader may rather raise the objection that the argument above is an example of a *Tu Quoque* fallacy: an appeal to a person's hypocrisy or incoherence to conclude that the person's claims are false (Van Vleet, 2021). The flawed argument could in this case be paraphrased thus: "Grünbaum, who probably considers the Big Bang hypothesis to be scientifically supported, is applying double standards when charging the inferences on the Rat Man of being pseudoscientific; therefore, the inferences on the Rat Man must be scientific".

The response I am giving to Grünbaum would indeed consist just in Appeals to Consequences or to Hypocrisy had I not provided some justificatory material around Explanationist and Bayesian inferences in historiography, and before that around what psychoanalytic inferences carry of them. Clearly, I have been much more descriptive than justificatory in this chapter. A complete justificatory account would depend on a careful formalism and on an involvement with the main objections raised against IBE and Bayesianism[64]; I have left this long and complex account for another moment lest the present chapter ended up vaster than it already is. Discussions about the reasonableness of the models at stake, though, are found in several of its corners.

As to the justification of the Method of Difference, I have taken it as a given. I have found no compelling argument to disagree with Edelson (1986) when he states that "the canons of eliminative inductivism are the canons of scientific reasoning and method" (p. 232). I have not taken Grünbaum's model of scientific reasoning and method to be anachronic, as many analysts and philosophers have. On the other hand, I have taken his versions of clinical-psychoanalytic inferences to be unfaithful to the actual (and the potential) ones. My general response to Grünbaum is that such inferences are actually and potentially consistent with his precepts. And, independently of his critique, that clinical psychoanalysis displays a very reasonable method, as reasonable as historiographic methods; a method that can evolve just like historiographic methods have evolved.

Now, a word on my responses to some grünbaumian and pre-grünbaumian charges. The risks of confirmation bias and question-begging inferences in clinical psychoanalysis were by and large countered by the description/prescription given of the interaction between its core, particular and categoric hypotheses. The *post hoc ergo propter hoc* and the thematic affinity fallacies became unthreatening with the revelation that the thematic affinities usually elected by analysts for the inference of particular hypotheses are repetitive and specific ones and with the possibility of devising control-groups for the inference of categoric hypotheses. His general charge that the foundations of psychoanalytic theory – its core hypotheses – cannot be tested at all inside the clinical context was also questioned. The sampling bias issue was not tackled in

[64]We might have to go, for example, through Lipton's (2004) attempts to overcome each of the great objections raised against the reasonableness of IBE – the "catch-22", the "Hungerford's objection", the "Voltaire's objection", the Van Fraassen's objection of "Underconsideration" (a.k.a. the "bad lot objection") – and through the dispute between Frequentists and Bayesians.

this chapter, because the treatment to it would have been just the following: if we assume that mental illnesses are "natural experiments" (see p. 80), that is, that the difference between neurotics and "non-neurotics" is not radical, but comes down to the degree to which certain mental features are manifested, and that the evidence collected for clinical research should not come only from private practice, but also from various kinds of institutions, the sampling bias issue becomes a hollow one.

Finally, I should point out that what this chapter has demonstrated is that the inferential method in clinical psychoanalysis is *reasonable* or valid, *not* that it is *consistent* or reliable. Through the elements I have presented, I am not able to sustain that every analyst, of every background, would translate this method to action in the exact same manner. Other resources and approaches would be needed to characterize the method at such a level of detail that everyone trained in its application would, faced with the same evidence, arrive approximately at the same conclusions. In other words, the arguments above may be a prelude to resolve the Consensus Problem (Seitz, 1966), but they do not resolve it. At this point, I am somehow reproducing Michael's (2015) corollary that psychoanalysis is not a pseudoscience nor a full-blown science because, although it "uses a legitimate form of reasoning" (p. 143), it still "needs to earn the credibility that comes with being a science by showing itself to be epistemically reliable" (p. 144). It follows that the arguments of this chapter do not fully satisfy Schurz's (2014) intersubjectivity assumption of scientific practice (see p. 87). However, "psychoanalysis has the potential to be a reliable epistemic enterprise" and "steps in this direction are being taken" within the field (Michael, 2015, p. 150).

To be honest, philosophical work alone could never deliver that level of detail; only empirical means could prove a significant consistency in the wide application of a certain meticulous version of the clinical-psychoanalytic method. We have seen in Chapter 2 the initiatives of Luborsky and Bucci in this direction and, recently, Kaszubowski (2016) proposed "a formal model for free association as a mean to build rigorous relations between clinical evidence and psychoanalytic theory" based "on probabilistic topic models, such as Latent Dirichlet Allocation"; he even evaluated the model "with a clinical case study based on the audio recording and full transcription of 47 psychoanalytic sessions" (p. ii).

I believe that, for the Consensus Problem to meet its end, clinical expertise and empirical studies should keep informing epistemological reflections, and vice-versa.

Let us now face Grünbaum's famous contamination charge.

6 THE PROBLEM OF DATA CONTAMINATION BY SUGGESTION

6.1 INTRODUCTION: THE IDEA OF SUGGESTION IS INTANGIBLE IN GRÜNBAUM

In Grünbaum's critique, we barely see him giving substance to the competing explanations that would render psychoanalytic explanations suspicious or improbable. There can be diverse non-Freudian explanations for symptoms, healings, dreams, parapraxes, jokes, memories, fantasies, etc.; human behaviour can have chemical causes, recent mental causes and conscious mental causes, for example. It can, alas, be a product of suggestion. But what is suggestion, exactly? And could a deeper understanding of this term do something for the epistemology of clinical-psychoanalytic research?

Grünbaum examines the concept of placebo in many works, including some independent of his critique (Grünbaum, 1981, 1986b); by contrast, he takes the concept of suggestion for granted. While he does a thorough reframing of the placebo concept, showing why the influential definition of Shapiro and Morris (1978) was unhandy, it is only in offhand manner, in the middle of an argument, that he adumbrates what he means by suggestion – as if the notion were not as confusing as the notion of placebo. Considering that most of his arguments against psychoanalysis depend on that notion, this fact is worthy of analysis.

It is also worthy of analysis that he did not mention any theory about the causes and the nature of suggestion. Did he take for granted the *psychoanalytic* theory of suggestion so that he could internally refute the psychoanalytic solutions to the problem of suggestion? Or, if he had in mind other theories, why did he treat it only as an epistemological hindrance, as only an alternative explanation for a range of phenomena? Why did he bypass its psychosocial explanation?

There is yet the possibility that he thought of his discussion on the concept of placebo as sufficient for his use of the concept of suggestion. After all, both that discussion and this concept are used as premises in the falsification of the NCT. In this case, the two concepts would be related or even mingled in his argument. He writes, for example, about "the therapist's suggestion or some other placebo factor" (Grünbaum 2015, p. 17), indicating that suggestion is a kind of placebo-factor.

A placebo is, according to the philosopher, a therapy in which the healing is caused by one or more of its incidental factors and is not caused by *any* one of its characteristic factors (Grünbaum, 1981). The theory of the therapy is what defines which of the therapy's components are the hypothetical causes of its aimed healing – the characteristic factors of the therapy – and which should not be (Grünbaum, 1981).

But our common-sense tells us that, in a placebo therapy, the incidental factors that cause the healing are psychosocial; and, in its positivistic eagerness, the definition defended by Grünbaum does not make that clear and mandatory. Let us suppose a rare, but possible, case in pharmacology: the excipient or the adjuvant of the medication being tested is the only cause of the healing[65]. By Grünbaum's definition, this medication would be a placebo, for the theory of the therapy did not consider as characteristic the substances used as excipients or adjuvants. But our common-sense is hurt in conceding that this medication would be a placebo; we would say simply that the cause of the healing was one substance rather than the other. Hence, Grünbaum's definition should change to: a placebo is a therapy in which the healing is totally caused by one or more of its *psychosocial* incidental factors.

Placebo phenomena may, only then, come to be related to suggestion, as shall become clear with our approaching discussion. More accurately, it can be taken as a kind of suggestion – not the contrary, as in Grünbaum's excerpt above –, a kind of suggestion that happens in a context of suffering and therapeutic aims. Now, in this case, only as far as Grünbaum said something about the causes and the nature of placebos, we can say that he did the same with suggestion. In his 1980 paper on psychoanalysis, while discussing Jerome Frank's work, he writes that the incidental factors of the placebo hypothesis could be "the arousal of hope and the mobilization of a sense of mastery in the patient" (Grünbaum, 1980, p. 342). Further, in 1981, he writes about "interesting conjectures" as to the identity of the incidental factors: a "psychogenic activation of the secretion of substances" such as pain-killing endorphins, interferon and steroids (p. 160). In his last publication on psychoanalysis, he talks again about "the mobilisation of the patient's hope by the therapist" (Grünbaum, 2015, p. 20).

This incorporation of concepts, however, is not supported by some factorial analyses:

> […] there seems to be a common assumption that placebo effects are due to suggestion. From this perspective, it may come as a surprise that empirical research has failed to show any clear relation between suggestibility and placebo effects (Evans, 1989). This lack of correlation, however, becomes a little less surprising in view of the relative lack of correlation even between different kinds of suggestibility (Eysenck & Furneaux, 1945; Gudjonsson, 1989). In fact, there seems to be no empirical support for a unitary trait of suggestibility. Similarly, research on placebo effects has failed to find any stable trait of placebo reactivity. This raises the question whether there are any stable personality traits involved here, or whether both suggestibility and placebo effects are very much the products of temporary situational factors and temporary mental states (Lundh, 1998, p. 32).

[65]Unless the excipient and the adjuvant were also tested as independent variables, the experimenter would never uncover this causation.

As the author says above, studies indicate that "suggestibility" is one name for a lot of different traits; and, as long as this concept and its empirical translation remain open to discussion, we can consider inconclusive the empirical refutations of the placebo-suggestion affiliation. To be sure, in an experimental study some years after the citation above,

> individual differences in suggestibility were found to significantly contribute to the magnitude of placebo analgesia. The effect was graded with respect to suggestibility and verbal expectancy levels. Significant pain intensity reductions were observed in highly suggestible subjects, and to a lesser extent in mid suggestible ones, who received suggestions presumed to elicit a high expectancy of drug efficacy. These results indicate that both suggestibility and verbally induced expectancy of drug efficacy can have an additive effect on placebo analgesia (De Pascalis, Chiaradia & Carotenuto, 2002).

This is consistent with the proposition that placebo phenomena are part of the concept of suggestion. If we wish to advance on this point, though, we must start to wonder what this last name could possibly mean.

This chapter deals with the threat of data contamination by suggestion in clinical psychoanalysis. I shall first delve into conceptual and theoretical facets of suggestion. From this, I shall show that cases of cognitive suggestion do not pervert the primary evidence in clinical psychoanalysis – that they are, actually, primary evidence itself. I thereby conclude that those facets, a blind spot in Grünbaum's critique, deflate that hypothesis of contamination. I also explain why, precisely because suggestion phenomena are primary evidence, analysts should regulate the cognitive suggestion they have the power to foster. Finally, I propose the rigorous implementation of a "cognitive baseline" as a means to regulate and analyse cognitive suggestion. This is again a weakening of Grünbaum's claim that psychological hypotheses cannot be cogently tested in a clinical context.

6.2 WHAT IS SUGGESTION?

6.2.1 The Concept of Suggestion

6.2.1.1 History

Suggestion-related phenomena have been documented since the magical, pre-scientific, times of human history. To agree with such claim, one has to sustain a distinction between these phenomena, the interpretations of these phenomena, and the use of the word "suggestion" to name, in the 19th century, one of these interpretations, the theoretical charge of which grew ever more along the emergence of dynamic psychiatry (Ellenberger, 1970). Let us keep that distinction in mind and embrace both the anachronism of finding the concept of suggestion where it was not yet born and the advantage of having a broader and more substantial idea of this concept. Therefrom, it is fair to say that, in the pre-modern methods of healing and spiritual improvement, from the Palaeolithic to the 18th century, from shamans to magnetists, many concepts were invoked to encompass the same thing we, offspring of the Enlightenment, place under the concepts of suggestion and placebo.

But we still struggle to understand this thing. In fact, science has not done enough for suggestion and placebo: now and again, these concepts show us hazy and inscrutable faces. It is no surprise, then, that the same evasive moves that – we have argued in the previous section – can be found in Grünbaum are also found in many scholarly contexts.

The path of suggestion in human history can be tracked in the monumental book on the origins of dynamic psychiatry, "The Discovery of the Unconscious", by Henri F. Ellenberger. This path starts in the dawn of human life with the idea of magic, of fundamental, powerful and mostly invisible forces that would rule the cosmos and that could be mastered or at least manipulated. This idea seems familiar to us and it is indeed; according to Ellenberger (1970), magic is nothing but a "fallacious anticipation of science" (p. 35), and magical performances, nothing but inadequate endeavours to exert power over nature; but "whereas science is 'neutral' and can be applied toward good and evil ends, magic is usually more strongly divided into 'bad' or 'good'" (p. 35), the former creating disease and the latter eradicating it. The enchanted *Lebenswelt* is universal among pre-modern cultures, and persists, timidly but forcefully, in our "enlightened" era.

The French author contends that a systematic study of magical medicine would for sure elucidate the phenomena we today call suggestion and autosuggestion. For certain civilizations in certain times, the magicians' power is more than formidable; they would produce and cure a number of symptoms and diseases and bring a person back from the verge of death – or push the person toward it. This is a creed, but also, in a way, an objective truth.

Ellenberger (1970) refers to a report made by the anthropologist and doctor Herbert Basedow, published in 1925, of how Aboriginal men of Central Australia reacted when realizing they were the target of a murderous spell. In the cultures of that territory, when an man's enemy points him a stick or a bone while speaking the right words in the right manner, such man would become aghast and pale, his hands trying to expel what was presumably cast into his body; then his face would become distorted, as if he were stricken with palsy, and his body would start to tremble and twitch, making him fall to the ground; he would demonstrate great agony, would moan and refuse to eat. If a healer, a *Nangarri*, does not come to a man in this state and work to bring a counter-spell, the man dies within a few hours. The *Nangarri* performs rituals and with a mouth-sucking method draws the evil out the man's body; the cursed man sees what was extracted from his body and, delighted, starts to recover.

With another report, Ellenberger (1970) demonstrates once again how extreme the reach of a suggestion response can be – even if there is no enchanter present. When stating that "a severe disease or even an acute psychogenic death can result from the infringement of a taboo", he is not merely describing a disease theory from an exotic culture; instead, he is stating "an actual fact, confirmed by many reliable eyewitnesses" (Ellenberger, 1970, p. 22). A missionary in the French Congo, Reverend Grébert, reported in 1928 the case of a pupil who was struck by convulsions, perishing from it a while later; the pupil had just eaten bananas cooked, as he had learned, in a pot previously used for yuca and his parents had informed him of a taboo dictating that he would die if he ever ate that root. The convulsions were supposedly caused by the thought that he violated an ancestral command and should be punished for such, and the emotional shock that developed from that thought ended up reproducing the fantasized penalty.

The author still compares this case to the common response of individuals from Polynesian cultures after they have violated a taboo, highlighting the suggestive character of this response:

> In Polynesia there have been frequent reports of psychogenic deaths resulting from the violation of a taboo, although the features differ from those common in Africa. Death occurs in a less dramatic way, more slowly and quietly; the patient lies down, refusing nourishment, and dies within a few days. What is important here is not so much the violation of a taboo, but rather that the violation has been exposed and made public and the offender has thus been made an object of shame (Ellenberger, 1970, p. 23).

Suggestion would probably be "the most important agent at work in the practice of magic", specifically of magical medicine (Ellenberger, 1970, p. 35). It was only cultivated for so long because it significantly works, and it only works because the sorcerer, the sufferer and the whole community of which they are part believe that magic art is a powerful tool to deal with the hostile events of life. The beliefs of everyone involved in the magic procedure are crucial for its efficacy, and the ensuing sense of control over nature helps to create and maintain a strong communal bond (Ellenberger, 1970).

In pre-modern methods of healing, the diseased are more confident and hopeful about the *person* of the healers than about the substances and procedures handled by them. The main agent of the cure would therefore be an image of personal vigour, with sheer skill and literacy having only a supporting role in the production of this image. Ellenberger (1970) cites Alphonse

Maeder, according to whom the religious healer is but a vessel for the "archetype of the Saviour", from which the self-healing tendencies of the diseased are awaken and developed.

It is from this primordial medicine that dynamic psychiatry derives, and one can argue for a continuous movement between magic and exorcism, exorcism and magnetism, magnetism and hypnotism, and, finally, between hypnotism and the modern dynamic schools (Ellenberger, 1970). Before we go to the century where exorcism and magnetism clashed with each other, the 18th, let us note that a precursor of the psychological interpretation of magnetic phenomena – which in its turn led to the scientific (or non-animistic) interpretation that was condensed in the concept of suggestion during the 19th century – emerged in the 16th century with the old concept of "imagination".

Philosophers and physicians of the Renaissance "became very interested in a power of the mind, *Imaginatio*, a term that held a much broader meaning than it does today and included what we call suggestion and autosuggestion" (Ellenberger, 1970, pp. 111-112). Ellenberger cites Montaigne, for whom imagination, as can be read in his "Essays", was responsible for what was commonly ascribed to magic; it could be responsible for the contagious character of human emotions, for every type of disease, for someone's death and even for the transformation of one sex into another. The interest in imagination prolonged to the 18th century, with the "widely read and quoted" book by Lodovico Antonio Muratori, "On the power of human imagination[66]", published in 1745 (Ellenberger, 1970). Muratori considered dreams, visions, delusions, fixed ideas, antipathy (a name for phobias), and somnambulism as instances of imagination.

However, some obscure explanations for these peculiar phenomena still grew on the soil of the 18th century – there was still Father Gassner, a famous remainder of the Dark Ages, mobilizing and expelling evil spirits (such as melancholia) with prayers and faith, and Doctor Franz Mesmer, the father of animal magnetism, whose explanation for the great effect his healing method had over people relied on a mysterious substance that would flow between his body and the body of his patients. These names, Gassner and Mesmer, rhymed not only in regard to their phonemes, but also in regard to their deeds and their obscure explanations for these deeds; yet Ellenberger (1970), considering Mesmer a relatively enlightened man, places the birth of dynamic psychiatry in the dispute that took place between them.

[66] "*Fantasia*" was the Italian word translated, by Ellenberger, to "imagination"; it could also be translated to "fantasy".

But the historian also admits that the first to entertain psychological hypotheses (in their modern sense) for all that demoniac and juicy phenomena was the Marquis de Puységur, one of Mesmer's apprentices. Puységur found out that somnambulism could be induced and interrupted voluntarily (an "artificial somnambulism" that decades later came to be called "hypnosis"), opening a path for great secrets of the human mind; he started to hold that the uses of "magnetism" should only be therapeutic; and, last but not least, he "realized the vanity of Mesmer's teaching of the physical fluid" with the understanding "that the real agent in the cure was the magnetizer's will" (Ellenberger, 1970, p. 72). This last idea was present in the second edition of one of his books, published in 1809 – we may thus walk toward the 19th century.

The renaissancer enthusiasm with the concept of imagination had been replaced for some time by Mesmer's fluidic theory but, in the 19th century, that concept came back to light under a new brand, "suggestion". Suggestion (along with autosuggestion) "came to designate the entire realm previously covered by the notion of imagination" (Ellenberger, 1970, p. 149). It would be no surprise, then, that it was at first a very wide psychological concept, wider perhaps than it is today.

The Scottish philosopher Thomas Brown, for instance, proposed in 1820 a classification of mental phenomena according to which "Simple Suggestion" should be understood as

> a tendency […], by which feelings, that were formerly excited by an external cause, arise afterwards, in regular successions to each other, as it were spontaneously, or at least without the immediate presence of any known external cause (Brown, 1820, p. 199).

That is to say – in line with Brown himself –, the concept of Simple Suggestion would bluntly consist in what we know as Association of Ideas[67]. Simple suggestion would manifest primary laws, whose content Brown exhausted by explicating the law of resemblance, that of contrast and that of nearness in place and time; he stressed that similar orderings of the grounds from which one idea is aroused from another in our minds, one such ordering having been famously presented in the work of his fellow-countryman, David Hume, actually date back to Aristotle.

Following Puységur's turning point, more men introduced new psychological concepts and methods into the study of magnetism. There was for instance Abbé Faria, who claimed that the process of magnetization was due more to the subject than to the magnetizer; he later claimed that some magnetized subjects were more responsive than others. François-Joseph

[67] He argued that the term "association" could be taken as denoting only *idea-based*, *substantial*, mental connections, previously strengthened by a subject's *actual perception* of a concomitance/contiguity; so, he favoured the term "suggestion", which would only imply the *crude fact* that *any* mental event tends to lead to – "to suggest" – another in certain ways, a fact that would not demand a mysterious background (a Comtean stance before the "Course of Positive Philosophy", it seems). (See Brown, 1820, pp. 333-357).

Noizet, who watched Faria's demonstrations and whose teachings were embraced by Liébeault, the founder of the famous Nancy School, can be seen as the bridge between the New Magnetism of Puységur and the golden decade of suggestion and hypnotism that so much influenced Freud, the 1880s. In his historical account of suggestion and hypnotism, Pierre Janet stated that Puységur, Faria, Noizet and many others were not mere precursors, but the actual founders, of the science of those two objects; they have found that the mutual psychological influence between the patient and the magnetizer had a causal burden, and that this influence endured outside the limits of the séance, a phenomenon that was called post-hypnotic suggestion, and very little was added to this until the 20th century. "Posthypnotic suggestion", assents Ellenberger (1970), "had already been described in 1787 and were well known to Faria […]" (p. 76).

In the 1880s, the marvellous phenomenon of posthypnotic suggestion was the centre around which the masters of hypnotism revolved. There were, then, two grand theoretical and methodological schools of hypnotism: the Nancy School, whose "spiritual father" was Liébeault and whose intellectual leader was Hippolyte Bernheim; and the Salpêtrière School, founded by the vigorous personality of Jean-Martin Charcot (Ellenberger, 1970). Each school had its perspective as to the causes of the diseases healed through post-hypnotic suggestions, thereupon also as to the grounds of such healings. Their teachings and disputes had a great impact on the mind of the young Freud – he attended both Nancy and the Salpêtrière to learn from Bernheim and Charcot, and was the first to translate their works from French to German.

When the doctor Liébeault started his hypnotic trials, in the decades preceding the 1880s, any medical man deciding to work with hypnotic methods would have irremediably "compromised his scientific career and lost his medical practice" (Ellenberger, 1970, p. 85). For more than 20 years, his fellows considered him a quack. From his sources, Ellenberger (1970) described those trials thus:

> Liébeault hypnotized the patient by ordering him to look into his eyes, suggesting that he was getting increasingly sleepy. Once the patient was slightly hypnotized, Liébeault assured him that he was relieved of his symptoms. Most of his patients were poor people from the city and peasants from the neighbourhood, whom he indiscriminately treated with the same method no matter what disease they suffered from – arthritis, ulcers, icterus, or pulmonary tuberculosis (p. 86).

Unexpectedly, a respected professor of the University of Nancy became a pupil of the presumed quack and a promoter of the presumed quackery from 1882 on. Bernheim introduced Liébeault's methods in his University's hospital and, in 1886, published a textbook on their "Suggestive Therapeutics" and on the theory that sustained it; for the book, both men became famous.

Liébeault held that both the falling asleep of the hypnotized subjects and the peculiar relationship they developed with the hypnotizers (the rapport) occurred only because the subjects concentrated their attention on an idea (the idea of sleep or other), that is to say, only because suggestion was operative on them; what is more, suggestion was, for him, the one thing that distinguished hypnotic sleep from natural sleep. Bernheim observed that the hypnotizing and the suggestive healings were easier with people long used to obey rules and orders, such as factory workers and old soldiers, and harder with people of higher classes. Interestingly, with the time he realized that waking suggestion was as effective as post-hypnotic suggestion; the Nancy School came to use the term "psychotherapeutics" to refer to the former. Above all, the School defended that suggestibility and hypnotic states were far from rare or pathological matters. "In its wider sense", writes Ellenberger (1970), "the Nancy School was a loose group of psychiatrists who had adopted Bernheim's principles and methods" (p. 88). Boris Sidis and Auguste Forel, authors of classical books on suggestion[68], belonged to such group (Ellenberger, 1970).

The Salpêtrière School, on the other hand, "was strongly organized and headed by a powerful figure" (Ellenberger, 1970, p. 89). Charcot made history by investigating hysteria with the lenses of a 19th century neurologist. He described the whole process of the hysterical crisis with the detail of a taxonomic biologist and demonstrated, in spectacular performances, that one could produce or extinguish symptoms in hysterical patients by hypnotizing them and by dictating to them that their symptoms would appear or disappear; based on such evidence, he attributed the origin of hysteria to post-hypnotic autosuggestions. He assumed, nevertheless, that the natural hypnotic state during which the hysterical autosuggestion occurs is the result of an ill-constituted nervous system. In the Charcotian understanding, thus, hypnosis would have nothing to do with a strictly psychological and perfectly normal order, as it had for the Nancy boys (Ellenberger, 1970; Schmidt, 2017).

In fact, there were overt theoretical quarrels between the schools. Helping to bring the ostracized hypnotism back to scientific concern, Charcot presented his ideas on the subject to the *Académie des Sciences* in 1882; a while later, in 1884, Bernheim disseminated the work of Liébeault to the medical community. This was the spark for "an embittered struggle" between the men (Ellenberger, 1970, p. 87). Bernheim insisted that the hypnotic state was just one among all the states induced by suggestion, and that every human life was suggestible to some degree; in sum, that the Charcotian linkage between hysteria and hypnosis was ill-founded.

[68] We will analyse their definitions of the concept of suggestion in a moment.

Further, he claimed that the symptomatic stages of the hysterical crisis described in the Salpêtrière were mere artifacts, that is to say, that they were fabricated by the regental climate of the hospital – among the thousands of patients he had hypnotized, Bernheim caustically declared, the only one evidencing these stages was a woman who had spent three years in the famous Parisian hospital. This charge seems fair enough: old and young women were there piled, some exhibited for numerous spectators, like prima donnas, in Charcot's clinical lectures; because of his authoritarian conduct, his staff never dared contradict him and complied to his expectations; he used to rehearse the demonstrations and to discuss the patients' cases having them present in the room; etc. (Ellenberger, 1970).

The concept of suggestion was very popular by the end of the 19th century and was conjured to explain facts ranging between miracles, wars, and accomplishments of the sensibility and the intellect. But this produced a side effect. The wide and promiscuous use of the term made it lose its meaning (Ellenberger, 1970). Both pioneers and immediate followers became aware of this and attempted to define the concept, as we shall see in the next section. The lack of consensus remained along the 20th century, when suggestion phenomena started to be investigated under a more behaviouristic approach. In effect, as late in the 20th century as in 1989, Vladimir Gheorghiu asserted that:

> The development of empirical goals in research led to the rapid decline of efforts to conceptualize the phenomenon. This was even true for [Alfred] Binet (1900), who realized that suggestion (suggestibility) is a multi-faceted phenomenon. […] Even in contemporary research on suggestion phenomena, for example in the well-known works of Eysenck (1947), Eysenck and Furneaux (1945), and Stukat (1958), attempts to conceptualize suggestion and suggestibility are non-existent. Both Eysenck and Stukat primarily attempt to explain the factors of so-called primary and secondary suggestibility. Considering the general aspects of suggestion and suggestibility, they maintain, in accordance with Binet's proposition, that the phenomenon in question is not a unitary one. Certainly, this statement is important for the entire research on suggestion, and, contrary to Eysenck's endeavours to classify suggestibility, it has never been seriously questioned. Consequently, we only learn what suggestion and suggestibility do *not* mean.
>
> None of the definitions prevail today. Older interpretations, such as those given by Bernheim, Janet, and McDougall, are quoted in order to call attention to a particular development in research on suggestion and hypnosis, or to point out an especially influential tradition (Gheorghiu, 1989b, p. 99).

The provisional picture given above is one on which obscurantism, transfiguration and controversy haunts the notion of suggestion throughout history, but also one which hints at its enduring meaning and its substance. Having such picture, we shall now struggle to define the concept of suggestion and to categorize its manifestations, so we can at last display the hypotheses relative to the mechanisms that underlie the suggestion phenomena. The intention is to revisit Grünbaum's charge with a fresh outlook.

6.2.1.2 Definition

"Suggestion" is the etymological kin of the nouns "gesture", "gestation", "ingestion", "digestion" and "register", to mention but a few. It derives from the Latin verb "*suggerere*", which is composed by the root "*gerere*" – "to bear", "to carry", "to bring", "to take on oneself", "to take charge of", but also "to do", "to perform", "to accomplish", "to put" – and the prefix "*sub-*" – which means "underneath", but also "lower in rank", "to a lesser degree", "forming a division into smaller or less important parts" and "with less than the normal amount of" (Partridge, 2006, pp. 1266-1268; Danner, 2014, pp. 341, 838). Some questions arise here. What *exactly* is borne or performed? An idea, probably. Underneath *what* this thing is borne or performed? Underneath our consciousness or our critical self, perhaps. Would it be borne or performed with a less than normal vigour, in a primitive, in a split manner? For certain, etymology alone gives us interesting paths to explore…

We may also, as custom dictates, consider the day-to-day usage of the concept. In everyday life, "suggestion" may have three different meanings: (1) that of a recommendation, advice or tip, of a proposition to be considered by someone, as in "May I suggest the *Escargot à la Bourguignonne*?"; (2) that of a hint or insinuation, of something stated or expressed indirectly, as in "You are obviously suggesting he is a boring person"; (3) that of an emergence of ideas or affects in one's mind from some stimulus or seemingly *ex nihilo*, in other words, that of an *evocation*, as in "A work of art must suggest elevated sentiments" or "The idea that it would be nice to be a woman having sex suggested itself to Schreber" (Oxford, n.d.). These meanings seem to be related to each other. The psychological or scientific meaning of our concept, the one we seek, also seems to be related to all the three – especially to the third. But we still need more resources to reveal, from our multifaceted concept, that properly psychological or scientific facet.

Indeed, Freud (1955c) pointed out that in the German language the word was succumbing to a "looser and looser meaning" and would soon "come to designate any sort of influence whatever, *just as it […][did] in English* [emphasis added], where 'to suggest' and 'suggestion' […] [corresponded] to the German *nahelegen* and *Anregung*" (p. 90); thereupon, he praised the scientific efforts of his time to establish a proper usage for the word, referring specially to William Mcdougall (1920), who defined suggestion as "a process of communication resulting in the acceptance with conviction of the communicated proposition independently of the subject's appreciation of any logically adequate grounds for its acceptance" (p. 10).

Actually, most of the scientists who chose to minimally articulate suggestion have attempted to fixate its meaning. Before McDougall's, there were definitions from Freud (1966) himself, in a preface to his German translation of Bernheim (1886), who, in his turn, defined it in 1891 and 1917; likewise, there were definitions in Sidis (1899), Binet (1900), Forel (1906) and Janet (1919). The latter dedicated 38 pages of his book "*Le médications psychologiques*" for such task: according to him, nowhere else a medical lack of accuracy was more obvious than in works on psychotherapy and in the uses of the words hypnotism and suggestion. He confronted himself with the latter's "too general and inexact interpretations" proposed during the two decades preceding his pen – among which he includes his own interpretations:

> […] I have attempted to define it several times. In each of these attempts, I have no doubt been led to conserve the same general conception, but I think, if I am not deluding myself, that I have obtained each time a little more psychological precision (Janet, 1919, p. 191)[69].

However, nobody was able to propose a definition cleaned up of controversy in the 20th century. In the Preface of the proceeding of the 1st International Symposium on Suggestion and Suggestibility, published in 1989, V. Gheorghiu, P. Netter, H. Eysenck and R. Rosenthal claimed that "suggestion" and "suggestibility" were confusing terms, mingled as they were with the ideas of obedience, persuasion, imitation, influence and hypnosis (Gheorghiu et. al, 1989). When it came to the "stepchild of psychology", a semantics "[…] [remained] for the most part unclear" (Gheorghiu, 1989a, p. 5), which seemed at odds with the fact that, "when research on suggestion [had] first [begun] […], the serious endeavour of most authors was to propose definitions" (Gheorghiu, 1989b, p. 99). Gheorghiu (1989b) further asserts that:

> One has to agree with Larède (1980), who notes that the definitions found in dictionaries, including the specialized ones, give the general impression "that […] one is dealing with an extremely complex phenomenon on which many different and often contradictory opinions exist. There seems to be no consensus except unanimity in many definitions accentuating the interruption of the concerned subject's capacity for critical disposition" (p. 26). The Dictionary of Psychology edited by the French psychologist Pièron (1963), is one of the few which expressly points out the vague and undefined character of the term "suggestion" (p. 99).

In recent times, Lacewing (2013b) – in a paper discussing the very same problem we are now discussing – still held that "there is no exact definition" (p. 721), but only a "general understanding", present in most theorists, "that the effects of suggestion are non-rational or irrational, and bypass critical reflection and conscious control" (p. 722). (Yet he thought best to propose a definition, which we shall analyse in a moment). This general understanding would imply that suggestion pervades unsuspected realms of human life, such as advertising and

---

[69] This is my translation from: "*[...] j'ai déjà essayé plusieurs fois de le définir. Dans chacun de ces essais, j'ai été amené sans doute à conserver la même conception générale, mais je crois, si je ne me fais pas illusion, avoir obtenu chaque fois un peu plus de précision psychologique*".

children's education, "and perhaps even forms a part of all communications intended to influence what people think" (p. 722). By citing Schwanenberg (1989), for whom suggestion and persuasion are continuous concepts – the latter resting more on the reflexive and "cortical" side of communication – Lacewing (2013b) seems to accept that the former concept does not deliver us strong, distinctive contours.

An old exchanging of Procrustean accusations between "suggestologists" has made the bed in which we feel the semantic aches of today. Some scientists were charged of overstretching the meaning of "suggestion", others of cutting off vital parts of the concept. Let us have an example of this with Binet's criticism of a classical definition ascribed to the Nancy School and with this criticism's appraisal by Charles Baudouin, a man who designated himself a member of the "New Nancy School". Binet (1900) writes:

> When does suggestion begin? By which feature is it distinguished from other normal phenomena that are not instances of suggestion? This definition is a great problem, and it has long been said that most people who use the word suggestion do not have a clear idea of it. We must obviously recognise as erroneous the opinion of a whole group of scientists for whom suggestion is an idea which undergoes transformation into an act; on this account, suggestion would be confounded with the association of ideas and with all intellectual phenomena, and the term would have a most banal meaning, for the transformation of an idea into an act is a regular psychological fact, which occurs every time the idea attains a sufficient degree of vividness. In the narrow meaning of the word, in its acceptation so to speak technique, suggestion is a moral pressure that one person exerts upon another […] Pressure means violence; as a result of moral pressure the suggested individual acts and thinks otherwise than he would if he were left to himself[70] (Binet, 1900, pp. 9-10).

The Nancy definition presented by Binet, argues Baudouin (1921), would have omitted that it actually assumes the transformation of the idea into an act as specifically "brought about by subconscious activity, […] [as] effected without the subject's being aware of it" (p. 28). According to Baudouin (1921), if this is added to the definition, there is no need to restrict the idea of suggestion to a context of command, of influence between two or more people[71]. That

[70] This is my translation from: "*Quand est-ce que la suggestion commence ? A quel caractère la distingue-t-on des autres phénomènes normaux qui ne sont point de la suggestion ? Cette définition est tout un problème, et on a dit depuis longtemps que la plupart des gens qui emploient le mot de suggestion n'en ont pas une idée claire. Il faut évidemment reconnaître comme erronée l'opinion de tout un groupe de savants pour lesquels la suggestion est une idée qui se transforme en acte ; à ce compte, la suggestion se confondrait avec l'association des idées et tous les phénomènes intellectuels, et le terme aurait une signification des plus banales, car la transformation d'une idée en acte est un fait psychologique régulier, qui se produit toutes les fois que l'idée atteint un degré suffisant de vivacité. Au sens étroit du mot, dans son acception pour ainsi dire technique, la suggestion est une pression morale qu'une personne exerce sur une autre […] Pression veut dire violence ; par suite de la pression morale l'individu suggestionné agit et pense autrement qu'il le ferait s'il était livré à lui-mème*"

[71] Actually, Baudouin (1921) charges Binet's account of also being loose, perhaps even more than the rejected account: "in conformity with his view, he treats as absolutely identical the words 'suggestibility' and 'obedience'" (p. 28).

way, the concept could include the relevant and kindred phenomena of "autosuggestion" without at the same time embracing an enormous number of psychological phenomena. For him, the secret, the bulk, of the phenomenon lies within the person that had the idea transformed, not on the person that ordered it to be transformed – even more so because such external master can be absent in pure autosuggestions. In sum, he says that suggestion *can* be coercive and forcible, oppressive and deceptive – but that this is *not* fundamental to the concept.

The array of definitions could only be mirroring the array of *explanations* devised for the pertinent phenomena since the 19th century. After all, as Baudouin (1921) stated, "questions of words are at the same time questions of things; a definition is a theory" (p. 21). This point is crucial and we shall discuss it soon, but it does not explain the whole of our Babel Tower. In spite of all the theories over a concept, there is always some agreement over *which phenomena the latter is applicable to* (otherwise it would be nonsense to sustain any dispute over its explanations and definitions). The problem, thus, could also reside in how much one is willing to expand the array of phenomena recognized as "suggestion". As was just argued, the reason why the concept disturbs some scientists is that it tends to imply too wide a domain, absorbing almost any psychological phenomenon and becoming a deceptive tool; for others, ascribing it to a narrower domain would do damage to its explanatory potency and its substance.

Our concept, though, is not alone in its misery. Any prudent philosopher is "border-bothered" by very fundamental concepts, such as "knowledge" and "justice". Huemer (2015) stresses that the analyses of such concepts have utterly failed in the 20th century, even though many brilliant, educated and aided philosophers have then seen such analyses as the central task of their academic careers; he claims that none of these philosophers has succeeded with an analysis "immune from the ingenious-counter-example-generators in other philosophers' brains" (p. 51). He is assertive enough to affirm that "no generally accepted analysis of any philosophically interesting term has yet been devised" (p. 52).

Aiming to explain what went wrong in the attempts, he comes across the theory of concepts presupposed by all of them. Attributed to Locke, three important tenets of such theory would be: that concepts can be introspectively examined; that most of them are composed of other concepts; and that definitions govern their application. This theory would ill-explain why in actuality conceptual analysis is so strenuous and why it has been so strongly driven by examples (Huemer, 2015).

According to this Lockean (also called "classical") theory, we should compare the properties present in a concept's definition to the properties of an object in order to decide whether this object is part of that concept's domain. According to Huemer (2015), however, in practice we ascribe a concept to an object *intuitively*, and only afterwards we produce and assess a definition having in mind its degree of accord with that intuitively correct usage. He understands concepts as

> intentional mental states that represent abstract properties. Forming a concept should be understood as a matter of drawing a boundary around a region in the space of natures, grouping together all the natures in that region and distinguishing them from everything outside that region. Just as in the case of physical space, there are infinitely many regions in the space of natures; however, only a limited number of regions will be recognized in any human conceptual scheme. Most regions, that is, will fail to correspond to any actual concept. Different conceptual schemes will draw boundaries in different places. These different ways of drawing boundaries are neither right nor wrong, though some are more useful than others (Huemer, 2015, pp. 57-58).

Our dispositions for drawing these boundaries would be ushered by the criteria of objective similarity[72], usefulness (that is, relevance to our interests), informativeness and wide applicableness; last but not least, they would be profoundly influenced by the speech community: "when one hears the word 'know' […] applied to a given case, one is influenced toward applying the word in other cases similar to that case" (p. 60).

These dispositions would finally explain why conceptual analyses with a Lockean background are nothing but doomed. First, we do not learn the concept of "knowledge", for instance, by reading a definition, but by developing dispositions through interests, observation and imitation; to elude all the counterexamples imagined by philosophers would almost surely mean to reach a definition with "so many complicated, abstract clauses that most readers would find it far more difficult to comprehend the definition than to comprehend the concept of knowledge itself" (Huemer, 2015, p. 64). Second, we should expect that most concepts correspond to regions with intricate and singular shapes, "and there is no obvious reason to expect them to be definable in terms of other concepts" (p. 64); the fact that our conceptual learning is influenced by our linguistic community reduces the chance even further. Third, our classificatory dispositions are not directly introspectable as such, since they are potential, not real, mental states; they must be activated to be accessed.

Based on this sophisticated reasoning, Huemer (2015) indicates methodological conducts for the philosopher who wishes to study a concept without the false hopes of the classical theory. The aim to clarify concepts need not be abandoned, he claims, but only pursued otherwise. He lists six conducts to attain such a clarity:

1) To offer examples in which the concept is appropriately deployed, indicating its breadth.

---

72 "Objective similarity plays an important role; we dislike categories that exclude some objects that are more similar to some items in the category than the items in the category are to each other" (Huemer, 2015, p. 59).

2) To distinguish the category in question from other categories that are related to but distinct from it.
3) To give verbal formulations that approximate the meaning of the target term.
4) To provide with the conditions that are either necessary or sufficient for a concept's applicability even when conditions that are both necessary and sufficient are out of reach.
5) To discuss the role the concept plays in human life, the reason why the category established by that concept is important, and the further implications of something's falling under that concept.
6) To taxonomize its referents, by discussing the logical features of the concept and by examining other general features of the phenomenon to which the concept refers.

We shall have in mind Huemer's point and the conducts he recommends in order to better understand the scientific concept of suggestion. The endeavour includes following a pair of preliminary steps: a list of intuitive instances and counterinstances of phenomena (or categories of phenomena) related to suggestion; a list of definitions of suggestion proposed by prominent "suggestologists" together with a schematic analysis of them. By implicitly and explicitly comparing these two sources, we shall deal with three disputes over the concept: the specifics of the claim that the phenomena under it are unconscious; whether the origin of the phenomena under the concept is mostly inter- or intra-subjective; and whether "suggestion" should refer to the stimulus, to the response or to the whole process between both. Finally, we shall propose, following Huemer, a "verbal formulation that approximate the meaning of the target term".

6.2.1.2.1 Preliminary steps

Below we find a plausible list of instances of the phenomena we may intuitively relate to the concept of suggestion alongside a list of instances of phenomena that, although close to the concept's "region" (in Huemer's sense), are clearly not part of it; we are naming the latter "counterinstances".

Table 4 - Instances and Counterinstances of Suggestion

Instances	Counterinstances
●Learnings that developed suddenly and imper-ceptibly. Example: an insecure man becomes able overnight to dexterously talk in public about a subject that is familiar to him. ●Passionate learning. ●The cognitive and affective effects of fake news/ political propaganda. ●The cognitive and affective effects of pseudo-scientific doctrines (astrology, for example). ●The so-called self-fulfilling prophecies. ●Purposeful but (at some level) unconscious alter-ations in somatic processes – but only in those somatic processes we can directly control (for example, moving limbs, breathing, etc.) or at least directly sense (for example, hearing, heart-beating, stomach-burning, relaxing, etc.). ●Alterations in any other somatic processes that were *somehow* linked, purposefully but (at some level) unconsciously, to mental representations[73]. ●Adherence to unusual and recently established rules. Example: compulsions and inhibitions in general.	●Learnings that depended on will, awareness, practice and habit; learnings of complex skills. Examples: to become able to play the piano, to walk, to drive, etc. ●Critical learning. The entertaining of ideas with-out being affected by strong values and senti-ments. ●Well-informed and logical thinking. ●Any alteration in those somatic processes we cannot directly control, directly sense or link in any way to mental representations (such as the processes of a gland secreting a hormone, of the immune system fighting a virus, of a cell starting a mitosis, etc.). ●The activation of instincts. ●Purposeful and conscious alterations in the kind of somatic processes we can directly control. Example: wanting to move an arm in a certain way and move it in a certain way. ●Adherence to institutional or cultural rules. Example: adherence to the rule of not taking clothes off in public spaces.

[73] For example, I do not directly control my liver or directly sense it (at least not when it is healthy), but I can imagine its place and its contours inside my torso, and I can imagine a blue and loving light cleaning it, or whatever. Maybe *some* physiologic process would happen in the region of my liver if I imagined these things, even though it would not be caused by a "blue light" and it would not do exactly what I expect it to do to my liver; anyhow, a global somatic process related to feelings of pleasure would almost certainly follow my experiment. Such a picture would certainly be related to our concept of suggestion.

●Placebo, open-label placebo and nocebo pheno-mena. The expectations of therapists and the effects such expectations have on patients and also on therapists. ●Successful paradoxical interventions: when the patient gets better with the therapist directing him or her to perform the very problem to be eradicated. ●Religious/magical experiences in general.	

Source: Elaborated by the author (2021).

Now that we have had a grasp of the concept's domain – that we have an idea of which kinds of phenomena should (plausibly) be included in the concept and which should not – we could go through some diligent definitions, part of them classical, of suggestion:

Table 5 - Definitions of Suggestion

Bernheim (1891, 1917)	"[…] the act by which an idea is introduced into the brain and accepted by it" (1891, p. 24). "Suggestibility is the brain's capacity to receive or evoke ideas and its tendency to realize them, to transform them into actions" (1917, p. 18)[74].
Freud (1966)	"[…] a conscious idea, which has been introduced into the brain of the hypnotized person by an external influence and has been accepted by him as though it had arisen spontaneously" (p. 77). "What distinguishes suggestion from other kinds of psychic influence, such as a command or the giving of a piece of information or instruction, is that in the case of a suggestion an idea is aroused [*erweckt*] in another person's brain which is not examined with regard to its origin but is accepted just as though it had arisen spontaneously in that brain" (p. 82). "Indirect suggestions [or autosuggestions], in which a series of intermediate links arising from the subject's own activity are inserted between the external stimulus and the result, […] can […] be described equally as physiological or as psychical phenomena, and the term 'suggestion' has the same meaning as the reciprocal arousing [*Erweckung*] of psychical states according to the laws of association" (p. 83).
Sidis (1899)	"[…] the intrusion into the mind of an idea; met with more or less opposition by the person; accepted uncritically at last; and realized unreflectively, almost automatically" (p. 15).

[74] These are my translations from: "*[…] c'est l'acte par lequel une idée est introduite dans le cerveau et acceptée par lui*" and "*La suggestibilité, c'est l'aptitude du cerveau à recevoir ou évoquer des idées et sa tendance à les réaliser, à les transformer en actes*".

Binet (1900)	"[…] is a moral pressure that one person exerts upon another; the pressure is moral, which means that it is not a purely physical operation, but an influence that acts through ideas […]. Pressure means violence; as a result of moral pressure the suggested individual acts and thinks otherwise than he would if he were left to himself" (p. 10)[75].
Forel (1906)	"[a kind of hypnotizing produced] through the psychical influence of one person on another by means of placing ideas before the latter, which the former induces the latter to accept" (p. 40). "By suggestion (dictation) one means the production of a dynamic change in the nervous system of a person, or of such functions which depend on his nervous system, by another person by means of the calling forth of representations (be they conceived or unconceived) that such a change is taking place, has taken place, or will take place" (p. 54).
Mcdougall (2001, 1920)	"[…] a process of communication resulting in the acceptance with conviction of the communicated proposition in the absence of logically adequate grounds for its acceptance" (2001, p. 74). "[…] a process of communication resulting in the acceptance with conviction of the communicated proposition independently of the subject's appreciation of any logically adequate grounds for its acceptance" (1920, p. 10).
Janet (1919)	"It is a particular reaction to certain perceptions, this reaction consisting in the more or less complete activation of the evoked tendency without this activation being completed by the collaboration of the rest of the personality" (p. 212). "[…] the provocation of an impulse instead of the thoughtful realization" (p. 227) [76].
Baudouin (1921)	"[…] the subconscious realization of an idea" (p. 29).
Coué (1922)	"One could define it as 'the act of imposing an idea on the brain of another person'. Is such an action really possible? Properly speaking, no. Suggestion does not actually exist by itself. It does not exist, and cannot exist except on the distinct condition, *sine qua non*, that it *transforms itself* in the other person's mind *into autosuggestion*: and this word we define as '*implanting an idea in one's self through one's self*' […] [or as] *influence of the imagination on the moral and physical being of man*" (pp. 21, 23).
Dalbiez (1941)	"[An] unconscious and involuntary realization of the content of a representation." (p. 114).
Gheorghiu (1989b)	"[With suggestion,] Whether directly or indirectly, authoritatively or persuasively, implicitly or explicitly, overtly or discretely, deliberately or unintentionally, the subject's behaviour will always be guided in a certain direction. […] […] in the case of the suggestion situation, alternative reactions are also possible which are parallel to the intended reactions. […] Should this alternative of *reacting differently* not be guaranteed because of a forced limitation on a person's freedom of action, i.e.,

[75]This is my translation from: *"[…] est une pression morale qu'une personne exerce sur une autre ; la pression est morale, ceci veut dire que ce n'est pas une opération purement physique, mais une influence qui agit par idées […] Pression veut dire violence ; par suite de la pression morale l'individu suggestionné agit et pense autrement qu'il le ferait s'il était livré à lui-même*".

[76]These are my translations from: "*C'est une réaction particulière à certaines perceptions, cette réaction consiste dans l'activation plus ou moins complète de la tendance évoquée sans que cette activation soit complétée par la collaboration du reste de la personnalité*" and "[…] *la provocation d'une impulsion à la place de la réalisation refléchie*"

	limitations set by fixed behavioural patterns or external constraints, then we are no longer talking about a suggestion situation. […] the effects of suggestion always appear to correspond to a replacement or substitution, […] [there is a] tendency to respond in accordance with the proffered solution *as if* alternative solutions were non-existent in the given suggestion situation" (pp. 100-101, 103). "*Suggestion is a demand situation which, founded on a substitutional or replacement* […] *process, can induce (conscious) uncontrolled responses. In the given situation, though, the subject must theoretically (potentially) also dispose of the alternative of being able to react in a different way*" (p. 105).
Lundh (1998)	"[…] a form of communication, or inter-personal priming, whereby (a) one person (the 'suggestor') intentionally or unintentionally influences another person (the 'suggestant') by means of verbal communication, non-verbal behaviours, and/or other contextual factors; (b) in such a way that the suggestant takes over intentions, feelings, beliefs, or desires from the suggestor, and (c) where this process of influencing relies on the automatic activation of meaning structures in the suggestant." (p. 25).
Lacewing (2013b)	"[…] suggestion comprises communications and features of the structure and setting of communication that, while bypassing the subject's critical and/or conscious reflection, lead to a change in their mental states (beliefs, memories, desires, etc.), mental state reports, and/or behaviour" (p. 721).

Source: Elaborated by the author (2021).

With the aim of exploring the definitions above, we shall classify them and thus be able to compare them by asking ourselves three questions:

1) which point of the suggestion phenomena each definition declares to be the *centre of the concept*;

2) which *kind of subjective factors* (inter- or intra-subjective) each definition considers to be accountable for the bulk of suggestion responses;

3) which *features* each definition ascribes to the concept.

There are three possible *phenomenological centres* for the concept of suggestion. One is the stimulus and atmosphere in which the suggestion phenomena begin – the famous "sleight of hands". Another is the response of the subject at which the phenomena end – the "twist of fate", so to speak. The third possible "centre" is the whole process, all the way through, including the private psychological process between stimulus and response – or "how the sleight was handed and twisted toward a new fate", so to speak. Below I classify the authors according to the point, among these three, that each one of them sticks to the concept; I also present textual fragments to substantiate the classification.

Table 6 - The Phenomenological Centre in Definitions of Suggestion

	Stimulus/Situation	Response/Effect	Process
Bernheim (1891)	"the act by which"		
Freud (1966)	"a conscious idea"		
Sidis (1899)			"the intrusion into the mind of an idea"
Binet (1900)	"a moral pressure that one person exerts"		
Forel (1906)		"the production of a dynamic change"	
Mcdougall (2001, 1920)			"a process of communication"
Janet (1919)		"a particular reaction to certain perceptions"	
Baudouin (1921)			"the subconscious realization of an idea"
Coué (1922)			"[the act] *transforms itself* in the other person's mind"
Dalbiez (1941)			"[a] realization[77] of the con-tent of a representation"
Gheorghiu (1989b)	"*a demand situation*"		
Lundh (1998)			"this process of influencing"
Lacewing (2013b)	"features of the […] setting of communication"		

Source: Elaborated by the author (2021).

[77] As in Baudouin (1921), "realization" here is qualified as unconscious; for this reason, I am considering such term in its sense as a process, not in its sense as a product.

Five authors think “suggestion” refers to stimuli or situational features, only two think it refers to a response and six, that it refers to both and to what happens between both, namely, to a process.

Exploring the definitions further, we may find the two possible answers the authors give to the question about *the kind of subjective factor* accountable for the bulk of suggestion responses. Some consider the factors on the hypnotizer’s or suggestor’s[78] side to be the most determinant of the suggestant’s behaviour; for these, the inter-subjective dynamic is the crucial one. Other authors hold that the suggestant’s psychological structure is to be regarded as what made the greater difference to the response: that the intra-subjective dynamic is what counts the most. Below I classify the authors according to this dichotomy.

Table 7 - Inter or Intra-subjective Significance in Definitions of Suggestion

	Inter-subjective	Intra-subjective
Bernheim (1891)		“an idea is introduced [passive voice] into the brain"
Freud (1966)	“introduced […] by an external influence”	
Sidis (1899)		“the intrusion into the mind of an idea”
Binet (1900)	“one person exerts upon another”	
Forel (1906)	“influence of one person on another”	
Mcdougall (2001, 1920)	“a process of communication”	
Janet (1919)		“a particular reaction to certain perceptions”
Baudouin (1921)		“the subconscious realization of an idea”
Coué (1922)		“condition, *sine qua non*, that it *transforms itself* […] *into autosuggestion*”

[78] From here on, I shall use the terms “suggestor” and “suggestant” as they are used by Lundh (1998).

Dalbiez (1941)		"[a] realization of the content of a representation"
Gheorghiu (1989b)	"*a demand situation*"	
Lundh (1998)	"one person […] influences another person"	
Lacewing (2013b)	"suggestion comprises communications"	

Source: Elaborated by the author (2021).

There is more balance here: seven authors think it is the inter-subjective dynamic what decides the existence and the specifics of suggestion responses, while for the remaining six, it is the intra-subjective dynamic of the suggestant the most accountable for the responses.

Finally: each definition proposes a number of features for suggestion phenomena; I have abducted nine of these features that were more or less common to all of them. Below we can see whether, for each definition, suggestion: involves an influence through ideas; has an unconscious, uncritical or automatic character; ensues a significant change in the suggestant's mental state; is described as the turning of an idea into an action or somatic state (or as a "realization" of an idea); depends on what was already conditioned in the mind of the suggestant, being thus seen as a mere activation of a disposition; belongs to the sphere of communication; involves the intrusion or introduction of an idea; includes the acceptance of this idea by the suggestant; is restricted to the sphere of non-coercitive influences, that is, to influences before which it is possible for the suggestant to react in dissonance with the suggestor's command.

Table 8 - Features in Definitions of Suggestion

	Be	Fr	Si	Bi	Fo	Mc	Ja	Ba	Co	Da	Gh	Lu	La
Influence through ideas													
Unconscious/ uncritical/ automatic/etc. character													
Significant mental change/ Diversion from usual conduct													
Turning of a representation into an act or somatic state (realization)													
Activation of a tendency/ pre-existing structure													
Communica-tion													
Intrusion/ introduction of an idea													
Acceptance of an idea													
Lack of coercion (possible alternative reactions)													

Source: Elaborated by the author (2021).

None of the features is ubiquitous. The element most present is the one of “influence through ideas”, followed by “acceptance of an idea”, and the least is the one of “intrusion/introduction of an idea”.

By implicitly and explicitly comparing these two sources, we shall now deal with three disputes over the concept: the specifics of the claim that the phenomena under it are unconscious; whether the origin of the phenomena under the concept is mostly inter- or intra-subjective; and whether “suggestion” should refer to the stimulus, to the response or to the whole process between them. Finally, we shall propose, following Huemer, a “verbal formulation that approximate the meaning of the target term”.

6.2.1.2.2 The Unconsciousness of Suggestion

Baudouin (1921) minds that, while both suggestion and the will are both active processes “which […] [go] on in the interior of the individual, and whose starting-point is an idea” (p. 326), the former cannot be supported by consciousness all the way, having instead an essentially subconscious mechanism. He puts, finally, that it “does not bear its full fruit except on condition that it be not confounded with the will” (p. 326).

In researching the suggestibility of witnesses, Gudjonsson (1989) conceptualizes a difference between compliance and suggestibility. If consciously wishing to please the interviewer or to avoid the unease of disagreement, a witness is being compliant. He or she could only be called suggestible when privately accepting the idea elicited by the leading question of the interviewer or at least by believing it to be plausible.

Etchegoyen (2005) understands suggestion as one of the instruments for influencing the patient in psychotherapy, the others being support and persuasion, spelling out that “the basis of the suggestive method is to introduce in the mind of the patient, beneath what he is thinking, some type of judgement or affirmation that can operate thereafter from inside […]” (p. 312). Particularly important in Etchegoyen’s classification is the difference between suggestion and mere persuasion; the latter would name an *open and rational* influencing instead of a surreptitious one.

Lundh (1998), a philosopher interested in issues inside psychotherapy, affirms that there are necessarily “unintentional, non-rational processes” (p. 25) in the suggestant. He also defines our concept as a “process of influencing [that] relies on the *automatic* [emphasis added] activation of meaning structures in the suggestant" and "as a form of […] interpersonal priming" (p. 25), whilst "priming"

> refers to the influence that stimulus information presented to a person at one time will have on this person's way of interpreting new information at a later time, without his or her being aware of this. As pointed out by Fiske and Taylor (1991), "it is crucial in priming studies that subjects do not think that the primed interpretation comes to mind because it was previously provided to them […], but instead that subjects think the primed construct comes to mind because of the stimulus itself" (p. 258). In suggestion, also, certain beliefs, feelings, desires or intentions are primed in the suggestant, without his or her being aware that they have been implanted there by the suggestor (p. 27).

Lacewing (2013b) refers to a series of theorists (Freud, 1888/1966; McDougall, 1908; Eysenck, Arnold & Meili, 1975; Gheorghiu, 1989b; Levy & Inderbitzin, 2000) to assume a "general understanding" that the effects in the suggestant "are non-rational or irrational, and bypass critical reflection and conscious control" (p. 722).

Most of the definitions we have considered indicate that the concept of suggestion must refer to something happening somehow out of consciousness. Many who theorize on the matter hold the convention that processes similar to suggestion but without the quality of being unconscious should have another name. We have seen that etymology supports this convention: the prefix of "to suggest" is "*sub*", "underneath", alongside "*gerere*", to carry, bring or put (Partridge, 2006, p. 1268), and we could easily allow the thing under which the *gerere* takes place to be the consciousness. But what should be unconscious in suggestion? All of it, or only its causes, or its effects, or the intention to make it happen? And to whom should it be unconscious, to the suggestor (the person who suggests), to the suggestant (the person suggested to), or to both?

We should notice that the authors cited in this section are talking about the unconscious character of the operation *for the suggestant*. For the suggestor, on the other hand, the operation can be either conscious or unconscious. With the popular figures of the hypnotist and the advertiser in mind, it is easy to admit that the suggestor can act with a conscious intention; nevertheless, "unintentional forms of suggestive influencing may be expected to occur inevitably in all interpersonal functioning" (Lundh, 1998, p. 25). Lundh (1998) comments that in intentional suggestion the interpersonal priming "rests on a certain degree of communicative skill, general psychological knowledge, and empathy in the suggestor", whereas in unintentional, it is a byproduct of a "tuning" between meaning structures (p. 30).

Yet, we can think of two cases of suggestion in which the process seems to be conscious for the suggestant: in voluntary autosuggestion and in an open-label placebo (OLP). An OLP is a placebo-treatment in which the patient knows the treatment is a placebo and nonetheless has placebo-effects. A voluntary autosuggestion is one's realization of an idea from one's voluntary attention to this idea. Both cases indicate that someone can be suggested even when knowing that is being so. Does it mean that suggestion can involve conscious control?

Baudouin (1921) recognizes three kinds of suggestion: the spontaneous, the reflective, or voluntary, and the induced. The last kind would be synonym with heterosuggestion[79], the one occurring between two or more people; the spontaneous and the reflective would be both kinds of autosuggestion[80], that is, of suggestion one elicits on oneself. Baudouin's distinction between the two kinds of autosuggestions is inspired by Ribot's distinction between voluntary attention, the one presupposing reflection and conscious effort, and spontaneous attention, which in its turn is the attention driven by everything facilitating or hindering us in our aims. Reflective autosuggestion would be, thus, the one whose preliminary act of attention has been voluntary instead of spontaneous.

Considering the fact that in spontaneous autosuggestion an idea in able, under certain circumstances, "to release a force which, by means of subconscious activity, can realize the idea" (Baudouin, 1921, p. 143), in principle the same could happen with an idea that we get to choose or transform by voluntary attention. But, Baudouin (1921) contends, this transposition is not that simple.

Of the three laws "which relate to the preliminary conditions requisite for suggestion, those which show what characters an idea must exhibit if it is to bring about its own realization" (Baudouin, 1921, p. 143) – namely, the law of *auxiliary emotion*, the law of *reversed effort* and the law of *concentrated attention* –, only the second could be integrally determinative in voluntary autosuggestion. Although emotions do have a momentous role in the production of suggestions in general, they "cannot be manufactured to order" (p. 144) – thus, we could not count on their help in voluntary autosuggestion. The law of reversed effort would dictate that, when suggestion is propitious, all of the conscious efforts to *counteract* it tend to be reversed, or transferred, to it, *favouring* it instead. This should certainly be true for autosuggestion in general but would be obviously in conflict with its voluntary version now contemplated[81]. Finally, concentrated attention, as regards autosuggestion only, would have to be stimulated by the same person to be carried away by it, which accords with voluntary autosuggestion but not

[79]What we are also referring to as "intersubjective suggestion".
[80]What we are also referring to as "intrasubjective suggestion".
[81]It would be in such a conflict even when what we consciously or voluntarily struggle to realize is not in conflict with any of our spontaneous autosuggestions, for the idea that the will is being effortful presupposes the idea that it is overcoming a resistance, and so the idea of a resistance is inevitably realized in us (Baudouin, 1921, pp. 145-147).

with the law of reversed effort. Baudouin (1921) chooses to stick to the latter, and has to propose a substitute for the concept of voluntary attention in order to better describe what is at stake in voluntary autosuggestion.

The solution would lie, not in replacing autosuggestion for the will, but in "super-adding" one to the other, in realizing a condition "in which voluntary effort will be minimal, but which will none the less be quite as competent as attention to keep our mind occupied exclusively or almost exclusively with a particular thought" (Baudouin, 1921, p. 151). Voluntary attention should be replaced, he argues, by the processes of *collection* and *contention*. Collection would be a process of relaxation favouring an "outcropping" of the unconscious, which we can recognize today during "mindfulness" or "meditation" states; more democratically, it is the process that occurs between everyone's states of vigilance and slumber. As for contention, it would be a kind of serene or faint attention that can be trained during collection, "a crossways where two contraries meet" (Baudouin, 1921, p. 171).

Although vague, Baudouin's theorizing is just one of the pioneers attempting to grasp with scientific intentions what happens in autohypnosis in order to enable its practice, and autohypnosis, despite its magical undercurrents, is a legitimate scientific topic. Indeed, his theorizing could be applicable, for example, to the currently much-researched phenomenon of lucid dreaming. Perhaps his cloudy account is inevitable considering the nature of the subject matter, which is only approachable through introspection and intuition. Anyhow, the lesson we take from his theorizing is that voluntary autosuggestion is not *exactly* voluntary; so much so that, in the end, Baudouin prefers to call it "reflective autosuggestion" in order to be unaccused of promoting an oxymoron. Although starting with willpower, reflective autosuggestion only works if the person practicing it bows down, so to speak, to the unconscious.

OLP is a perplexing phenomenon in itself; but more perplexing is the fact that it was only recently recognized and taken as a research problem. So much so that common-sense still thinks that, in order to work, a placebo-treatment cannot be disclosed as such. There were first hints of the possibility of placebo-effects without concealment or deception in 1965, but only in 2010 it began to be properly explored and analysed by Ted J. Kaptchuk and his team (Kaptchuk, 2018).

Kaptchuk (2018) defends that, "given its unique features, OLP may involve unique processes and need innovative theoretical foundations" (p. 318). Such foundations could be the models of "prediction and error processing (PEP), Bayesian brain and embodied cognition" (p. 324). The PEP model holds that the brain processes information from the external world and the body through a set of probability-charged representations. According to the model, the brain, equipped with the experience of a lifetime, is continually predicting what it will perceive, and continually subjugating the bottom-up information to its predicting. Once in a while, it is faced with a surrounding and a body that intensely mismatch its predictions and, spending some energy, it is able to correct some of its mental representations[82]. In sum, it must decide how it is going to solve the discrepancy between predictions and prediction-errors.

Kaptchuk (2018) claims that OLP may involve a similar discrepancy: the patients are told they are taking dummy pills, and thus are led to predict that the suffering will not end; but the medical situation itself would trigger a contrary prediction. If it is presented a tree to one of a person's eyes and a face of Albert Einstein to the other, "there is tremendous neurological 'error' or dissonance", thus "instead of seeing some combination of both, a person sees either Einstein's face or the tree (Blake and Wilson 2011; Hohwy, Roepstorff, and Friston 2008; Lee, Blake, and Haeger 2005)" (Kaptchuk, 2018, p. 327). The same happens with open-label placebos: the brain should decide whether the symptoms are better or not, and it ends up assuming the former in most cases, for many possible reasons.

So, we have again an essentially unconscious process. Only the information that the suggestion will take place is conscious in OLP; a large part of its mechanism is not.

An there is more to this story residing in what Keptchuk (2018) called "the medical situation". There is an illocutionary force in saying "this pill won't do anything, but take it anyway" – especially in a medical experiment. From this, the subject could hear "I am a doctor, a scientist, and I believe in an alternative outcome; if I didn't believe in an alternative outcome, I would be cynical and cruel; but I'm a doctor, therefore not cynical and cruel"; or "I want you to take this pill even though I know it won't work, because I have a wish, a pure, wild, wish, that you get better". And so on: it would depend on the specific unconscious representations of the subject of the experiment.

From our analysis of the outlandish cases of reflective autosuggestion and open-label placebo, we can conclude that both the suggestor and the suggestant can be conscious of both the intention to promote suggestion and the information that a mental change is due to it. However, its *causes – the totality of its causes*, from its little triggers to the psychical and social conditions and processes of the persons involved – are always partly unconscious for everyone. As Baudouin (1921) observes, between the idea of a modification and its appearance, there is "a work of realization, performed unconsciously", a second phase which "occurs without our

[82] The model can remind us of some psychoanalytic notions – the conflict between the pleasure and the reality principles, the projective mechanism, and so on – and indeed it is used in the recent neuropsychoanalytic movement.

being aware of it" (p. 38). "Here, then", he concludes, "we have a causal chain whose two ends are within our grasp, but whose centre eludes us" (p. 38). Without this condition, we are not talking of suggestion, but of something else – perhaps of a wilful behaviour.

But is there such a thing as a purely wilful behaviour? Does a mental process of which we are fully conscious, from top to toe, even exist? Of course not. And, if we do not wish to fall into the pit where suggestion has too wide a domain, where its attribution is allowed to all sorts of mental phenomena, we should simply conclude that suggestion is a *level* of motivation in general, a level with a specific kind of functioning; a level predominating in some cases of motivation more than in others.

According to this last point, then, Etchegoyen's "persuasion" (see above in this section), like reflective autosuggestion and open-label placebo, would also be conscious and rational only in the surface. It surely involves a process calling, not only for the attentive and fair part of people, but also for their unconscious meaning structures. Etchegoyen (2005) is citing a classification from the literature, and in the context of psychotherapy he criticizes it, claiming for instance that, although "[Paul] Dubois always tried to differentiate his method [of persuasion] from that of support and suggestion, asserting that persuasion is linked to the rational process, to the patient's reason", his method "is always charged with affectivity", with foundations that are "more rationalizations than reasons" (p. 313).

6.2.1.2.3 The Origin of Suggestion

Is a suggestor indispensable for a suggestion to occur? Does the concept imply the presence of a prestigious or intriguing figure giving a command? Equally important is the question: if such a figure is indeed indispensable, what is more determinative to the phenomena, the words, gestures and postures of the suggestor or the constitutions and dispositions of the suggestant? Any definition of our concept must bear answers to such theoretical questions.

Baudouin's answers are that suggestors are much less determinative to the phenomena than the subjects, and that suggestors could even be absent in a process of suggestion, simply because: 1) It is possible for a person to promote a suggestion in itself (autosuggestion); 2) In cases where one person promotes a suggestion in another (heterosuggestion), the command of the former is not pristinely followed by latter, in other words, the influenced person transforms the command given by the influencer in an idiosyncratic manner, sometimes even contravening it altogether (Baudouin, 1921). In general, I take Baudouin's argument to be very convincing and I think the insights of the New Nancy School which he represents should be included in any definition of suggestion. But I also think he misses an important point, perhaps deliberately, for the sake of rhetoric; he does not discuss the plausible fact that autosuggestion is covertly dependent on an authority figure, a figure which once was bulky but which in the moment of autosuggestion is, of course, a fantasy.

Two phases can be distinguished in the phenomena "universally recognized as belonging to the domain of suggestion" (Baudouin, 1921, p. 21): the mind of the subject accepts an idea that the operator proposes or imposes; this idea is transformed into an action, and the object of the idea is realized. But people would diverge over which of the phases is "essentially and truly characteristic" (p. 22). "Public opinion" would choose the first phase in this respect, holding that suggestion "consists in the forcible control of the comparatively feeble will of the subject by the comparatively powerful will of the operator" (p. 22), and so would some scholars such as Binet (see above in this chapter). This wing defends that the second phase, with its astounding and accurate realizations, is only possible through combining domination and hypnosis, and that for this to be possible a suggestor would be indispensable.

However, this theory would simply be "invalidated by the facts" (Baudouin, 1921, p. 23). At the very least, autosuggestion is a real thing, but Baudouin (1921) wants to resort to a more compelling premise. He tells us about the most impressive results to which the practice of hypnosis has ever arrived – operations over the mechanism of childbirth, in general over many important organic processes – and he defends, with case reports, that a subject can reach similar results when isolated and unaided, that is, without hypnosis and without a suggestor – in other words, he shows that, sometimes, just "the idea of an organic modification can produce that modification in the individual who thinks the idea" (p. 26). What is more, "this action of the idea may be more powerful and more widespread in an isolated subject than in a hypnotized subject" (p. 26).

He also presents along the work we are citing the argument that, even when external triggers have a notable role in a suggestion, the mind of the suggestant is not a cognitive *tabula rasa*, and always makes of the trigger something it is not in itself[83]. A picture may bring to us a sentiment or memory, and

> we feel that the picture is the cause, the "occasional cause" as Malebranche would have said, of the memory or the sentiment. But the deeper cause is to be found in unconscious or subconscious work whose existence we can merely suspect. Something has been stirred in the depths of our personality (Baudouin, 1921, p. 33).

[83] That is, something it is not *intersubjectively*.

By the same token, the author argues that, in hetero- or induced suggestion, the suggestant may even ignore the command, so, in this case, "it would be better to compare the subject's mind to a soil which may be sometimes suitable and sometimes unsuitable for the implanting of the seed of suggestion sown by the operator" (p. 240). He presents less poetic, more evidential versions of such argument, and so do very recent publications by other authors; we shall discuss such versions in a section below about the intra-subjective causes of suggestion.

Anywise, the pair of arguments should make his reader accept that the characteristic element in the process of suggestion is in fact the second phase, the realization of the idea, its transformation into an action.

> The conclusion must therefore be drawn that the presence of a suggester is not essential to suggestion; it is enough to have a subject. In other words, suggestion cannot be defined as a phenomenon of transference wherein the starting-point is the consciousness of the operator and the terminus the consciousness of the subject. It must be defined as a work which proceeds wholly within the subject. If we do not allow ourselves to be repelled by barbarisms, we may find it convenient to say that suggestion is not an "inter-individual" phenomenon but an "intra-individual" phenomenon. Once for all, we must distinguish between the idea of suggestion and the idea of submission, of dependence upon another's will. We must not confound suggestion with subjection (Baudouin, 1921, p. 27).

And, inspired by Coué, he utters some aphorisms containing this radical conclusion: "autosuggestion is really the prototype of all suggestion" (p. 26) and "suggestion is […] nothing more than autosuggestion" (p. 326) are examples. As we have seen above, he does not think that the lack of a commander makes the concept too wide, for still it would expel from its fiefdom the conscious and wilful behaviours, or at least the conscious and wilful levels of behaviour (see p. 224).

In general, I agree with his radical conclusion but I think there is an important caveat to it. Let us follow a logical and theoretical consequence of this conclusion and see where it leads us. Let us accept that every suggestion is, at bottom, a process of autosuggestion. If an external figure does not determine this process, by what exactly is it determined? In autosuggestion, the suggestor and the suggestant are the same person; the essential idea contained in the notion is the idea of one giving commands to oneself. What would then determine such internal commands? Well, it could only be something "bossy" in the person's mind. If it helps, we are free to imagine a bossy homunculus inside the person's head, but of course this would only be a metaphor. But we are not free to deny that there is in the autosuggesting person some psychological structure determining how a specific trigger should be transformed into a compelling task and that this structure, in its turn, was once determined by the person's interaction with people, with "external", actual people.

There is a theoretical presupposition here, the presupposition that the limits and detours we impose to our own behaviour more or less outside consciousness are not innate, or at least not purely innate, but largely acquired from social interactions occurring from babyhood on. This *could* be related to what psychoanalysis tells us regarding the development of the superego or the mechanism of transference, but it need not be; this is, in truth, a truism, found in many different psychological theories of human development.

Humans are social animals. Even in pure autosuggestion, the kick for an idea to realize itself must come from human figures; in pure autosuggestion, specifically, such figures are no longer made of flesh and bones, but of meticulously choreographed neural impulses. We read a text about some disease and therefrom present nocebo effects; this could not have happened without the more or less unconscious idea that it was written by specialists (perhaps wearing white coats) who would surely see the signs of the disease in us were we to meet them in person. When we aim at calming our own breathing rhythm, we try to access a part of ourselves (or one of our selves) which (or who) wishes to calm *someone* down and which (who) succeeds. This part of ourselves is not innate; we have *learned* such behaviours with an actual person who once wished to relax us and succeeded. During suggestion, every event is “humanized”.

What, then, should *radically* determine the process of suggestion, if not external figures? Well, internal ones. And where do they come from? From the lifelong interaction between the suggestants’ dispositions and the many flesh-and-bone figures that at least once in the suggestants’ existence were entitled to give them orders; that is, from *external* figures of authority *made internal*.

6.2.1.2.4 The Centre of the Concept

The analysis of the classical definitions presented above has demonstrated a confusion related to the use of our concept. Whereto is it justifiable to point and say “this is suggestion”? Should the term refer to the stimulus, to the response or to everything happening between each of them, including themselves? Is it the sugar pill given by the figure in white coat, the cure of the headache, or the psychological process that makes sugar pills and white coats become cures? Is it the political propaganda, its interpretation by citizens, or the hate crimes? Is it the analyst’s frowning, the patient’s anxiety because of it, or something else?

I think we should choose the *process between the stimulus and the response* as the centre of the concept as much as it is possible. This is not to condemn everyday uses like “this advertisement is full of dangerous suggestions” or “he would not have exhibited a behaviour like this in other circumstances, so, it was a suggestion", but only a reasonable attempt to avert confusion among scientists and philosophers. I think we should adopt the process as the centre because this adoption: kills controversy; accords with the idea that unconsciousness is an important feature of the concept; lastly, accords with the idea that the intra-subjective dimension of suggestion supersedes but does not exclude its inter-subjective dimension.

The idea of a process should end with the existing controversy among suggestologists because it is all-embracing; to make sense, it must include both the ideas of stimulus and response. Most importantly, it must include a hint of the explanation of what happens in the interval. If one definition is motivating the solution of theoretical problems, and others are not, the former should obviously be favoured over the others.

By the way, to do justice to suggestion's complex determination, we could distinguish a host of elements inside the phenomenon's starting, unfolding and ending. Let us suppose a hypnotizer is showing a yam to a hypnotized subject and aiming to convince him/her it is a delicious apple that should be eaten immediately. We would have the thing to be interpreted and/or handled, more specifically its form or presentation, which we could call the "stimulus" in the strict sense – the yam put before the subject's eyes. We would have the "command", the suggestor's literal speech or gesture expressing how the thing should be interpreted/handled: "This is a sweet, sweet apple; why don't you eat it?". We could also crop the "message" of *what* should be done – this thing should be seen as an apple and the apple should be eaten[84]. On the side of the suggestor, lastly, we would have his/her "tokens of prestige", such as his/her clothes, countenance, posture, voice, relationship to the place where the suggestion-pair is in and to other people there, which would work as an additional message about *why* the *what* should be done: a message like "You better please this powerful person…". And, of course, we have the side of the suggestant: his/her personality, with its fantasies and schemas enabling a certain level of "suggestibility" (that is, of propensity to be influenced) and of "rapport" (affection and confidence toward the suggestor). We have, at the end of the line, the suggestant's behaviour showing how the command and the tokens of prestige were actually introjected. Be it as it may, the term "process" would embrace any element we could possibly distinguish in the phenomenon of suggestion.

The use of the term "process" also weds well with the ideas presented in the preceding sections. First, with what we have concluded about the unconsciousness of suggestion. Some

[84]The "message" is not always a plain reflection of the command itself; see, for example, what happens in so-called "paradoxical interventions" (Weeks & L'Abate, 1982).

elements in the process can be conscious, but never the whole process; if we focused on one of the elements that can be conscious, we would not be sure whether or not we would be talking about suggestion. Lastly, the use of the term is competent at condensing the idea that even if, in heterosuggestion, what happens in the subject is more important than what happens from and around the operator, what happens from and around the operator is not at all negligible.

6.2.1.2.5 A Good-Enough Definition

We can now propose a definition for our concept. Suggestion is a *partially unconscious process in which certain intra- and intersubjective circumstances cause a subject's mental state to conform itself to his or her representations of an authority's wish*, as much as his or her mental and/or physiological constitution allows; it is a kind of process that occurs in any instance of motivation, predominating in some of them.

This definition is hopefully reaching the intuition of those familiarized with the concept and also reflecting the main theoretical conclusions we have arrived at up until now. Importantly, it expresses the "every suggestion is at bottom autosuggestion" maxim – through the elements "intra-", "itself", "representations", and "as much as his or her mental constitution allows" – without forgetting the pivotal role of a now-fantasized-but-once-animated figure of authority.

Baudouin (1921) reminds us that a definition is a theory. Our definition is surely sketching a theory, but it does not specify the intra- and intersubjective conditions for suggestion. Let us go through some of them in order to completely establish what suggestion is – and then head, finally, to showing the fundamental weakness in Grünbaum's contamination argument.

6.2.2 The Conditions for Suggestion

6.2.2.1 Conditions External to the Subject

In the beginning of the 20th century, heterosuggestion was regarded an extremely complex phenomenon, one affected by numerous and obscure factors (Baudouin, 1921). And, as late as in 1989 Gheorghiu asserted that, from the 1900 pioneer book by Alfred Binet ("*La suggestibilité*") on, three questions had remained alive in the research on suggestibility:

> 1. Should suggestion be understood as a single component or as a multi-dimensional phenomenon?
> 2. Are there relationships between suggestibility and other personality characteristics or between suggestion and other psychological phenomena?
> 3. Is it possible to identify factors involved in the production of responses to suggestion [i.e., to suggestive actions]? (Gheorghiu, 1989a, p. 12).

Indeed, this last question (and its relation to questions 1 and 2) has remained until our days painfully difficult to answer; notwithstanding, it has also been poorly tackled. Not long ago, Peter Halligan and David Oakley assented to Bisiach and Rusconi (1990) in that "it might be the case that we find certain aspects of suggestion puzzling, because we do not find the whole business of suggestion puzzling enough" (Halligan & Oakley, 2014, p. 118). The same authors claim that

> despite being a common and significant feature in many forms of human behaviour, (Schumaker, 1991) as well as a key element responsible for generating the broad range of subjective experiences and behaviours produced in hypnosis (Kihlstrom, 2008), suggestibility has received comparatively little attention (Gheorghiu et al, 1989; Lundh, L., 1998; Kirsch et al, 2011; Michael, Garry, & Kirsch, 2012; Schumaker, 1991) […]
> […] fundamental aspects of suggestion and suggestibility have been relatively unexplored, and "the importance of treating suggestion as an important domain in its own right has been largely ignored" (Kirsch et al, 2011) (Halligan & Oakley, 2014, p. 105).

Despite this deficit, here are some empirically substantiated and relatively well-established points regarding the conditions related to the process of suggestion.

It is now clear for the reader that the factors enhancing the probability of suggestion can be tracked "around" the suggestor or "inside" the potential suggestant. A subject conforms to an authority's wish more easily in a situation of ambivalence or inconsistency, but a situation cannot boost a suggestive response without there being in the subject a specific meaning and motivating structure. A discussion of the strength each of these poles, the external and the internal, has in the causation of suggestion is cardinal for the argument of this chapter.

That a situation of *ambivalence* or *inconsistency* is a precondition for suggestion is "one of the few findings that have remained undisputed in research on suggestion" (Gheorghiu, 1989a, p. 19). The finding date back to the classic experiments of Muzafer Sherif and Alfred Binet. Sherif (1936) demonstrated that the less precise and clear the structure of a stimulus before us, the higher the probability that we attune to someone else's interpretation of this stimulus. His team presented to subjects a tiny point of light that, in total darkness, elicited a false perception of motion – a so-called autokinetic visual illusion. The subjects were asked to estimate the distance apparently covered by the point of light, alone and in a group; some were asked in the group, then alone, others were asked alone, then in the group. As the subjects tended to keep the group's reports of the visual illusion or to conform to the group's reports of the visual illusion, the researchers concluded that an unstructured situation is a good condition for suggestion.

> In this situation, within certain limits, there is no “right” or “wrong” judgment. One subject demonstrated this spontaneously during the experiment, in spite of the fact that he was not supposed to talk: “If you tell me once how much I am mistaken, all my judgments will be better.” Not being sure about the correctness of his judgments, the subject feels uneasy. This we know from the introspective reports. […] In the group situation the members of the group tend to structure the situation by converging toward a common norm in their judgments. If in the beginning of the experimental session they start with divergent judgments, in the course of the experiment they come together, the divergent one feeling uncertain and even insecure in the deviating position of his judgments (p. 107).

Binet (1900), in his turn, made experiments in which “the objective was not so much to utilize absence of structure in the stimulus […], but rather […] to create a state […] that would impede the use of adequate control mechanisms” (Gheorghiu, 1989a, p. 19). He has investigated the effects of so-called “leading questions”, questions phrased in a way as to imply expectations and premises. Another technique he utilized involved casting doubt on the subjects’ correct judgements or making incorrect judgements of a clear state of affairs. In sum, it is classically established that whatever installs uncertainty about the reports of our experiences, be it a technique or a random state-of-affairs, cleans the stage for a suggestion to act. “Binet (1900) emphasizes that suggestibility concerning a particular point is in inverse relation to the person’s degree of certainty on this point. The fainter or more ambiguous a stimulus, the more easily it can be given a new interpretation” (Gheorghiu, 1989a, p. 19).

It is difficult to put aside the thought that this basic character of suggestible situations is also one of the basic characters of a psychoanalytic treatment. Let us compare the first reports and associations that patients present during treatment to the “stimulus” of Sherif’s experiment, for example. For sure, such reports and associations are quite ambivalent and inconsistent and the patients do not know how to make sense of them; indeed, it is reasonable to claim that every patient’s motivation to start a treatment is related to this last fact. On their part, analysts, according to most approaches, should not undo the uncertainty of the patients right away; it is held that giving premature and complete explanations to patients invigorates their repressive mechanisms. The outsets of psychoanalytic treatments, therefore, may well be considered excellent lands for the cultivation of suggestion.

One more word about Binet’s “leading questions”. As indicated above, they are questions phrased to promote specific responses by implying expectations and premises. The expectations of a question are expressed in its words, its syntax and even its intonation; by controlling such elements, we are able to subside or eliminate certain interrogative expectations. On the other hand, there is no such thing as a question without half-hidden premises. A question with two alternatives, for example, has a surreptitious premise: that there is no other alternative in that context. Preventable or not, the elements of leading questions may be well or ill supported and, clearly, well supported expectations and premises within an interrogation are just signals that the respective investigation is not in its initial phase (Gudjonsson, 1987).

The effects of leading questions in the context of police interrogations have been objects of research since the end of the 20th century. In 1984, Gisli Gudjonsson published a suggestibility scale to evaluate complying responses to misleading questions, which he called "Yield" responses, and complying responses to negative feedbacks of the questioner – that is, changes of report following the questioner's overt dissatisfaction and/or pressure before an answer – which he called "Shift" responses. Since the application of the scale showed that "Yield" and "Shift" were not consistently correlated in the same subjects, Gudjonsson (1984) proposed that one kind of response should be theoretically separated from the other, as they would reflect different tendencies.

Research has also demonstrated, however, that the pair of responses, as a cluster, did not correlate to sensory or motor types of suggestibility. Hence, Gudjonsson (1984) coined a term to refer to a subject's tendency to "yield" or "shift" while in an interrogation: interrogative suggestibility. Gudjonsson and Clark (1986) could then propose a theory about the conditions that, in a witness or a suspect, would enhance the chances of them accepting incorrect premises and expectations. There would be three of such conditions: uncertainty as to the correct answer to the questions presented; confidence on the interrogator; expectations of giving satisfactory answers (Gudjonsson, 1987).

Szpitalak and Polczyk (2015) take this last condition to be the most vital one, for if there is only uncertainty about the answer and confidence on the interrogator, the subject is likely to answer a truthful "I don't know". "If the witness of a given event is convinced that lack of knowledge about the event is unacceptable", on the other hand, "they may fill the gaps in their memory using information from sources other than the event itself" (p. 2). The two psychologists, then, designed an experiment to test whether to reduce the answer expectations through the technique of reinforced self-affirmation (RSA) would reduce interrogative suggestibility as well. The technique of RSA consisted in asking the subjects of the interrogation to record their greatest achievements and in giving these subjects positive feedback about their performance in a memory task. The hypothesis was confirmed: both Yield and Shift were lower in the RSA Group in comparison to the Control Group. "This concurs with the suggestion that greater self-confidence helps people resist the effects of both misleading questions and negative feedback" (p. 6). Close to the end of this chapter, we shall discuss the epistemological consequence of researches such as this one.

6.2.2.2 Conditions Internal to the Subject

We have just discussed the "external" pole of causation in the process of suggestion. If we thought this pole were the only one or the only significant one, we would probably assume that *any* command is undertaken by suggestants, *in the most undistorted manner*, in every case where the right external circumstances are present. Suggestants, thus, would appear to us merely as responsive fleshes, as sharply programmed robots. But we know that, besides the framework and the actions on the side of the suggestor, the subjectivity of the suggestant is an important determinant of the process we are now discussing; what is more, both anecdotal and experimental evidence indicate that, in this area, *internal conditions are more determinant than external ones*.

The experimenter Gheorghiu (1989a) agrees with it in commenting that, although always reliant on other people and groups, individuals are reliant "to a still higher degree […] on their own personality traits as well as interests, attitudes, and belief systems" and that all of the latter "– according to the situation – induce them more or less strongly to react to suggestive stimuli" (p. 43). He claims that

> […] the demand nature of a suggestion situation cannot be adequately defined without taking the subject into account. The suggestion stimulus in itself is not yet self-evident, credible, surprising, infectious, etc. The prestige of the suggestor by itself can sometimes evoke suggestible responses more strongly than the spoken word. But there is no prestige until the subject attributes it to the suggestor. Lewin (1946) views the demand characteristics – and the environment factor in general – as variables which are principally not independent of the person (Gheorghiu, 1989b, p. 104).

Little wonder, then, that certain terms usual in theories on suggestion indicate its connection to the dimension of motivation, and that "the formation of a suggestive situation and its feasible effects has often been related to motivational factors (Allport, 1961; Krech & Crutchfield, 1948; McDougall, 1908; Stokvis & Pflanz, 1961; Stukat, 1958)" (p. 102).

The most basic fact that empirical research on suggestion inform us is that different subjects embody suggestions to very different degrees, even under the same situation. This fact, in itself, already eliminates the hypothesis that the external pole is fateful. We could assume, though, that there are subjects with constitutions predisposing them to comply indiscriminately – in any "suggestible situation" – and straightforwardly – undertaking the exact behaviours that some command tells them to undertake. But empirical evidence eliminates this hypothesis as well.

Since the time of the magnetizers, Ellenberger (1970) teaches us, there is an understanding "that no person can be hypnotized against his will" (p. 114) and that hypnotic suggestions are "not necessarily forced upon the subject" (p. 150). Pioneers such as James Braid were aware of the existence of autosuggestion already before the Nancy School. Already in that first stage of research,

> imperative suggestions were found to work best with persons who occupied subordinate positions in life and were accustomed to obeying orders (soldiers and laborers) or with people whose willpower was weak or who were eager to submit their will to that of the hypnotist. But even in such cases, the power of imperative suggestion had its limitations. When used with a person who was unwilling to submit, it was found that there was either no success at all, or else only a temporary removal of the symptom, which then reappeared or was replaced by another one (Ellenberger, 1970, pp. 150-151).

By the interwar years, Baudouin (1921) argued that even one of the first laws of suggestologists – that suggestibility is higher in a slumberous ("hypnotic") state than in a waking state, that inside slumberous states a subject bears more chances of realizing whatever the operator's command turns out to be – is liable to variation. He contends that every suggestologist has come across subjects "who are less obedient in induced sleep than they are in the waking state" (p. 239) and also across some "who are most docile in the hypnotic state, [but] none the less quite unexpectedly impose a plea of exception to certain suggestions" (pp. 239-240).

> When a subject who is customarily docile opposes a plea of exception to some particular suggestion, we must not attempt to explain this refusal as a deliberate exercise of will. The question is always one of the soil on which the seed has fallen. We shall find, on close examination, that the suggestion has involved some infraction of the subject's deeply-rooted tendencies, that it conflicts with his character, with his inveterate habits (Baudouin, 1921, p. 241).

Baudouin (1921) explains these apparently contradictory facts with a "theory of autosuggestion", which by now we may know by heart: the command of the operator can only be realized "on condition that it is accepted by the subject's mind and transformed into an autosuggestion" (p. 240).

During World War II, Thomas Coffin and his colleagues experimentally demonstrated that the stimuli to which subjects respond compliantly bear clues about who the subjects already were before their compliant responses. Their research was undertaken in the winter of 1939-1940, when, Coffin (1941) had learned "from casual conversation" with college folk, "opinion on the war was less unified" and "there were individuals representing all shades of opinion toward the belligerents" (p. 32). The experimenters tested 226 subjects on their attitudes toward the Allies' and the Germans' political stances and actions; then, they showed political propaganda sympathetic to each of the sides, and asked the subjects to label them "true", "false", "probable" or "improbable". The results indicated a close relationship between previous attitude and acceptance of propaganda: the strongly pro-Allies group accepted over four times as many pro-Allied propaganda as did the strongly pro-German group, while neutral groups accepted an intermediate number of both (p. 45).

The mention of this old study may conjure up current psychosocial discussions involving "fake news", the name the 21st century chose for "political propaganda". After the 2016 and 2018 presidential elections in the United States and in Brazil, respectively, many had the intuition that fake news are not sorts of evil spells, that voters are not blank slates, and that the idea of deceit does not provide the key to understand the far-right's return to official politics. It was largely discussed that the ethical slips and the paranoid narratives implicit in the fake news were already part of some citizens' minds. This hypothesis would explain why so many were accepting improbable and extreme ideas and speaking up for them in so short a time. Fake news would thus be mere triggers – triggers to just a part of the citizens' subjectivity, for sure, but still mere triggers.

More recently, Rotaru and Dafinoiu (2014) experimentally investigated the connection between people's attachment styles and their responses to nonhypnotic suggestions. The experimenters departed from the idea that people's "significant models on the personal value of the Self and the readiness and responsiveness of the Other" would "influence the magnitude of […] [their] response to suggestion" (p. 199); attachment styles would be, thus, meaning structures which are activated during suggestion. There would be four adult attachment styles: the secure style, presenting low anxiety and low avoidance; the anxious-preoccupied, presenting high anxiety and low avoidance; the avoidant, displaying high avoidance and low anxiety; last but not least, the anxious-fearful style, which is one of high anxiety and high avoidance. Their results indicated a connection between attachment avoidance and low suggestibility – "explained by the fact that attachment avoidance involves an overestimation of personal competence" (p. 210) – and also another, incongruent with their previous hypothesis, between high attachment anxiety and low suggestibility. This would render people with anxious-fearful attachment style the least responsive to suggestion and the secure participants the most. The authors explain this result by proposing

> that a higher anxiety does not involve higher reliance on the Other as compensation for a negative model of the Self. On the contrary, the instability of the attachment figures is likely to

> generate the feeling that the others' responsiveness is unpredictable, which leads to a marked distrust expressed in all social relations of the adult (Rotaru & Dafinoiu, 2014, pp. 211).

Such experiments highlight the motivational factors that assign to a figure a prestigious character and to a situation a suggestive character. But, beyond this, subjects in a process of suggestion also tend to *alter the given message or command* through their previous memories and other mental resources, disambiguating the natural ambiguities of the message, making a specific sense of it and extensively elaborating on it.

An experiment undertaken by Etzel Cardeña and Devin Terhune, the results of which were published as recently as 2019, showed that people's nonhypnotic baseline experiences and expectancies correlate moderately to strongly with their spontaneous, unsuggested, idiosyncratic experiences during a neutral hypnotic procedure. Before being hypnotically induced, participants were asked to complete a Phenomenology of Consciousness Inventory (PCI; Pekala, 1991) to register their expectancies about how their experiences would be in a deep level of hypnosis, as well as to register their actual experiences during "a nonhypnotic baseline condition involving a 2-minute eyes closed resting period in which they were asked not to practice meditation or hypnosis but just to sit quietly and relax" (Cardeña & Terhune, 2019, p. 7). Then, a hypnosis induction was administered "with the single suggestion to go into a deep state of hypnosis" (p. 7), during and after which the participants were asked to report their experiences (the after-report also had the help of the PCI). The results proved that

> multiple dimensions in both baseline and expectancy experiences correlated moderately to strongly with the experiences during hypnosis. The correlations between baseline and hypnosis tended to refer more to everyday experiences, such as affectivity and memory, whereas those between expectancies and hypnosis also referred to alterations of consciousness. Response expectancies for the impact of hypnosis predicted experiences during hypnosis even after controlling for baseline and hypnotizability. The results overall suggest that spontaneous experiences during neutral hypnosis are driven by a confluence of variables, including a general propensity to experience alterations of consciousness as well as expectancies regarding one's experiential response to a hypnotic procedure, partly informed by baseline resting state experiences (Cardeña & Terhune, 2019, p. 18).

Hypnotized suggestants make their own "contributions" to the suggestor's hypnotic command even when this command is far from "neutral", for example in the hypnotic induction of a specific symptom, such as blindness, limb paralysis, memory loss, hallucination, analgesia, pain, etc.. Halligan and Oakley (2014) tell us that

> […] [targeted] suggestions [to generate clinical analogues] can only convey a selective and small number of propositions – whereas the subject's responses often produces [sic] a much richer, more comprehensive set of experiences and responses (see supplementary data in Oakley & Halligan, 2009) requiring subject's *to go beyond and supplement the specific suggestion with existing or prior knowledge* [emphasis added] to produce the eventual clinical analogue. Evidence from placebo and nocebo studies also suggest that the subject's cultural knowledge regarding medicine (e.g., beliefs) and its practitioners may not only provide for therapeutic

> interventions but also help to generate conditions of illness by activating latent expectations (Hahn, 1997) (p. 116).

This supplementation is known since the time of Auguste Forel, reminds us Baudouin (1921), as Forel argued about "how the subject completes the suggestions made to him, completes them with the aid of elements borrowed from his own personality" (p. 237). Forel noted this while inducing hallucinations:

> if to a subject in the somnambulist state we propose a negative hallucination (that is to say, if we suggest to him that he is unable to see a real object), there ensues a hiatus in his visual field, and he invariably fills in this hiatus in some way, with the aid of a positive autosuggested hallucination. In like manner, we cannot induce a positive hallucination without this involving for the subject the production of a negative autosuggested hallucination indispensable to the occurrence of the positive hallucination (Baudouin, 1921, p. 238).

After praising the French suggestologist, Baudouin (1921) presents another anecdotical evidence from the sweat of his master, Coué. After a command to see an apparition clothed in white on the right-hand window pane, each of a group of Coué's suggestants would describe it in a different manner and some of them would not see it on that specific window. Baudouin (1921) took such episodes as proof that what is realized in suggestion is not the will and the thought of the operator, but the interpretation of his words by the subject – the images the command evokes in the subject's mind. He writes the ill-recognized obvious: "There is no question of the conduction, as by invisible wires, of the operator's thought into the subject's brain[…][; it is] the subject's own thought is conducted by the subject's own nervous system" (p. 239).

In sum, a suggestion response shows not only that it has *something* to do with the subject's attachment-related and mnemonic history; if we just regard carefully the context in which it happened and the response excesses and absences, it also shows *details* of this history. If we just look and hear, we realize that the subject is not passive in suggestion at all.

6.3 IS ELIMINATING SUGGESTION REALLY DESIRABLE?

6.3.1 Clinical Psychoanalysis Is All About Suggestion

Let me come back to our definition and main theoretical point: suggestion is the partially unconscious process, current in all instances of motivation but preponderant in some of them, whereby our mental states change in conformity with what, according to our fantasy, an authority figure expects from us; it always has an external trigger, but the core of the explanation for our suggestion responses lies in our psychic structure. This means that such responses, although seemingly just complying to an external command, are largely expressing aspects of the subject's mental life.

We could take the concept of "caregiver" to be inside the scope of the concept of "authority" – or we could consider them synonyms. In our childhood, a caregiver is a person upon whom we depend to survive; positively, as this person keeps us healthy and makes us thrive, or negatively, as this "terrible giant" refrains from harming us. We have the instincts and the intelligence to realize this, and to know that their warnings, lessons and wishes should be followed; thereby, we overcome the challenges of existence and maintain a rich source of wisdom and power by our side. An authority is nothing more than a caregiver in our mental fauna, and vice-versa – for why would we be instruments of someone's wishes if our life did not depend on this? We may conclude, thus, that it is the subject's history of relations with caregivers, which is ingrained in his or her psychic structure, that tells us *if* and *how* a command, proposition or stimulus will be processed by him or her.

Now, according to psychoanalysis, the patients' symptoms, parapraxes, dreams, jokes[85], etc., are ultimately caused by their childhood fantasies around a caregiving figure, specifically by their precarious theories on how to please this figure in order to conquer or keep its love. Our unconscious formations, thus, would simply be attempts to work through our flaws with the others' demands.

We are here pushed to the unorthodox conclusion that the phenomena toward which the wits of analysts are dragged could be confounded with the set of suggestion phenomena. Panic attacks, opportune oblivions, sexual dreams, a guffaw, are all examples of spontaneous autosuggestions. Such phenomena are nothing more than unconscious processes whereby mental states conform to a representation of the wish of an authority/caregiving figure.

Note that I am not merely defending that symptoms, dreams, etc., are *analogous* to suggestion, *nor solely* that the suggestion of common-sense, induced suggestion (see p. 221), could be *explained* by psychoanalytic theory. The point I am defending now is more radical: I am defending that *the domain of suggestion phenomena comprises all of the domain of clinical psychoanalysis and could be merged with it*, since any human behaviour with an unconscious motivation could be understood as a suggestion. Every "psychoanalytic" mental product is a

[85]A "joke" in the context of unconscious formations would include not only its creation, but also its appreciation and sharing.

suggestion, that is, an influence digested by our subjectivity. The two domains may have different names, but they have the same elements.

If the reader is not convinced by my conceptual and theoretical study of suggestion and thinks that the latter should be restricted to heterosuggestion, or even further, to intentional heterosuggestion, still we could say that it is a candidate to be explained by the psychoanalytic theory.

Clinical psychoanalysis is all about suggestion; at the very least, analysts should be interested in explaining it. Either way, the factors stirring a process of suggestion cannot be seen, in clinical psychoanalysis, as – to use the experimental jargon – extraneous variables, as Grünbaum presumes they must be. They must be seen as relevant variables – as, precisely, the (or some of the) variables under scrutiny along the treatment of a patient. Cognitive conformity is indeed an intrusive process in any experiment that does not investigate it, and Grünbaum thinks this is enough to claim it is impertinent to clinical psychoanalysis as well. But how could an ordinary-yet-mysterious psychosocial mechanism be something impertinent for an analyst?

So, when there is evidence that cognitive suggestions took place in clinical psychoanalysis – in which case, according to Grünbaum, it would be evidence of some *contamination* of the clinical data – the researcher-psychoanalysts should not necessarily declare a "quarantine"; for the objects of the researcher-psychoanalysts are, rightfully, "infections", "transmissions" and "incubations" and their matter, how such "virulence" became possible.

Cognitive suggestions are as analysable as any symptom or dream, whether because symptoms and dreams are suggestions ("spontaneous suggestions"), whether because "induced suggestions" can be considered an unconscious formation as well. Why was the interpretation *x* – from the analyst, from a friend, from a book, etc. – accepted by the patient? What is the subjective context that made it possible? How does the patient make sense of it? What is the position the patient wishes to put the analyst in by accepting *x*? What does the patient wish the analyst to know and not to know? All this may be asked. The problem is not if cognitive suggestion has occurred, but if it was not perceived and taken as valuable data by the analyst.

Let us imagine a patient who is also an undergraduate student of some social science and who has heard something about psychoanalysis. He would come to therapy one day and say: "I decided to read Freud's study of the Rat Man this week, and now I am sure I am an obsessional neurotic". From then on, his reports would start to resemble the ones contained in the case study. Now, this would only be an epistemological problem if the analyst in question did not understand that it matters less that the patient's reports are inauthentic, artificial, etc., than the facts that: the patient went after that text; he found it very interesting; he compared himself with certain traits of the Rat Man; and he felt the urge to tell his analyst about it. Many unconscious motives would be compatible with such facts, and the analyst in question would have to gather more data about the patient to know which one is the best explanation for the patient's behaviour. This example shows, anyway, that suggestion in general is more an opportunity than a problem for the analyst-researcher.

This is a deflation of the threat of cognitive suggestion; a recognition that it is a strawman, stuffed with the same straw of the land he stands upon, and a recognition that the straw of both the land and the man may be harvested.

Of course, Grünbaum could reply that this is forceful under the condition that psychoanalytic theory be right about the phenomena it predicates, including about suggestion; that this is not the case, and, further, that our understanding of suggestion, even if inspired by many non-psychoanalytic accounts, is also wrong. However, as discussed at the beginning of this chapter, he did not disclose too much of what he was taking "suggestion" to imply, both conceptually and theoretically. Moreover, the point now is not that psychoanalytic theory is right, only that, according to its own standards, it cannot in principle be jeopardized by suggestion phenomena.

Since Freud, analysts recognize that their very act of trying to know patients affects them – an aspect some social scientists nowadays call "reflexivity" (Frosh & Baraitser, 2008). Indeed, one of the functions of the concept of transference was to show that this reflexivity, a potential epistemic obstacle, could be transformed into a potential epistemic advantage when "included in the database" and interpreted. To be coherent with our previous discussions, "transference", the transference of a childhood representation to a fresh representation, could be taken as an *internal cause* of suggestion. Freud even affirmed that analysts should operate transferential suggestions to aid the patients to overcome their resistances (e.g., Freud, 1963a). Baudouin (2015) was especially explicit about this point:

> In many cases, a subject has nothing whatever to say regarding some memory or association; or the subject has merely the vague sense that there is a reminiscence which cannot be clearly recalled. In such instances, I have sometimes taken out my watch and, showing it to the subject before laying it on the table, I have declared that the missing memory would become perfectly clear in five minutes. We would then go on talking of other things, the subject not consciously watching the passage of time; but after the lapse of five minutes, neither more or less, our conversation would be interrupted by the explanation: "Ah, now I remember!"
> Above all, suggestion can aid the clarifying process of the analysis if the operator assures the subject, once for all, that the latter's reminiscences will be revived with increasing facility, and that dreams will in future be remembered without any difficulty. A direct suggestion may even be given that the subject will be *less and less inclined to repress* what is now hidden in the subconscious. (p. 112)

Freud's discretion explains why this old recognition that transference and suggestion are auspicious words was forgotten in most responses to Grünbaum. Freud struggled to have his figure unattached to that of the "charlatan", to have his method and treatment as respected as other medical treatments and scientific methods of his time. His time and his community told him this could only be done by moving his method away from the suggestive ones. Because of the Father, "suggestion" became a dirty word for the Children. Grünbaum was able to bewilder analysts because he took advantage of this historical aversion and of the conceptual and theoretical blur it ensued.

But before "suggestion" became for Freud a dirty word, before he even inaugurated psychoanalysis, the doctor related it to what he would later treat as the domain of unconscious processes. In the preface to his German translation of Bernheim's 1886 book "*De la suggestion et de ses applications à la thérapeutique*", he attempts to integrate Bernheim's and Charcot's different understandings regarding suggestion phenomena (Freud, 1966). In the accommodating arguments of this preface, Freud (1966) shows that, were his time's medical climate different, he would have been able to someday explicitly defend that his dynamic unconscious is entirely reducible to the domain of suggestion processes. There he states, for example, that "there is little to be gained by calling suggestions 'obsessional ideas' […]" because "it seems likely that more light will be thrown on obsessional ideas by comparing them with suggestions than the other way round" (pp. 76-77). In other parts of the text, he is even more assertive in this respect.

We have seen earlier in this chapter how, along the 1880s, two suggestion doctrines have become influent and their intellectual contacts, conflictive. According to Freud (1966), Bernheim thought that *hypnotism phenomena are caused by suggestion*, and that suggestion is a *conscious* idea that an *external* influence introduces into a subject's brain. Years later Freud would propose that something could be mental and unconscious at the same time, but for the European science of 1888 "conscious" was all that "mental" could mean; so he summarized Bernheim's position as that which took hypnotism phenomena as essentially *psychological* – as linked "with familiar phenomena of *normal* [emphasis added] psychological life and of sleep" (p. 75). Charcot, on the other hand, would have theorized *at least some of the manifestations of hypnotism as caused by pathophysiological mechanisms*, namely, displacements of excitability in the nervous system which are *out of consciousness' reach*. Again, for that context's state-of-the-art, the mind was one thing and the non-conscious nervous processes, another, because this was before Freud himself equated the mind with the conscious representations *plus* the non-conscious nervous processes that precede and condition such representations. In reading his 1895 "Project for Scientific Psychology" and comparing it to the rest of his writings, one may see that he had non-conscious nervous processes in mind when proposing the concept of unconscious representation, and that he never renounced his initial understanding that the nature of unconscious representations is essentially neurological (Caropreso & Simanke, 2011).

Well, Freud (1966) notes, if Bernheim and his Nancy School are right, Charcot's observations and the theory they support about the regular symptomatology of hypnotism in hysterical patients are worthless, for then "every physician would be free to produce any symptomatology that he liked in the patients he hypnotized" and the conclusion would be that "it would have no characteristics of its own" (Freud, 1966, p. 78). Here we see Freud aware of the epistemological catastrophe ensued by the idea of a universal, undiscriminated and straightforward suggestibility. If Bernheim is right, says Freud (1966), what we learn with Charcot is not "what alterations in excitability succeed one another in the nervous system of hysterical patients in response to certain kinds of intervention", but only "what intentions Charcot suggested (in a manner of which he himself was unconscious) to the subjects of his experiments – a thing entirely irrelevant to our understanding alike of hypnosis and of hysteria" (p. 78).

Freud (1966) is not willing to abandon the robust epistemic consequences of Charcot's premise, but, as Bernheim's psychological approach to hypnosis and psychopathology still strikes him as fruitful (he translated the man's book to German, after all), he adumbrates a synthesis of the two theories. Charcot should be right because the main points of his descriptions (anaesthesia, paralyses, contractures, etc.) parallel "reports coming from past times and from distant lands" (p. 79); hysterical major hypnotism, thus, should have a physiological nature. But the phenomena of "normal minor hypnotism" would indeed, as Bernheim defended, transpire through psychological means – through suggestion. Freud (1966) concludes that "there are both psychical and physiological phenomena in hypnotism, and hypnosis itself can be brought about in the one manner or the other" (p. 81). But this dualism leaves him discontented, and he strives to "give some indication of the connecting link" between the two dimensions of hypnotism (p. 82).

He then reaches the same principle defended throughout this chapter, even though he does not generalize it to every person like we do in here. In a fraction of cases of suggestion, he affirms, "the part played by [the] external stimulus is evidently smaller and the part played by the physiological condition of the subject […] is greater than in the […] [rest of the] cases"

(p. 82). It would be a question then "not so much of suggestions as of stimulation to *autosuggestions*" and the latter, "as anyone can see, contain an objective factor, independent of the physician's will" (p. 83). He brings psychopathology closer to the notion of suggestion once again with a claim that hysteria would be better explained by an inclination to such autosuggestions than by compliance to the physician (p. 83).

He calls them *indirect suggestions*, because between its external stimulus and its result there would be "a series of intermediate links arising from the subject's own activity" (p. 83); indirect suggestions would "reveal a connection between various conditions of innervation or excitation in the nervous system" (p. 83), containing thus a great epistemic value.

The reader may interrupt my hermeneutical proof: "but the nervous system has nothing to do with the principle defended in this chapter, that suggestion is not extraneous to research on unconscious motivation; you said it yourself, Freud's mind-body monism and his concept of unconscious mind related to it were not yet fully formed in 1888". They were *partly* formed, though:

> Indirect suggestions […] are […] psychical processes; but they are no longer exposed to the full light of consciousness which falls upon direct suggestions. […] Indirect suggestions or autosuggestions can accordingly be described equally as physiological or as psychical phenomena […]
> […] There is no justification for making […] a contrast […] between the cerebral cortex and the rest of the nervous system: it is improbable that so profound a functional change in the cerebral cortex could occur unaccompanied by significant changes in the excitability of the other parts of the brain. We possess no criterion which enables us to distinguish exactly between a psychical process and a physiological one, between an act occurring in the cerebral cortex and one occurring in the sub-cortical substance; for 'consciousness', whatever that may be, is not attached to every activity of the cerebral cortex, nor is it always attached in an equal degree to any particular one of its activities; it is not a thing which is bound up with any locality in the nervous system (Freud, 1966, pp. 83-84).

If Freud thought that some suggestions are indirect, that is, only stimulations to autosuggestions, that the latter depend on the nervous constitution of the individual, and that the psychical sphere cannot be separated from this nervous constitution (a psychical-nervous constitution he would later call the subject's unconscious), then he thought that some suggestions are richly informative – that they could indeed be objects for a scientist interested in the mind-body interface. He writes that one can sustain an antithesis between the psychical and the physiological phenomena of hypnosis only if suggestion is understood as "a directly psychical influence exercised by the physician which forced any symptomatology it liked upon the hypnotized subject", but that this meaning vanishes when one realizes that "suggestion *only releases sets of manifestations which are based upon the functional peculiarities of the hypnotized nervous system* [emphasis added], and that in hypnosis characteristics of the nervous system other than suggestibility make themselves felt as well" (p. 84). In the end, the term "suggestion" would have, for him, "the same meaning as the reciprocal arousing of psychical states according to the laws of association" (p. 83).

It is clear, thus, that the core of our deflation argument can be derived from the first sprouts of psychoanalytic theory, even though some orthodox Freudians would shiver at our suggestion that this theory is, after all, all about suggestion. But those who get chills with this idea today "are like the child who wanted to go to the Midnight Mass if only it could be celebrated in the daytime" (Baudouin, 2015, p. 115).

6.3.2 Clinical Suggestion Must be Reduced, Estimated and Explained

What does this deflation imply? It surely does not imply that analysts are freed from dealing with the problem of the "contamination" of their patients' thoughts and reports by interpersonal influences, in other words, the problem of induced cognitive suggestion. If anything, it indicates the exact opposite: as instances of induced cognitive suggestion should be seen within the borders of their profession as primary data, they should devote a great deal of attention to such instances. A whimsical guest is no less problematic than an intruder.

Analysts should explain such instances just like they explain other instances of their patients' behaviour. They should, in Freudian terms, "analyse" induced cognitive suggestion. But before they are able to analyse it, they have to know: (1) whether it has occurred; (2) whether it was fostered/triggered inside or outside the clinical context. The true challenge in Grünbaum, thus, is his argument that the clinical method *does not* allow us to know (1) and (2). *Could* the clinical method allow it, though?

Analysts should know (2) to keep track of the influence they exert through word and deed over their patients' cognition. To be clear, they should reduce this influence as much as possible and estimate the features and forces of the cognitive suggestions they could not help but promote. After all, the conclusion of our previous discussions over the conditions for suggestion was only that environmental conditions determine it less than personal ones, not that environmental conditions are negligible.

But, if the domain of suggestion and the domain of clinical psychoanalysis are the same, as argued in the previous section, why should analysts reduce and estimate their influence over patients? If suggestion, in the wide meaning we have vindicated for it, is *the* explanatory target of analysts, why can analysts not loosely foster suggestion, as if they were experimenters? There is a therapeutical/ethical argument dictating the reduction of certain suggestions: all the pernicious ones, spontaneous or induced, triggered inside or outside the clinical context. (It should be reminded that, according to the deflation argument, a symptom is a suggestion). But now we are only in the epistemological dimension of clinical psychoanalysis, and in need of an epistemological justification for the intuition that analysts should try to control their cognitive influence over patients. From the epistemological perspective, extraclinical cognitive suggestions are valuable data; so, why can intraclinical suggestions not be fostered with the aim of making patients dispense more and more valuable data? Let us imagine that an analyst said to his patient she has "penis envy", even if he did not consider this to be true; according to the arguments presented in this chapter, he would gain valuable information about the patient's psyche just from how she reacts to this "interpretation".

Let us leave aside the matter of how unethical this would be, that is, let us consider only the epistemological perspective. If the deflation argument implies that analysts could attempt to obtain information about the patients' inner life by making them "gestate" any idea and by observing what comes out of them afterwards, then the argument should be restricted. Analysts indeed should carefully reduce their influence over patients, and the epistemological justification for this is not that this influence would corrupt the primary data, as Grünbaum affirms. The justification is simpler: it would add factors where there already are too many.

A hypothetical analyst who raised data and tested hypotheses this way, by fostering in patients cognitive suggestions, would be comparable to a hypothetical historian who included, in his writing about the events for which some documents he found are evidence, the history of his discovery of the documents, as well as the history of all the events leading to that discovery, from past to present – or comparable to a hypothetical experimenter who studies suggestion and does not control for experimenter expectancy effects so that she is able to include such effects in her database. In the three hypothetical cases, a recognition of the reflexivity between subject and object grants dispensable complexity to an already complex picture. Analysts should simplify their variables to do simpler and better research. Lacan (2007) is right when, in "The direction of the treatment…", he affirms that the more the analyst's being is involved in analysis, the less sure he is of his action (p. 491).

Finally, analysts should estimate, or identify, the cognitive suggestions they could not help but induce because this is what allows them to properly explain such suggestions.

In sum: it is not epistemologically desirable (nor possible, for that matter) to ban cognitive suggestion from clinical psychoanalysis; but cognitive suggestion is only epistemologically desirable if explainable; analysts should be able to "analyse" it in case it transpired or showed within the session and, to do that, it is advisable that they try not to "contaminate" the patient's speech beyond the inevitable.

But how can they try that? And how can they know (1) and (2)? In Grünbaum's argument, the clinical context would not allow us to know (1) and (2) because it would not allow a proper application of the Method of Difference. We would not be able to know whether the patients' behaviours were caused by their unconscious constitution or by clinically induced suggestions because a patient could not be compared with him- or herself. We have seen in Chapter 4, though, that Grünbaum is wrong in his argument that the Millian Method is not applicable in the clinical context: a patient *is* comparable to him- or herself.

6.4 HOW COULD THE ANALYST REDUCE, ESTIMATE AND ANALYSE CLINICAL SUGGESTION?

I shall now formalize a technical device that could help analysts to shrink their role as inducers of cognitive suggestions as well as to know if and how clinically induced suggestions have occurred. Part of such a device is already drawable from many traditions of research and therapy in psychoanalysis, but not all analysts adopt it. What I intend now is to argue it would be reasonable for analysts-researchers to adopt it, at least from an epistemological point-of-view. I shall recommend that analysts establish a "cognitive baseline" in order to overcome what is left of Grünbaum's problem of cognitive suggestion. A "cognitive baseline" would be analogous to the therapeutic baseline of single subject research, and would be devisable through an appreciation of the empirical research on the "external" causes of suggestion and through a discussion on the concepts of "neutrality", "abstinence" and "non-directivity".

The research of Szpitalak and Polczyk (2015) cited above have demonstrated that subjects that were asked to record their greatest achievements and were given positive feedback about their performance (a technique called "reinforced self-affirmation", or "RSA") present less "Yield" and "Shift" responses during their subsequent interrogations in comparison to a control-group. Therefore, they confirmed that, of the three conditions boosting the acceptance by witnesses and suspects under interrogation of incorrect premises and expectations (Gudjonsson & Clark, 1986) – uncertainty as to the correct answers, trust in the interviewer, and expectation that they will know how to answer the questions – the most critical one is the last, that is, expectation to succeed, for if it is lacking interviewees feel secure in answering an honest "I don't know" (Szpitalak & Polczyk, 2015).

If the beginning of a psychoanalytic treatment is a context of uncertainty, and a context of uncertainty favours suggestibility (see pp. 230-231), could analysts, thus, neutralize this condition through a conduct that, like the one involved in the RSA, stimulates self-confidence in the patient? The notion of interrogative suggestibility can easily be applied to clinical psychoanalysis, but the lessons clinical psychoanalysis can draw from the many experiments based on that notion are not entirely clear. Although it was first conceived inside criminology, interrogative suggestibility would in theory be active in every formal questioning about one's past experiences and events, like the questioning that occurs in a clinical context (Gudjonsson, 1987). But what could an analyst do with the results on the causal relation between RSA and interrogative suggestibility, for example? Perhaps start to think that clearly corroborating the patients' reports and associations is necessary, an important part of the technique, after all, if there is no intention of collaborating with their suggestibility? How could a study like this be accommodated in the psychoanalytic theory and technique?

These questions need no answers, though, for already their presumption is part of the solution to our problem: the presumption that *something* can be done with cognitive suggestion in the clinical context. Cognitive suggestion is no elusive and probably all-pervasive process, a sneaky demon, as Grünbaum seems to argue; it and its conditions can be operationalized. Now we have the knowledge (experimental knowledge, as Grünbaum appreciates) that certain conditions can diminish it in an investigative interview similar to the psychoanalytic. The mere possibility that experimental research advise clinical psychoanalysis on the matter of cognitive suggestion is enough to undermine Grünbaum's argument. If his argument is, roughly, "we cannot say for sure that there is suggestion in clinical psychoanalysis, we can only say that it is impossible to know whether there is suggestion or to control for it in that context", mine is the opposite: "we *cannot* say that there is *no* suggestion in clinical psychoanalysis, but only that it is *in principle possible* to know whether there is suggestion or a condition for suggestibility and to control for it in that context". I am, thus, aligned with the arguments given in the BBS debate by Fine and Forbes (1986), Holt (1986), Spence (1986) and Gauld and Shotter (1986) (see Chapter 2). Before them, Glymour (1982) shortly gave this same kind of response to Grünbaum's charge of contamination. He asserts he does not see "that the experimental knowledge we now have about suggestibility requires us to renounce clinical evidence altogether" and that, anyway, we could "seek for finer knowledge of how much error is introduced in clinical proceedings, and in what circumstances" (p. 30).

One could say that, if the RSA counteracted the climate of uncertainty inherent at the beginning of a psychoanalytic treatment, neutralizing thus the climate's favouring of suggestion, then the treatment would cease to be a psychoanalytic one; that the patient's anxious uncertainty is the fuel of the "analytic work", and that premature interventions like reinforced self-affirmations run the risk of reinforcing the patient's repressions. Is there a conflict or inconsistency here? Does RSA prevents suggestion on one side and promotes it on the other?

As Laplanche and Pontalis (1988) bluntly put, "from the dynamic point of view, the treatment relies basically on the existence of suffering brought about by frustration […]" (p. 3). They explain: "the analyst must make sure that the quantities of libido released by the treatment are not immediately redirected toward a fresh cathexis of external objects; they must so far as possible be transferred into the analytic situation", being thus "deprived of any occasion for discharge other than through verbal expression" (p. 3). At the beginning of the treatment, analysts should not promote in patients any satisfaction that is not the satisfaction of speaking about their open, unresolved, conflicts. Which kind of satisfaction would an RSA technique promote, one that shuts down the unconscious or one that opens it up?

Many have claimed it, including Freud: analysts should be *abstinent*; they should be overall *neutral*. But what are properly abstinent and neutral analysts allowed to do, especially at the beginning of a treatment? Laplanche and Pontalis (1988) tell us that neutrality is "one of the defining characteristics of the attitude of the analyst during treatment" (p. 271). By refraining from counselling the patient and from directing the treatment according to some religious, social or ethical ideal, analysts are being neutral in the *moral* sense of the concept. They can be neutral in the *affective* sense – "as regards manifestations of transference" (p. 271) – which is the same as being *abstinent*. According to the rule of abstinence, analysts "should refuse on principle to satisfy the [...] [patients'] demands and to fulfil the roles which the patient[s] […] [tend] to impose upon [...] [them]" (p. 2). Finally, and this is what interests us the most now, analysts are neutral in the *epistemic* sense when they do not, "*a priori*, lend a special ear to particular parts of […] [the patient's] discourse, or read particular meanings into it, according to [...] [their] theoretical preconceptions" (p. 271).

Let us remind what we have concluded about analysts who lend a special ear or eye to certain areas of the patients' behaviour *a priori*. We have seen in Chapter 4 (see pp. 132-133) that the analysts' attention is not *exactly* free-floating (setting aside the matter of whether a perfect free-floating attention is somehow achievable), for, faced with a host of mental products, they positively discriminate the ones that exhibit, according to the core hypotheses of psychoanalytic theory, the most promising general themes; for example, a comment on family conflicts would draw more of their attention than a comment on the room's pretty couch. Besides, analysts do read particular meanings into the patient's discourse according to core hypotheses, although in a sense they do it *a posteriori*. We have seen how psychoanalytic theory, in an interaction with the total evidence raised by the patient, guides the generation of the competing "meanings" (mental schemas) among which the most explanatory will be the championed one. I have argued that such uses of core hypotheses are harmless "biases" (they are actually heuristics), for their loose constraints do not prevent the evidence from teaching and surprising analysts or from showing them fallible in their presumptions of what is relevant or probable for the patients' unconscious motivations – even if, I would add, this heuristic come to influence the patient's productions. Last but not least: in our model, anyway, core hypotheses are supposed to have a robust independent support. On the other hand, a use of categoric hypotheses to guide clinical inferences represents indeed a dangerous bias (see pp. 118-119), which could influence the patients and complexify the evidence to an excessive degree ("contaminate" the evidence). We must, therefore, qualify Laplanche and Pontalis' definition of (epistemic) neutrality: analysts may loosely select data according to the core of the theory and still be neutral, but they indeed cease to be so if they "read particular meanings into it" before they have contact with an expressive amount of evidence.

We could distinguish four attitudes through which the analyst can be cognitively directive, hence potentially suggestive, in clinical psychoanalysis: the attitude of *raising themes from scratch*, that is, of leading the patient toward ideas and/or actions he/she has never considered before[86]; the attitude of *establishing links* between ideas and/or actions the patient has presented in session and recognized to be ever-present in his/her life; the attitude of *manifesting that certain ideas and/or actions of the patient have some relevance or significance*, that is, of reinforcing certain *categories of ideas*; the attitude of reinforcing in the patient some kinds of *mental functioning*, and consequently of causing him/her to inhibit other kinds of mental functioning, an attitude of which the RSA technique and a statement of the "fundamental rule" are instances. Here we have an ordinal scale of the analyst's directive attitudes, from the most directive to the least.

In my understanding, only the last two, the ones boosting certain categories of ideas and kinds of mental functioning, should be deemed *neutral attitudes*, because the risk they pose of contaminating the patients' productions is little. If they come to induce suggestions, the latter will be almost pure autosuggestions – a kind, we have seen, very interesting for analysts.

Now, as the two directive-but-neutral attitudes could include the RSA technique, the use of the latter would not contradict the rules of neutrality and abstinence, and the issue we have

[86]At least as far as the analyst knows.

raised is a false one. RSA counteracts interrogative suggestibility, experimental research suggests, and it would not be a hindrance to the psychoanalytic work were it applied in psychoanalytic interviews. Inviting the patients to trust their own memories and to feel some pleasure in talking about them is not something repressive. It represents well, after all, the fact that extraclinical studies have the potential to solve the problem of clinically induced suggestion – and it makes us wonder about the numerous studies that could be planned and undertaken with this purpose in mind.

Actually, there have already been studies indicating the various factors potentially at play in clinical psychoanalysis that reduce suggestibility – factors, thus, that analysts could strengthen further in their practice. Lacewing (2013b) cites some of these studies. He cites Gibbons and McCoy (1991), whose results "indicate that being directed to attend to one's mental states […] decreases susceptibility to suggestion" (Lacewing, 2013b, p. 736); here we would have a reinforcement of a primary kind of mental functioning, cited above among the analyst's less directive attitudes. He also refers to Gudjonsson (2003), who demonstrated "a highly significant relationship between suggestibility and coping strategies" (Gudjonsson, 2003, p. 395). According to Gudjonsson (2003), coping strategies that decrease suggestibility

> involved a critical analysis of the situation and a facilitative problem-solving action. Common self-statements of this group were the following: cognitive, 'I can't be expected to know all the answers', 'Some of the questions were not in the story', 'I am sure I have done as well as anyone else'; behavioural, 'I tried to stick to what I remembered', 'I looked critically at each of my answers', 'I tried to look at the situation objectively' (p. 396).

Lacewing (2013b) seems to suggest, therefore, that such self-statements are already potentially at play in psychoanalytic interviews, but they could, of course, be fostered.

Laplanche and Pontalis (1988) claims that neutrality, anyway, is "an ideal to be aimed at rather than an absolute injunction" (p. 272). Neutral would be the quality, not of actual analysts, but of their *function*, and this would mean that they should not make themselves "felt" in their own "psycho-social specificity" (Laplanche & Pontalis, 1988, p. 272). To give a more epistemic sense to this function, ideal or attitude, though: neutrality could be understood as that which is operative when analysts manifest to patients that they can endure the incompleteness of the data and the underdetermination of any hypothesis one of the parts of the therapeutic couple may come up with to explain the data. A neutral analyst leaves judgment in suspension, strives so that his or her mind does not lock his judgments about the patient within a rigid framework, a framework dictated by his or her "psycho-social specificity". This requires both intellectual and emotional work: theory, logic and creativity on one side, in order to remain attentive to relevant data and open to surprising data, as well as to make the pool of candidate explanations as large as possible, and an ingrained habit of postponing intellectual satisfaction, of keeping the truth-seeking angst at an optimal level, just like the good scientists and philosophers do. If this work is honestly done, the patients' clinical heterosuggestions will not be numerous. Wallace (1986) contends that,

> although the analyst's eccentricities are not totally banished through analysis, they are considerably muted; he gains a heightened awareness of his personality traits and manner of impact on others. This, coupled with mastery of technique, minimizes the extent to which the analyst's behaviour skews or distorts the manifestations of the unfolding transference and maximizes the degrees to which he can treat the patient's behaviour as intrapsychically determined (p. 63).

At some point, though, analysts must bring their judgements to the earth and communicate interpretations to the patients, and this should involve the two more directive attitudes cited above: the attitude of leading patients toward brand new ideas and actions, and the attitude of establishing links between their perennial ideas and actions. This is because the primary aim of clinical psychoanalysis is not to be an arena for psychological research, but to alleviate the patients' suffering. Laplanche and Pontalis (1988) indeed tell us that even orthodox psychoanalysts "waive the rule of complete neutrality on the grounds of its being neither desirable nor practicable" (p. 272) in certain cases – "cases involving anxiety in children, the psychoses and certain perversions" (p. 272), for example. But mild neurotics should also have their analysts waiving the neutrality rule, simply because it is impossible to attain insights and working-throughs from absolute neutrality:

> Although the analyst's neutrality and (relative) verbal passivity minimize his influence on the patient's productions and hence give the analytic procedure some crude approximation to a controlled experiment, the analyst cannot simply sit there silently, month after month, and expect the patient, through sheer free association, to arrive at a coherent and accurate formulation of his case. […] it is precisely some of his most significant and painful issues about which the patient will *not* want to talk and around which he will have erected all manner of defenses; the analysts must sensitively, though consistently, direct the analysand's attention to his defensive maneuvers, the anxieties motivating them, and the affect-laden memories and fantasies defended against. In short, it would be absurd to argue that the only way the analyst could obviate the charge of suggestion is to never interpret (Wallace, 1986, p. 101).

Could there be a compromise between this therapeutical injunction and the injunction posed at the outset of this section – that analysts should reduce and assess the cognitive suggestions they induce?

Wallace (1986) argues that "at issue is *the point at which interpretation occurs*" (p. 101). Well, *until communicated interpretations are inevitable*, analysts may safeguard the epistemic dimension of the clinical encounter by having a neutral attitude. If they can sustain this attitude for a significant time at the beginning of the treatment, they will have at their disposal a *cognitive baseline* to compare with the second, more directive moment of the treatment. They can, then, come to know whether they have induced cognitive suggestions and how.

This cognitive baseline is analogous to the therapeutic baseline advocated in N=1 research on psychotherapeutic interventions. To study whether a therapeutic intervention is really effective, researchers may regularly measure one subject's target-behaviours (the behaviours that should be affected) for a significant segment of time where the intervention under study is not yet delivered. The values of this first moment are called the subject's "baseline". It allows researchers to use the subject as his or her own control: if, in the segment of time after the intervention is delivered, the values of the target-behaviour are considerably altered *in comparison to the baseline*, according to the aim of the intervention, it can be said that the intervention was relevant for the subject's improvement. If such a control is not available, measures of improvement are compatible with the null hypothesis that such an improvement has other causes and would have been obtained regardless of the intervention.

An analogous procedure could be undertaken to estimate and analyse the suggestion that transpires in the clinical context. A strictly "neutral" phase could be sustained by analyst-researchers at a prolonged opening of their treatments, so that the patients' behaviours meticulously registered in this phase can serve as a control for the moment following the deliverance of more directive interventions. It would be a control because the chance that it contain suggestions fostered the analyst-researchers would be very low, allowing them to examine whether and how the suggestions of the second, post-interpretation phase were fostered inside or outside the clinical context.

Let us suppose a team of analyst-researchers have recorded and transcribed the sessions of a patient and wish to entertain the hypothesis that his last post-interpretation reported insight has all over it the fingerprints of the analyst who conducts his treatment. They check if in the transcripts of the baseline sessions there exists some evidence that is perfectly consistent with the insight. If it exists, it is reasonable to conclude that the analyst was not a corruptor, but just a facilitator of the patient's mental products. If otherwise they note a significant qualitative difference between the baseline evidence and the non-baseline insight, they may examine whether some elements of the insight are similar to highly directive elements of the analyst's preceding interpretations. If this is the case, the analyst in question may be asked to stop manifesting such suggestive elements. Then, if the patient ceases, with time, to manifest any behaviour with such elements, it is reasonable to conclude that that insight was indeed a product of a clinically induced cognitive suggestion and that it brought superfluous (but still analysable) complexity to the investigation of that patient's particular unconscious motivations. If, on the contrary, the patient keeps manifesting behaviours with such elements, it is reasonable to suppose that – given all that we have presented about suggestion – even if such behaviours could not have come to existence without the "hand" of the analyst – and this counterfactual would actually advocate in favour of her competence – the team would have valuable evidence at disposal, for such behaviours would be largely an effect of the patient's psychic structure. If the analyst indeed had brought up something that is slightly discrepant with the baseline evidence, and if this was digested by the patient, we may remind of Baudouin's metaphor: the seed found a good soil.

In any of the possible outcomes, anyway, a cognitive baseline would allow the team to know what is seed and what is soil. In any of the cases where the baseline is different from the non-baseline as regards an interesting behaviour, it is possible to distinguish what of this behaviour could be a consequence of the patients' "raw", pre-intervention personalities, from what of it has been largely provoked by the analysts' deliberate or undeliberate whims. And what analysts have largely provoked in them can also be evidence (although a superfluously complex one) if analysed side-by-side with the baseline. The patients' fantasies about authority figures could be in question then.

Much is left operationally undefined here, for example how we could know when baseline evidence is consistent with non-baseline evidence. Like the one in Chapter 4, the procedure here is reasonable but vague, and its application by different researchers could ensue different conclusions about a same material. But it is a start, and if its details were properly defined we could have a properly Millian solution to Grünbaum's charge that it is impossible to know whether the data from clinical psychoanalysis is corrupted to the point that it loses its value as probative evidence. With this procedure and our previous deflation of suggestion *qua* epistemic intruder, the picture painted by Grünbaum becomes much less intimidating.

To conclude, one could say that this procedure is no sheer novelty for analysts that are not formal researchers and whose rigorous investigation would only serve the mental health of their patients. Wallace (1986) asserts that Freud was indeed liable to the charge of suggestion with some of his premature interpretations and reconstructions, but that, ordinarily, analysts would not

> dispense with an inductive testing (i.e., history gathering, as well as the observation of transferential themes, patterns of associating, and so forth) of [...] reconstructions before feeling confident enough to communicate them to the patient who, in any event, would be resistant to hearing them early in therapy (pp. 37-38).

Complete interpretations[87], he says, would be "delivered only after there is ample evidence for them, only after one has gleaned considerable history and encountered the same themes, patterns, and sequences many times"; they would not be "hurled at the patient like bolts from the blue" (p. 101).

6.5 CONCLUSION: A RESPONSE TO GRÜNBAUM

In this chapter, we have seen a long conceptual and theoretical discussion challenging Grünbaum's argument that psychoanalysts have no means to ensure that their clinical data are not vitiated by suggestion and thus that the hypotheses generated and tested through this data will always be deceptive. The strategy was first to understand what suggestion is, how it operates and how corruptive it can be, and then to verify whether there is indeed no way to observe it and reduce it in the clinical context; in other words, to check for the real nature and dimension of the problem and only then deal with it. Throughout our conceptual and theoretical discussion, we have seen a deflation of the corruptive potential of suggestion – but we have not seen its disappearance. Cognitive suggestion remained a problem, but a small and manageable one.

At the beginning, I argued that in Grünbaum's critique there is no effort to define suggestion or to cite scientific studies on suggestion phenomena. I concluded that his conceptual and theoretical elucidation of suggestion is poor and that, as most of his critique depends on the infamous term, it would be worthwhile to delve into its conceptual and theoretical dimensions. We then went through a refreshing history of the concept, through the complicated process of its definition and through some surprising facts about its causes and effects. It was defined as a partially unconscious process in which a subject adjusts his/her mental state to the wish of an authority figure as the subject mentally represents it. Evidence for the intra- and intersubjective causes of the process was presented, and thereby what appeared to be a hardly traceable and extremely corruptive process became the opposite – a controllable and epistemically revealing process. For example, experimental research on interrogative suggestibility and on the influence of the subject's expectancies in hypnotic responses show respectively that it is possible and feasible to reduce suggestibility in an unbalanced relationship and that suggestion response reflects the suggestant's mental structure more than the command given by the suggestor.

[87]Interpretations "speaking to the historically determined motives as they manifest themselves in the past, present and transference" (Wallace, 1986, p. 101).

With this in hand, I first argued that suggestion is not an intruder in clinical psychoanalysis, in fact that the domain of clinical psychoanalysis could be merged with the domain of suggestion phenomena – if it is assumed that the primary fantasies and the complexes clinical psychoanalysis talks about are in their turn invariably about attempts to please a caregiving figure in order to conquer or keep its love. It follows, first of all, that Grünbaum's argument is much less worrisome than it looks.

But we have seen that, if induced suggestion – the classical kind, the only kind that could be considered a "contamination" – is not an intruder in clinical psychoanalysis, it is at least a whimsical guest. We have seen that analysts should be able to know whether it has occurred and whether it was fostered inside or outside the clinical context in order to embrace it as evidence; also, that analysts should strive to reduce it in order to prevent their evidence from becoming unduly complex. At last, I indicated that the adoption of a "cognitive baseline" can help analysts to track induced suggestion and that the experimental studies on the "external" causes of induced suggestion can help analysts to reduce it. In sum, we can be aided by the Method of Difference and by experimental studies to manage what is left of Grünbaum's contamination problem. Clinical data are not irredeemably deceptive, after all.

In the next chapter I shall argue, however, that some "contamination" of unconscious motives is necessary for mental healing.

7 THE PROBLEM OF THE LINK BETWEEN TRUTHFULNESS AND THERAPEUTIC EFFECT

7.1 INTRODUCTION: A SECONDARY BUT STILL IMPORTANT KIND OF EVIDENCE

An analyst enunciates to a patient what she has privately inferred regarding the unconscious determinants of his mental products; she *interprets*, in classical jargon. He seems to pay attention. Perhaps, rather, the patient himself enunciates to the analyst something he has learned in the last week about the clandestine motives of his erratic and painful behaviour. What can she expect then to happen to his mental health in the short and in the long run because of such overt speech?

Now backwards. Following many sessions and (what seemed to the analyst) striking exchanges, the patient reports that his panic attack did not show up in contexts it usually would have shown up; or that he started to have the courage to make love, to party and to be productive; or that he has been sensing in the last days that his mind is clearer and his conduct more confident. What can the analyst infer from this? Does this mean that what has been said in the last sessions was epistemically valuable? What would the evidential value of such therapeutic effects be? And, were things not that positive, would nontherapeutic or antitherapeutic effects be indexes of a lack of truthfulness in the past analytic encounters?

The problem of the link between manifestations of truthfulness and a subsequent therapeutic effect in the clinical context is the last of the problems abducted from Grünbaum's criticism of psychoanalysis that we shall try to clear up. It is derived from his discussions on the Tally Argument.

Independently of Grünbaum's debunking of the Tally Argument, if we had followed him in considering it the most fundamental argument in the epistemology of psychoanalysis only because it was presumably so for Freud, we would be in trouble. For a fallout of that tenet is the idea that the main way analysts could obtain reasons to infer that the content of an interpretation represents part of the patient's psyche – that is, to infer they have diverted from noxious contaminations and defective reasonings – is by enunciating random interpretations and waiting until one of these causes a therapeutic effect. We would have to conclude that the test of hypotheses in clinical psychoanalysis could be or become nothing but a bizarre game of trial-and-error.

Another fallout of the tenet that the Tally is a most fundamental Argument is the impossibility of testing the NCT in a clinical setting. In order to test if the enunciation of truthful stuff is indispensable for the healing of patients, it is indispensable that we are able to check whether a true utterance is really the case – if all therapeutic effects were preceded by true utterances, and all non-true utterances were followed by non-therapeutic effects, the NCT would be thereby supported. But if, as clinical researchers, the Tally were our most cherished principle, how could we check truthfulness without becoming circular? For, then, we would not be able to test the NCT with another, independent, method, but only with the NCT itself[88]!

Of course, a reasonable psychoanalysis could never work and has never worked by praising the Tally Argument as the One and Only Epistemic Saviour! As we have seen in the last two chapters, its clinical method has actual or potential devices to overcome a host of inferential fallacies and the threat of evidence contamination quite independently of any therapeutic consideration.

Nevertheless, there are some reasons to reflect if and in what way therapeutic effects are indexes of truth. Let us remind ourselves that it is only because the "objects" of a research are part of a context of illness and care that it must employ something called "the clinical method". In such a context, to alleviate distress is of course a primal ethical concern. Therefore, clinical researchers have assuredly more chances of acting ethically in pursuing their questions and answers if they are able to harmonize both their therapeutic and epistemic aims as much as it is possible. It remains to be learned how much it is possible.

A second reason to think about the "alethiometric"[89] nature of therapeutic effects is so to speak historical. As Grünbaum (1984) and others (e.g., Rubovits-Seitz, 1998) have taught us, one of the criteria to accept or reject an interpretation in clinical psychoanalysis has historically been the behavioural reactions to it, which includes symptom remissions. This has been so probably because of the harmonizing urge just mentioned; even though psychoanalysis has been taken to other realms, it has been forged under the pressure of a medical look. It is the epistemologist's task to figure out whether analysts have been correct in using such a touchstone.

We should risk presenting still a third reason: that it looks nice to our 21st Century Western moral self that truth and happiness can be compatibilized; for most of us tend to reject both happiness founded upon lies and paralysing suffering as a result of having faced things as they are.

---

[88]Of course, Grünbaum was aware of this impossibility. That is why he invoked comparative quasi-experimental researches to demonstrate the problematic character of the NCT: he knew he had to attack it from the outside if he wanted to preserve his whole argumentative strategy (see pp. 12-13).

[89]This neologism was inspired by Philip Pullman's fantasy books (Pullman, 2003).

Despite the fact that the evidential worth of therapeutic effects is secondary in clinical psychoanalysis, thus, we have good reasons to reflect upon if and how an enunciated truth is related to a therapeutic effect in the patient. We shall do it in what follows, and we shall conclude that the point that forces us to admit that positive therapeutic outcomes constitute a complex and misleading evidence is exactly the same point that shows us that boosting truthfulness increases the chance of success in any psychotherapy.

7.2 FACT AND VALUE: DISTINCT AND ENTANGLED

Let us imagine an analyst saying to her patient: “You are trying to avoid the painful idea that you and your wife have become no more than good friends”. How would this inferred state-of-affairs sound to the patient? Would it sound to him simply as an interesting fact, a riddle to be solved? Or would it rather sound like “You should feel free to speculate about this idea, because human bonds are indeed always changing”, or maybe “Do not avoid this idea, your sexual life should never be sacrificed this much”? Could it also sound like “It is cowardice not to admit this to yourself and to your wife” or even like “These dangerous ideas will not save your marriage”?

Before we deal with the problem of the link between what is true and what is good, we must deal with the broader problem of the link between fact (whether true, false, probable, improbable, possible, impossible, etc.) and value (whether positive, negative, neutral, etc.). Do analysts imply, indicate or incite affective “shoulds” or “should-nots”, ideas of wrong and right, of health and illness, of beneficence and maleficence – to sum up, do they imply, indicate or incite *values* – by their *very acts of investigating and describing their “objects”*? Are they theoretically able to separate facts from values in their exchanges with the patients?

By the way, what should be their aim in this respect: to separate or to mix up?

In the last chapter, we have seen that an analyst could control for her contaminating moves in the clinical process by instituting an initial phase in which she must be abstinent and neutral as much as this is possible; she could then compare the patient’s responses from such epistemic baseline to his responses from a late phase, a phase where she is capable of intervening more elaborately; that way she can track and analyse contaminated mental products in the patient. The analyst is *capable* of intervening more elaborately only in a late phase because nothing significant and correct can be inferred from a scant amount of clinical material. But she also *must* do so – after all, the patient is there to have his sufferings alleviated, not to be just inspected and bidden farewell!

We arrive thus at a radical turning point on our epistemological course. The problem is no longer how to know an "object" better, but how to keep knowing "it" well even when "its" essential properties *change*. A negative character of the clinical influence, expressed in the term "contamination", makes sense only in the initial chunk of the clinical process; the analyst must track if and how she influenced the patient but she cannot utterly refrain from attempting to influence him! Thus, to answer the question above: the analyst's aim is both to *separate* and to *mix up* fact, what-is, and value, what-should-be.

But, again, are analysts theoretically able to separate facts from values? Are mere acts of investigating and describing, of stating facts, *by themselves* able to incite values, to indicate what is virtue and what is vice? Two concrete sets of cases, one in natural science and another in art theory, indicate that the puzzle of the fact-value connection admits no extreme guesses.

A picture of an exclusively material universe can appear as morally degrading or absurd to a spiritualist or idealist philosopher. But science promoters with a poetic verve, such as Carl Sagan, have demonstrated that materialism can also inspire sacred and mysterious sentiments and a sense of connection, meaning and moral responsibility. He gives us glimpses of what a post-modern materialist religion would look like. In "Cosmos", for instance, he writes:

> I am a collection of water, calcium and organic molecules called Carl Sagan. You are a collection of almost identical molecules with a different collective label. But is that all? Is there nothing in here but molecules? Some people find this idea somehow demeaning to human dignity. For myself, I find it elevating that our universe permits the evolution of molecular machines as intricate and subtle as we (Sagan, 1980, p. 127).

And, in his "The demon-haunted world":

> Science is not only compatible with spirituality; it is a profound source of spirituality. When we recognize our place in an immensity of light years and in the passage of ages, when we grasp the intricacy, beauty and subtlety of life, then that soaring feeling, that sense of elation and humility combined, is surely spiritual (Sagan, 1997, p. 32).

Here we have a case in the natural sciences in which the intrinsic features of the exact same fact (leaving aside the matter of its truth) appear for some as *good* features and for others as *bad* features. This proves some sort of independence between facts and *specific* values. It leaves us wondering whether enunciated facts can be totally deprived of *some* moral colour, though. Values are crawling underneath the skin of natural-scientific researchers since the outset of their work, already in their selection of objects and problems. An astrobiologist would not look for traces of life in the universe if she really – deep within her limbic system – considered the possibility that alien life would pose a threat to life on our Earth, in the manner of a Lovecraftian "colour out of space" (Lovecraft, 2016); it is reasonable to infer that she has only become an astrobiologist because she hopes that her possible scientific findings will enrich Earthling life, both technologically and "spiritually" (in Sagan's sense, at least)[90].

But what if the moral colour is, as the colour of the Lovecraftian tale, undefinable? When representing experiences of internal and external worlds, the artist always expresses a moral stance toward them; to put it roughly, the concrete shape the artist gives to the experience is always a message of "this is good" or "this is bad" (even though such stance is usually nuanced). But to track what is the case in a specific work of art - to track whether the opus is apologetic or critical regarding its theme - is not easy.

More than ten years ago, there was a funny episode in the Brazilian film market demonstrating the significance of the issue. Telling the story of an officer of Rio's Military Police who tortures and kills people in the favelas, Captain Nascimento, the film "Elite Squad" (Padilha, 2007) was debated and praised both by progressive and fascist minds. The same – let us put it thus, *aggressive* – scenes were taken by some as an acid exposure of the crimes the Military Police in Brazil commits every day, such as torture and racism, and of how monstruous a neoliberal and colonial system can become, and by others as representing the deeds of assertive, masculine heroes, ridding the "good neighbourhood" of the "bad guys". Left-wing cinephiles were secretly baffled at the joy some spectators took from the film, wanting to ask them: "Why did you like this film again?". And themselves: "Am I really a good person for having loved this film?".

A psychoanalyst would say that we could track the moral perspective of artists by grasping their work as a whole and by investigating their personal lives. In the case of H. P. Lovecraft, it is a consensus that his work is racist and we would not even need to go through his life to be sure of it. In other cases, the moral diagnosis may be quite surprising, as (we have seen it in chapter 4) a psychoanalytic inference has the potential to be. Let us consider, for instance, the works of the filmmaker Lars Von Trier. There are in almost every one of his films (if not in all of them) increasingly tormented women who at some point become victims of sadistic men; I have always seen his works as criticizing misogynism, almost as misandrist manifestos. But as I learned one thing or two about his personal life (Macnab, 2011) and watched his last film, "The house that Jack built", I am considering the hypothesis that the sight of suffering women is for him a source of pleasure, not revolt[91].

[90] Unless, of course, she has a sociopathic personality.

[91] I oversimplify the man's unconscious with such a comment, but I intended only to illustrate the complex link between fact and value in art; this was not an attempt to show how inferences in "applied psychoanalysis" work.

Some art theorists would say that the perspective of the artist matters less than the perspective of the art consumer; that what is really relevant in art is how the work is interpreted, how it reverberates and influences people. But then another popular dilemma appears on the horizon. Are art consumers able to separate the personal and ethical sides of the artists from the latter's works? Are we at risk of incorporating the values of the artists we appreciate? Here we can refer to what we have concluded in the last chapter, that interpersonal influence is both limited and enriched by the psychological features of the person that is to be influenced. So, our answers to these questions should be that it all depends on who the art consumer is. In reading Lovecraft, persons who have worked-through their aggressive drives along life are able to have epistemic lessons about the paranoid attitude that sustains racism and xenophobia, and moral lessons of how unfortunate, how hellish, this attitude is. It is not too bold to affirm that his cosmic horror can be an opportunity to observe dangerous racist and xenophobic fantasies, and even to dissolve the ones arising in ourselves and in our fellows.

By showing that people may attribute opposite values to the same fact, that it may be inevitable to attribute some value to a fact, and, moreover, that a value attributed to a fact may be far from transparent, even idiosyncratic, these cases in natural science and in art indicate a *theoretical, but not a practical, divorce* between facts and values. Let us now see if this is corroborated by the philosophical debate on the matter. There are roughly two sides in this debate: a side defending that even the theoretical entanglement of fact and value is inevitable; another defending that no entanglement of fact and value is ever fateful.

Ontologically and conceptually, "what-is" is clearly something other than "what-should-be". The former is actual, the latter is potential; the former is imposed to us subjects, the latter may be reached through our effort. Indeed, the philosopher George Moore calls attention to the fundamental mistake committed when one slips from what-is to what-should-be – the Naturalistic Fallacy. It consists in inferring that "X is good" merely from X's natural features; in other words, in bearing the idea that X's naturalness would entail X's goodness, or that anything that is natural must also be deemed good. It would be a fallacy, according to him, because "purely factual premises about the naturalistic features of things do not entail or even support evaluative conclusions" (Ridge, 2019)[92].

[92]Following other authors, Ridge (2019) objects that Moore's Naturalistic Fallacy has an inadequate name, for it is not a mistake inherent and exclusive of naturalistic stances – a point that Moore admits – nor is it a fallacy in its strict sense. According to Moore, the ones who hold that goodness is the property of having been commanded by a supernatural God would also be committing the Naturalistic Fallacy. Thus, having merged naturalism with reductionism, Moore's actual and forceful argument would be that the attempt to reduce moral properties to anything else will always be unprosperous; that goodness

The sky is blue: does it imply that the blueness of the sky is right, good, pretty? Are the works of art that mirror nature and cherish formal fidelity more graceful and transformative than the works of art that do not? If a neo-Darwinian scientist says that apes are our close relatives and that their social organization is hierarchical, with alpha males and the like, does it mean he is advising us to stop struggling against our deep nature, and to accept an unequal society? Or is he giving us resources to overcome this deep nature – if he is indeed claiming it to be our deep nature, considering that there are matriarchal species of apes, and that we are not 100% apes after all – and to work toward fairer societies?

The Frankfurtian and post-structuralist traditions share almost the same mistake described by Moore. It is assumed in such traditions that the mere enunciation of a "fact" is screaming "this is natural, this is good, do not fight it!", thus essentially boosting a power structure. But, diversely, neo-Frankfurtians and post-structuralists take this naturalness to be just an *illusion* of naturalness. Thus, they hold, we may use the weapon of the Enemy against Him, creating fact-performances to dissolve oppressive power and boost the power of the oppressed; for them, apparently, some people, the oppressed, are ethically authorized to perform the Naturalistic Fallacy, while the oppressive people must be framed for such a fallacious conduct. Again, they think there would be no such thing as naturalness, no such things as facts, linguistic mirrors of imposing states-of-affairs; but their slipping from fact to value is very similar to Moore's description.

Let us go back to our neo-Darwinian scientist. A Foucauldian activist would say that he is distilling chauvinist piggishness by talking about the hierarchical apes and about how similar apes are to humans. Well, he *might* be a propagandist hiding behind the glow of science. But is he *certainly* so?

In fact, the Darwinian theory admits at least two opposite morals. There has been a conservative, Spencerian[93], interpretation of the theory, according to which brutal competition is the condition *sine qua non* for the improvement of our lives, be it in nature or culture; the survival of the fittest and the perishing of the weaker would secure that future generations be

should be seen as *sui generis* and irreducible. But, as the God-example shows, not all reductionists are naturalists, and vice-versa (as some naturalists hold that moral properties are natural but irreducible). Further, Ridge (2019) claims that the arguments Moore considers fallacious are actually benign enthymemes: we should understand "X is pleasant, therefore X is good", for example, as suppressing the evaluative premise "Whatever is pleasant is good". Non-naturalists could accept this premise, by the way, as a mere link between a natural and a non-natural property; what they should reject is only the identity between them. At bottom, Moore would be pointing *this identity* as a problem, which would make of the Fallacy not a fallacy, but just a metaphysical or semantical confusion.

[93] From Herbert Spencer.

better constituted than they are now. Natural selection would entail the goodness of *difference* and *dominance*. On the other hand, there has been a "granola" interpretation of Darwin, according to which the basic tenet of his theory is that all the living and non-living beings are – literally – our kin; that each of us is a member of a great family encompassing the entire cosmos. The lack of a Great Creative Purpose, the lack of a Superior Father or Mother, would entail the goodness of *fraternity* and of *equality within diversity*. There may be, of course, other, more nuanced, moral interpretations of the theory born in Down House – an also an amoral interpretation of it. Anyhow, this contrast should make us grant the hypothetical neo-Darwinian scientist the benefit of the doubt.

Indeed, these *applications* (political, in the case above) of science is not what science is all about. Schurz (2014) teaches us that there is one stage in the scientific endeavour that can and must be value-neutral; it is the stage in which the scientists give us their reasons for ascribing a degree of likeliness to given theses, the eminent *context of justification*. He concedes that science could not exist without epistemic goals and values (as seen in Chapter 3), hence that such an impartial stance is relative to value-assumptions that are external to science; so, *the actual demand is that the context of justification be free from* just *science-external, or non-epistemic, value-assumptions*. However, Schurz (2014) adds, both the stage preceding this context, where the relevant problems are "constituted" – the context of discovery – and the stage following it, where well-tested findings are used for different social purposes – the context of application – are without a doubt directed by non-epistemic values.

Still, this should not weaken the impartial character of the context of justification. In this respect, Schurz (2014) speaks of a "retro-corrective reference of the context of justification [CJ] to the problem selection in the context of discovery [CD]" (p. 43). Even if, in a scientific research, the stage of problem selection (in CD) constricts the forthcoming results, having thus a great effect on CJ, the latter may dictate that the former be revisited, widened and refined. This way, factors at first excluded from the CD may be later found central to the phenomenon under study. Thus, the selection and the parameters of the CD must be criticized and corrected according to the impasses and surprises that emerge in the CJ, even when this competes with the political goals of the research project. He provides a somewhat Freudian example of such process:

> For a long time, classical medicine had a purely biological and physiological orientation; to a certain extent this still applies today. Medicine of this kind already abstracts away from psychological parameters in the context of discovery. In the meantime, however, we know many organic diseases with psychological causes, and the significance of psychosomatic interaction cannot be dismissed out of hand. In order to be able to examine the question of psychosomatic connections, the restriction to biological and physiological factors in medical models must be

> given up, and the area of psychological phenomena included. In other words, the original and externally motivated restriction in (CD) must be dropped because of knowledge discovered in (CJ) (p. 43).

Schurz (2014) is cautious enough to emphasize that the purge of science-external values from the CJ does not always take place, it having then the character of a *reasonable norm*. It is a norm that the scholars in the social and legal sciences who think it impossible to separate facts from values, as a matter of coherence, do not follow. The author we are citing thinks, on the other hand, that "with enough verbal care it is always possible to separate descriptive from normative claims" (p. 44).

He is able to demonstrate this point also in the context of application, by formalizing a kind of inference that is central in such context: the Means-End Inference. With a Means-End inference, the scientist proposes an indispensable or an optimal means to realize, in a specific circumstance, an end that is externally given, for example by politicians. As this end is dictated by a *fundamental* norm, it cannot be scientifically favoured. But, *in the case where this fundamental norm is presumed*, scientists can underpin a *derived* norm – as if claiming "if you want this, you should do that". The inference has the following structure: if a client or community takes an end E to be desirable, then the client or community should put in practice the means M, for the evidence produced and appraised by scientific methods indicates that M is an indispensable or optimal means to attain E.

Schurz (2014) makes an important caveat: working for a value-neutral CJ is not the same as refraining from ethical and political engagement. The reasonable norm of value-neutrality dictates only that "the scientist [separate] […] her scientific knowledge from fundamental value assumptions which she assumes in means-end inferences, and thereby [that she relativize] […] her means-end inferences hypothetically to these value assumptions" (p. 76).

Decades ago, in a book critical of the Frankfurtian account of the Freudian theory, Keat (1981) was defending with the same stamina as Schurz that science, including the part of it that is concerned with human affairs, could be value-freed. Keat (1981) claimed that there is a *logical* independence between the criteria to assess how valid a scientific theory is and the acceptance or rejection of moral or political commitments. To demonstrate this, he gives a formidable example of Schurz's "verbal care" cited above – of how it is possible to replace "any concept that appears to express or presuppose a particular normative attitude […] by one that does not do this" (p. 42). (Although unable to prove that this is always possible, he claims he has produced a satisfactory result in all of his attempts to drain away values from scientific concepts.) Let us have a moment to appreciate his example:

> Suppose that someone claims that there is a relationship between the degree of "industrial unrest" and the "alienating" character of the work involved in a certain kind of society. Some people might object to the phrase "industrial unrest" on the grounds that it indicates a standpoint which values order and discipline, and is thus implicitly critical of the activities to which it refers. To meet this, I would propose replacing "degree of industrial unrest" by something like "the number of strikes and level of absenteeism"; and if the latter part of this new description were found objectionable, for rather similar reasons, it might be replaced by "the number of days un-worked but not due to illness or injury" – and so on. Others might object to the term "alienating" because it expresses a negative judgement based on an unacceptable ideal of human labour. In which case, I would propose an alternative such as "the lack of control exercised by workers over economic production", and so on. By these kinds of reconstruction of the initial statement, I suggest, it is possible to make its scientific assessment independent of potentially challengeable normative claims. And if this is so, the force of this objection is considerably weakened (p. 43).

Like Schurz, Keat (1981) holds that to recognize such logical independence is not tantamount to be impelled to expunge moral and political attitudes off of scientific concepts; for him, there would be no reason why a scientific theory should not bear such attitudes, no reason to establish a value-free lexicon for it. His point is, rather, that there remains in it a parcel we could call the theory's scientific core that may be assessed for its epistemological character alone.

Keat (1981) offers us a couple of other formidable illustrations in favour of his point. For the first illustration, he invites us to consider the scientific claim that, in a particular case, an unjust economic structure caused a certain political change. Now, there is no consensus over the concept of social justice, and each one of the concept's accounts is composed of different principles indicating how social goods such as income, rights, power, etc., should be distributed – that is, of different ethical, value-laden principles. Even so, he argues, if one defines which principles of justice are at stake in a claim like this and determines how much a particular distribution of social goods conforms to these principles, independently of one's penchant for these principles, then its explanatory character can be assessed in a value-free manner.

The second illustration comes from Ernest Nagel (the same Nagel that foreshadowed most of Popper's criticism of Freud). Nagel (1961 as cited in Keat, 1981) refers to the medical concept of anaemia to defend the difference between "characterizing" and "appraising" value-judgements and thereby to defend the chance of a value-free social science. To judge whether an organism is in an anaemic state by counting the number of red cells it has available in a sample of its blood would be "characterizing". Even though various different numeric definitions of an anaemic state may exist, once one of these definitions is picked it is possible to scientifically determine how anaemic an organism is. On the other hand, to judge an anaemic state could also be “appraising” in case it is judged an undesirable or pathological state for having disabled the organism to perform certain crucial functions related to the *norms* of survival, reproductive fitness and so on. The moral here is that a characterizing judgement does not hold on to an appraising one. “Nagel then goes on to argue that the same analysis can be given for many concepts employed in the social sciences: for instance, the judgements that certain attitudes or actions are mercenary, deceitful, cruel and such like” (Keat, 1981, p. 41).

By defending the *logical* independence between facts and values, Keat (1981) is criticizing both positivist and Frankfurtian stances regarding their connection. Unlike positivists, he thinks that science can include values and be an ally to human thriving; he sympathizes with the project of a social theory that is valuative and attempts to overcome social vices. However, unlike a Frankfurtian such as Habermas, he thinks it is possible to assess a scientific claim independently of the values it may bear, and that it is a mistake to adopt distinctive criteria of scientific validity, criteria based on the successful realization of certain values.

If we think of positivists and Frankfurtians as opposing extremists as regards the fact-value issue, Hilary Putnam has, like Keat, a moderate stance; unlike Keat, though, he is more inclined to the pole where fact and value are indistinguishable. Putnam (2002) debunks the positivist idea “of an omnipresent and all-important gulf between value judgments and so-called statements of fact” (p. 8) and holds that, in most human instances, these two aspects are so entangled with one another that the attempts to separate them become onerous, if not altogether futile. His background aim then is to oppose the ones who claim that ethical questions in welfare economics cannot or need not be resolved through reasoned arguments; the link between fact and value, thus, would be “no ivory-tower issue”, for it would relate to “matters of – literally – life and death” (p. 2).

Putnam (2002) argues that, in a positivist framework, the fact-value dichotomy is “omnipresent” and “all-important”. “Omnipresent”, he explains, as, according to the framework, it should cover the whole universe of meaningful judgements regardless of the area they pertain to; if a judgement could not be attributed to any of the sides of the dichotomy, then a positivist would feel impelled to conclude that it was ambiguous – that its utterer had just a faint idea of what she intended to mean – or that it was not a *real* judgment at all. And “all-important”, he explains, because it would enable the prowess to solve all philosophical problems at once (except the technical ones, such as the ones self-imposed by the positivists). According to the author, also the analytic-synthetic (or convention-fact) dichotomy was taken to be thus omnipresent and all-important until, just like the fact-value dichotomy, it was undermined by the exposure of the hybrid character of relevant cases.

In fact, Putnam (2002) teaches us that the pair of dichotomies have one more thing in common: both had their origins in Hume's "famous doctrine that one cannot infer an 'ought' from an 'is'" (Putnam, 2002, p. 14). Putnam (2002) calls it a "doctrine" because he takes it to be no mere claim about the validity of certain forms of inference, but rather a repository of metaphysical assumptions. According to Hume, a fact is no more than an assemblance of *sensorial data*, that is, of the visual, olfactory, tactile, etc., outcomes of our surroundings. This "pictorial semantics" dictates that, if there could be facts about virtues and vices, then virtues and vices would be picturable in the same way a table, for example, is picturable – with colours, textures, sounds, etc.. As this is not the case, ethics must have to do with something non-factual. In Hume's system, therefore, we produce a moral stance with the *sentiments* we attach to the facts (Putnam, 2002).

Later, with Kant, value judgements are seen as imperatives: as non-factual, like in Hume, but, rooted in Reason, also as unrelated to sentiments. Kant's Ethics also depends on strong metaphysical assumptions, and when his pure practical reason and, supported by it, his *a priori* ethics, became no longer credible notions, logical positivists decided to go back to an inflated version of Hume's doctrine (Putnam, 2002).

Under the pen of logical positivists, the fact-value dualism and the analytic-synthetic dualism boosted one another. A value judgement is obviously not a tautology (that is, not analytic) and the positivists' confidence over what exactly a fact was led them to conclude that it was not a fact as well (that is, not synthetic). Their picture of a fact was a "narrowly scientistic" (Putnam, 2002, p. 26) one – an authentic fact should be represented with a language resembling the language of *physics*. Standing thus outside the domain of the analytic-synthetic dualism, ethics would not be rationally discussable. But with science suffering radical transformations in the beginning of the 20th century, the positivists felt the urge to abandon their Humean notion of a fact as just a sensorial "impression"; the "very basis on which they erected the fact-value dichotomy" (Putnam, 2002, p. 21) was, then, demolished.

Putnam (2002) argues that

> The realization that so much of our descriptive language is a living counterexample to *both* (classical empiricist and logical positivist) pictures of the realm of "fact" ought to shake the confidence of anyone who supposes that there is a notion of *fact* that contrasts neatly and absolutely with the notion of "value" supposedly invoked in talk of the nature of all "value judgments" (p. 26).

He sustains, thus, that the dimension of fact is inflexibly interlaced with the dimension of value. Apart from showing the inconsistency and ludicrousness of the positivist program, he mobilizes three points in favour of this conclusion: the existence of epistemic values; the existence of thick ethical concepts; and the non-factorability of thick ethical concepts.

Many who regard science as "objective" and values as "subjective" would still be attempting to ignore that physical science presupposes epistemic values such as coherence, preservation of past doctrine, simplicity, beauty, naturalness, and so on. For Putnam's interlocutors, to discriminate hypotheses through such so-called values would be just one more way to attain objectivity. Although Putnam (2002) accepts that epistemic values are different from ethical values, he insists that both are still authentically and integrally values and, thus, that epistemic values are not mere indexes of truth. He stresses that we have no "way of telling that we have arrived at the truth *apart from* our epistemic values", that is, we cannot "run a test to see how often choosing the more coherent, simpler, and so on, theory turns out to be true *without presupposing these very standards*" (p. 32). It would be only through the very lenses of such values that we see them as enabling us to correctly describe the world.

Further, Putnam (2002) discusses the existence of so-called *thick ethical concepts*, and criticizes the allegation that such concepts could be satisfactorily disentangled. An ethical concept such as "cruel" is called "thick" in the literature for admitting a descriptive use as well as a normative and ethical use:

> If one asks me what sort of person my child's teacher is, and I say "He is very cruel," I have both criticized him as a teacher and criticized him as a man. […] "cruel" can also be used purely descriptively as when a historian writes that a certain monarch was exceptionally cruel, or that the cruelties of the regime provoked a number of rebellions (pp. 34-35).

By condensing and mixing fact and value, concepts such as "cruel", "crime", "generous", "elegant", "skilful", "strong", "gauche", "weak" or "vulgar" would disdain the fact-value dichotomy. "Not even David Hume would be willing to classify […] [them] as concepts to which no 'fact' corresponds" (p. 35).

Opposing arguments like the one we have seen in Keat (1981), Putnam (2002) holds that thick ethical concepts are not factorable, i.e., that it is not feasible to split them into two meaning elements, a descriptive on one side and a prescriptive on the other; he points to the unsuccessful attempts to define the descriptive meaning of a word such as "cruel" without using the word itself or one of its synonyms. In Keat's terms, he disagrees that the factual part and the valuative part of a statement are logically independent.

> For example, it certainly is not the case that the extension of "cruel" (setting the evaluation aside, as it were) is simply "causing deep suffering," nor, as [Richard] Hare himself should have

> noticed, is “causes deep suffering” itself free of evaluative force. “Suffering” does not just mean “pain,” nor does “deep” just mean “a lot of”. Before the introduction of anesthesia at the end of the nineteenth century, any operation caused great pain, but the surgeons were not normally being *cruel*, and behavior that does not cause obvious pain at all may be extremely cruel. Imagine that a person debauches a young person with the deliberate aim of keeping him or her from fulfilling some great talent! (Putnam, 2002, p. 38).

I would add one more premise in favour of an entanglement-argument like Putnam’s. It is the flagrant fact that the *context of a claim’s enunciation* – its source’s background interacting with its receptor’s and with the immediate context common to both – *unfailingly institutes some non-epistemic interest* related to the claim. This means that, even if we are able to isolate the alethic from the moral parcel of a scientific claim at an *ideal level*, as Keat (1981) has suggested, in *concrete life* this process can never be completed. For example, let us suppose I utter a loud “The sky is blue!” to a random person on the street. Perhaps my motive then would be innocent: a love for the Truth or a wish to inform that person the truth about the sky’s colour just in case that person were not aware of it. However, unless there were a public class or something as reasonable as a public class on the go (but then there would be explicit elements in the surroundings indicating so), we must agree that, in this case, a love for the Truth or a wish to inform would be very improbable motives. It would be understandable if that person questioned my mental health or attributed shady intentions to my action, such as harassing persons like him or her or preaching for a pagan cult. This is, of course, an artificial example. However, the fact is that, every time someone tells us something true, even when it is novel and useful for us, it is common wisdom to ask ourselves: why this to me; why this from him/her; why this here; why this now; why this in this manner; etc.. And the answer always has the following form: because this someone is inviting or coercing me to act, feel or think something beyond the content of the true utterance, something close to this someone’s interests. The same “The sky is blue!” gains a different tone if uttered to a toddler, while trekking, or in a philosophical paper, for example. It is not that *sometimes* there is a practical entanglement of fact and value, but *always*, for *facts are always enunciated in a specific context*.

Lacanians are familiarized with this point, and one of them, Slavoj Žižek, has called it the performative dimension of language; one of his examples is more realistic than mine and should help us better understand such dimension:

> Imagine a couple with a tacit agreement that they can engage in discreet extramarital affairs. If, all of a sudden, the husband openly tells his wife about an ongoing affair, she will have cause to panic: ‘If it’s just an affair, then why are you telling me this? It must be something more!’ The act of publicly reporting on something is never neutral; it affects the reported content itself, and although the partners learn nothing new by means of it, it changes everything. There is also a big difference between the partner simply not talking about secret adventures and explicitly stating that s/he will not talk about them (‘You know, I think I have the right not to tell you

> about all my contacts; there is a part of my life which is of no concern to you!'). In the second case, when the silent pact is rendered explicit, this statement itself cannot but deliver an additional aggressive message (Žižek, 2006, p. 18).

Having gone through some representative points of the philosophical debate over the fact-value issue, it is reasonable to keep our previous case-based hypothesis that the two dimensions are isolated at a logical level, but entangled at a concrete level. Thus, the disagreements over the issue come down to the following: *whether a concrete enunciation can have its epistemic parcel isolated, at least at a logical level*.

Considering the two poles of the debate presented above, *this seems possible*, but only *at some degree*. It seems that the epistemic parcel of a concrete enunciation could never be completely freed from values, but *it could be at least freed from non-epistemic and extreme values*. In spite of the existence of thick concepts, of interest-instituting contexts, of the all-pervasiveness of values, it seems reasonable to assume that, when what matters most is truthfulness, it is possible to communicate facts which imply only *epistemic* or *mild non-epistemic* values.

In sum, until the scrutiny of other cases, arguments and concepts tells us otherwise, we should keep three conclusions: that fact and value are theoretically, but not practically, distinguishable; that both the source and the receptor of an utterance are responsible for the specific way its factual and valuative dimensions are practically indistinguishable; finally, that it is possible to isolate a fact from extreme non-epistemic values.

How would some of such conclusions reverberate in clinical psychoanalysis? If we consider clinical psychoanalysis at its concrete level, and if, at a concrete level, a fact always indicates a value, then it is at least *possible* to promote therapeutic effects by telling patients true facts about their unconscious motives: the utterance of such true facts *could* indicate to patients that an attitude x before such motives is good and an attitude y before such motives is bad, thus influencing them to invigorate x rather than y.

However, we have also seen that a specific fact does not indicate a specific value, that is, the values a fact will incite cannot be predetermined. It is the actors and the situation related to a fact's utterance which determine whether it will be instituted as a positive, a negative or a more or less neutral fact. So, interpretive correctness is not enough to incite the values that will lead to whichever outcome an analyst considers therapeutic, and, in principle, interpretive incorrectness could, in certain circumstances, incite such right values. A true interpretation of an unconscious motive can produce unwished responses – an intensification of schizo-paranoid mechanisms, or a *passage à l'acte*, for example. Therefore, how could analysts indicate the right values through their interpretations? More urgently, what should analysts consider therapeutic, that is, what should a good life be for analysts, and for psychotherapists in general?

7.3 THE EFFICACY OF PSYCHOANALYTIC THERAPY: WHAT DOES IT TELL US?

In the last section, the conclusion that truthfulness is not necessarily linked to a therapeutic effect, and vice-versa, was reached along an abstract and wide path. Now a more concrete and specific path will conduct us to a similar place.

Along the last two decades, there has appeared a wealth of research dealing with the issue of the efficacy and effectiveness of psychoanalytic psychotherapy – generally called "psychodynamic psychotherapy", or "PDT" – and demonstrating it does as well as other kinds of psychotherapy. In a recent systematic review of the evidence for the efficacy of PDT, two of the leading researchers on the matter, Falk Leichsenring and Susanne Klein, concluded that evidence from Randomized Control Trials have significantly demonstrated the therapy's efficacy for treating a host of mental disorders, namely, "depressive disorders, prolonged or complicated grief, anxiety disorders, PTSD, eating disorders, somatoform disorders, substance-related disorders, and personality disorders, including both less severe (Cluster C) and severe personality disorders (BPD)" (Leichsenring & Klein, 2020, p. 118). In every study included in the review, either PDT was more effective than placebo therapy, supportive therapy or "treatment as usual", or it was as effective as cognitive-behavioural therapy (CBT) and pharmacotherapy.

There have even been indications that PDT does *better* than other psychotherapies. A few studies have demonstrated that it is in certain senses superior to CBT (Milrod et al, 2007; Clarkin, Levy, Lenzenweger & Kernberg, 2007). Considering that "most of the studies that found no differences in efficacy between PDT and another bona fide treatment were not sufficiently powered [i.e., did not have an adequate sample size]" (Leichsenring & Klein, 2020, p. 119), we may wonder whether more results like this could have emerged. Already ten years before Leichsenring and Klein's review, Jonathan Shedler listed meta-analyses of rigorous evaluations of PDT disclosing that its benefits "not only endure but increase with time", and that "the benefits of other (nonpsychodynamic) empirically supported therapies", on the other hand, "tend to decay over time for the most common disorders" (Shedler, 2010, pp. 101-102). Whether showing that PDT is equal to other approaches or that it is better than other approaches, such results contradict the researchers and therapists who restlessly claim that the efficacy of PDT is not empirically supported (Shedler, 2010; Leichsenring, Klein & Salzer, 2014). "Studies that demonstrate inferiority of PDT to alternatives exist, but are small in number and often questionable in design", which shows that research on therapy outcomes is subject to allegiance effects (Fonagy, 2015).

Scientifically-minded analysts may indiscriminately respond to Freud-bashers with the good news about the outcomes of CBT's oldest relative. Some may even use such results as a powerful premise for the argument that psychoanalytic *theory* is scientifically credible: psychoanalytic general theory would be true because the therapy based on it is efficacious and effective. Therapists and researchers unpersuaded by psychoanalysis, or simply more persuaded by other theories and approaches, may keep the terms of the dispute: they may conclude that this triumphant conclusion about the truthfulness of psychoanalytic anthropology and psychology is too premature *because* the proofs of the efficacy and effectiveness of PDT display problems of replication, method, scope, etc.. But what exactly would the efficacy and effectiveness of psychoanalytic therapy tell us about the truthfulness of its theory?

It is beneficial for the argument of this chapter to reconstitute the demonstration that, insofar as both sides above (which we still encounter with ease in informal academic debates) are taking that technological evidence to confirm or refute the general theory of psychoanalysis, both sides are misguided. Researches detecting the efficacy or inefficacy of PDT, and even the ones detecting its superiority or inferiority, are not able to tell us something clear about the truthfulness of that theory. Actually, every research on the efficacy of a certain psychotherapy is utterly unable to confirm or refute some general psychological theory.

To become convinced by this, we should embark on a *reductio ad absurdum* – we should discuss the conditions through which the efficacy of psychoanalytic therapy could indeed serve as evidence for psychoanalytic general theory and realize that, in the end, we arrive at an inconsistent picture.

As we have seen in Chapter 3 (see pp. 79-81), the psychoanalytic theory can be broken up into a general theory, a theory of illness and a theory of therapy. The resultant General Theory concerns neuropsychological, psychological, and anthropological facts about how human minds in general work and develop, about what is their ultimate nature, etc.. The Psychoanalytic Theory of Illness, in its turn, is at its bottom the General Theory plus some propositions expressing what is undesirable or wrong in human minds, namely, expressing what we should deem psychopathological and/or vicious – this time we have, besides facts, also *values*. The two theories are composed of both core and categoric hypotheses. The Theory of Therapy is logically derived from the Theory of Illness: if we know what causes the illness, we contradict or suppress it and we get a cure. Psychoanalysis *qua* therapeutic technique is just the operational translation of this therapeutical theory.

Thus, in assessing researches that compare the efficacy of different psychological interventions, we have to be mindful of: the concepts of *pathos* and *vice* implicit in the Illness Theory in question, hence in the corresponding Therapeutical Theory, that is, the *values* established by them; how was the Therapeutical Theory translated into a technique; and the degree to which this technique can be experimentally controlled, that is, the degree to which the therapist's conducts can be standardized. The whole path from evidence to proof is far from simple. Anyhow, according to this understanding, an efficacious therapy could well indicate a truthful Illness Theory and, by putting the latter's values aside, also a truthful General Theory – were it not some furtive complexities of the theory-shifting operations.

The most obvious of such complexities is related to the Grünbaumian concept of placebo (see p. 196-197). Even if a Theory of Therapy is well translated into a technique and this technique is more efficacious than others, it would still be possible that what this Theory avows to be the curative factor of the Therapy (the "characteristic factor", in Grünbaumian jargon) is not *actually* its curative factor – it would still be possible that another element intrinsically or contingently connected to the official factor is the real and single promoter of improvement. Such an unfortunate possibility is attributable to experimental vindications of any line of psychotherapy.

Moreover, in the cases where the "characteristic factor" of a psychotherapy is too narrow, we can say that it is *impossible* to completely isolate it for research purposes. For example, could cognitive-behavioural therapists artificially eliminate the possibility that what mattered in their efficacious interventions on the patient's thinking habits, conditioned stimuli and reward systems was not the "intervention itself" but the "confidence" this kind of intervention transmits, just because the therapists who make them and the patients who hear them "feel" them as scientific and pragmatic – therefore efficacious – enough? If they attempted to eliminate the possibility – for example, by manipulating persons' environments without communicating it to them with nice words, and hoping that results remained positive – the intervention would look nothing like a real cognitive-behavioural intervention (in this case, boosting the study's internal validity would completely ruin its external validity). Along the same lines, no Randomized Control Trial could ever isolate a psychoanalytic interpretation from the persons who enunciate it, or from the social impact of psychoanalytic discourse.

In the insight-based psychotherapies, such as the psychoanalytic, the operations from general to illness theory and from illness to therapeutical theory present an extra intricacy.

According to the psychoanalytic Illness Theory, a large part of what makes people suffer is the lack of certain knowledge, the knowledge about their own unconscious motivations; more accurately, people would suffer from "resistance", the failure to tolerate some facts about themselves. In the Therapeutical Theory, the analyst's interpretations and the patient's insights would serve to counteract "resistance". Well, such knowledge, such facts, found in interpretations and insights, are *literally part of the psychoanalytic General Theory*. So, if the possibility of placebo jeopardizes the truthfulness of any psychotherapy's general theory, it jeopardizes the psychoanalytic General Theory even more. For, even if insight-based psychotherapies are proved efficacious, the possibility of placebo puts in question the healing power not only of *insights in themselves* but also of the *content of specific insights*.

Having gone through some important caveats, let us now see under which conditions therapeutic efficacy could mean theoretical truthfulness in psychoanalysis:

1) IF psychoanalytic therapies are demonstrably, not just efficacious, but also more efficacious than other psychotherapies.
STILL, the advantage could be indicating only that psychoanalytic therapies are better manipulating some factor that bears no relation to insights – a placebo-factor, according to Grünbaum's understanding.
THUS, one would have to compare the outcomes of psychoanalytic therapies promoting insights to the outcomes of psychoanalytic therapies (artificially) eschewing interpretations (merely supportive psychoanalytic therapies).

But let us depart from the start. As Grünbaum has taught us, the empirical proof that psychotherapeutic techniques are uniformly efficacious refutes the NCT and is consistent with the hypothesis that psychoanalytic therapies are nothing but placebos (see p. 12). In truth, however, this proof had a much larger impact: it threatened the very idea and project of a scientific psychotherapy!

Saul Rosenzweig[94] is the most famous among the firsts to touch upon such uniform efficaciousness. Back when it still lacked any systematic proof, in 1936, the psychologist

[94]The same Saul Rosenzweig who sent Freud reprints of his experimental study on repression (Rosenzweig & Mason, 1934) and to whom Freud responded with a "masterpiece of succinctness" (Rosenzweig, 1992, p. 171):

> I have examined your experimental studies for the verification of psychoanalytic propositions with interest. I cannot put much value on such confirmation because the abundance of reliable

published a paper entitled "Some implicit common factors in diverse methods of psychotherapy", which is opened thus: "It has often been remarked upon that no form of psychotherapy is without cures to its credit. Proponents of psychoanalysis, treatment by persuasion, Christian Science and any number of other psychotherapeutic ideologies can point to notable successes" (Rosenzweig, 1936, p. 412). He observed that, from such successes, allegiants always concluded, even if implicitly, that their "ideology" is true and all others false, while "more detached observers" deduced that, "if such theoretically conflicting procedures […] can lead to success, often even in similar cases, then therapeutic result is not a reliable guide to the validity of theory" (Rosenzweig, 1936, p. 412).

This evenness of outcomes between psychotherapeutic approaches was later called "The Dodo Bird Verdict" by Luborsky, Singer and Luborsky (1975) in reference to the "verdict" uttered by the Dodo Bird in "Alice's Adventures in Wonderland" – "*Everybody* has won and *all* must have prizes" (Carrol, 2014, p. 30) – and cited by Rosenzweig as an epigraph to his 1936 paper.

There have been three ample explanations for the Dodo Bird discomfort. First, the explanation that the similar improvement in all psychotherapies is due to different factors. Second, that researchers have failed to measure the outcomes of the psychotherapies in a refined or accurate manner, thus missing important differences between these outcomes. And third, that the similar improvement in all psychotherapies is due to factors that are common to all of them; as each approach is substantiated by a distinct theory, it would follow that the proponents of most approaches are missing or ignoring these common factors (Grünbaum, 1980; Lambert, 2013; Lacewing, 2014b). The common-factors explanation was already explicit in Rosenzweig (1936). He writes:

> Not only is it sound to believe that the same conclusion cannot follow from opposite premises but when such a contradiction appears, as seems to be true in the present instance, it is justifiable to wonder (I) whether the factors alleged to be operating in a given therapy are identical with the factors that actually are operating, and (2) whether the factors that actually are operating in several different therapies may not have much more in common than have the factors alleged to be operating (Rosenzweig, 1936, p. 412).

Although the animal Dodo Bird is extinct, the Verdict with its name and the common-factors explanation of this Verdict are alive and well. We have seen studies indicating the superiority of PDT but, due to results contradicting this superiority (there are also studies indicating the superiority of CBT, for example) and to reservations about the methods employed

observations on which these propositions rest makes them independent of experimental verification. Still, it can do no harm (Freud, 1934, as cited in Rosenzweig, 1992, pp. 171-173).

in this kind of comparison, we are still unallowed to leave the Verdict behind (Shedler, 2010; Lundh, 2014; Leichsenring & Klein, 2020). In a significantly recent paper, Lundh (2014) holds that, "although a large number of therapies have been shown to be more effective than control groups, […] there is still no strong evidence than any of these therapies is more effective than any other" (p. 131). We have also seen an evidence-based defence that the outcomes of PDT last and increase more than others, so clearly the unrefined-outcome-measure explanation is currently under examination. However, the common-factors explanation remains strong.

And yet it has become less daunting over the years. Lundh (2014) proposes two different models through which we can understand the common-factors explanation – the Relational-Procedural-Persuasion Model (RPP) and the Methodological Principles and Skills Model (MPS) – and argues that only the RPP model presents pessimistic implications for the idea and the project of a scientific psychotherapy.

According to the RPP model, a myriad of psychotherapeutic approaches works equally well because they are all able to produce "(1) a good therapeutic relationship, which engages the client in (2) a certain therapeutic procedure and helps persuade the client of (3) a new explanation that provides new perspectives and new meanings in life" (Lundh, 2014, p. 139). Here, the explanation for the Verdict is that all approaches manage to motivate the clients to experience a procedure and to make them believe in a meaning system, and that this whole process is in itself curative; the specificity of the procedure is not assumed to have any potency and the meaning system may or may not have a scientific background. This model would have many controversial implications (Lundh, 2014).

> For example, because the effects of psychotherapy are attributed to non-methodological factors that are shared with a number of other procedures (alternative treatments, "occult" therapies, and religious conversions), the effects of psychotherapy cannot be said to be specific to psychotherapy as such, and psychotherapy cannot be expected to be more effective than "alternative" forms of treatment, provided that they contain the same common RPP factors (i.e., relation, procedure, and persuasion). This means that, if the RPP model is a sufficient model of psychotherapy, we would have to rethink the future of psychotherapy, and whether the present system for training psychotherapists is necessary (Lundh, 2014, p. 147).

The MPS model, on the other hand, establishes that the factors that are common to many psychotherapies and that are responsible for their similar success are indeed described in their theories, but of course under different names – and that some of them may bring more success than others. Lundh (2014) suggests that, if this model is arguably the best explanation of the Verdict, then researchers would have to try to identify such factors and their specific manifestations, as well as to compare the effectiveness of different combinations of them in different contexts. He gives three examples of widely shared principles that could confirm the

MPS model: the principle of facilitating new interpersonal learning experiences in the client (which would have its origin in psychoanalytic therapy); the principle of gradual exposure; and the principle of empathic listening and validation.

Having defined a placebo-therapy as one whose effects are due only to factors unrecognized by the theory that fundaments it, Grünbaum (1980) argues that the Dodo Bird Verdict and its common-factors explanation indicate that the psychoanalytic therapy is probably a placebo. The picture becomes more complicated with Lundh's argument that a powerful factor that is common to many psychotherapies could be recognized in their respective theories as curative, even if under distinct designations. But, if Grünbaum can sustain that for the Psychoanalytic Theory of Therapy the *only* factor that can alleviate a neurotic patient is a correct insight, his argument is nonetheless saved; for correct insights could never be a factor common to many psychotherapies. Soon we shall question the ideas that for that Theory a correct insight is the *only* curative factor, and that correct insights could never be common to many psychotherapies, but for now let us see where they lead us to.

Back to our IF-STILL-THUS argument, in affirming that, in case psychoanalytic therapies are demonstrably superior to other psychotherapies, for some reason they could be just manipulating some placebo-factor better than other therapies, we could take this placebo-factor to be anything that bears no relation to insights – be it something along the lines of the RPP model or something along the lines of the MPS model. In order to eliminate this possibility, thus, we would have to compare the outcomes of interpretive psychoanalytic therapies to the outcomes of merely supportive psychoanalytic therapies and see that the former are more curative than the latter.

In fact, there have been studies aiming at such a comparison. Lacewing (2014b) tells us that the most famous among the long-term studies comparing supportive and interpretive PDT was perhaps the Menninger Clinic's Psychoanalysis Research Project (PRP) and that, "as reported in Wallerstein (1986), […] [it] found no significant differences between the therapeutic outcomes of the two approaches – so insight apparently made no difference" (p. 163). He remarks, though, that

> this conclusion proved too simple, as more detailed analyses of the results showed. First, patients in this study with higher QOR [quality of object relations] did better with expressive approaches (Bateman & Fonagy, 2004). Second, complicating the picture further, Blatt (1992) found that "introjective" patients did better with expressive approaches, while "anaclitic" patients did better with supportive ones, a result confirmed by his own studies (e.g., Blatt, Ford, Berman, Cook, & Myers, 1988) (Lacewing, 2014a, p. 163).

And he concludes that,

> if the success of LTPP [long-term psychodynamic psychotherapy] rested on relationship factors only, this would not explain why interpretive and supportive forms of LTPP – which are differentiated precisely in terms of expressive interventions that are correlated with increased insight – produce different results for different types of patient. It is more economical to think that insight contributes to outcomes, indeed is a mediator, albeit one moderated in complicated ways by patient and relationship variables (Lacewing, 2014b, p. 163).

For the sake of our argument, let us now leave such complications aside and suppose that some research has demonstrated interpretive psychoanalytic therapies as being absolutely more efficacious than merely supportive ones.

2) IF psychoanalytic therapies promoting insights (through interpretations) are more efficacious than merely supportive psychoanalytic therapies.
STILL, it is possible that the interpretations made along the sessions *sounded* true to the patients, that the patients *felt* them as true, but that they are actually false (or perhaps true in a very distorted way, which also does not help us to solve our problem). In this case, the promoter of improvement would be the *performance* or *act* of the interpretation, not its content.
THUS, one would have to compare the outcomes of *psychoanalytic therapies with truthful interpretations* to the outcomes of *psychoanalytic therapies with bogus interpretations*, and, for this, one would have to find truthfulness-indexes that have nothing to do with therapeutic effects. (What I have defended in the last chapters is precisely that there are reasonable means to infer unconscious motives of patients without resorting to therapeutic evidence).

At last:
3) IF independent truthfulness-indexes are found and it is demonstrated that truthful interpretations are more therapeutic than bogus ones.
STILL, we could not discount the cases where consuming bogusness results in something we could reasonably call psychotherapeutic (for example, cases of wellness around the appreciation of charlatanism or of fictional narratives); nor the cases where non-alethic factors can be psychotherapeutic (for example, cases where one changes one's mattress or starts to eat more dark-green vegetables). (I should leave aside here the question of whether truthful insights can produce specific, superior therapeutic effects; this seems plausible only in a scenario where we can accept a specific ethical perspective, but it would be a very complex task to come up with a philosophical demonstration that a psychoanalysed person is better off than, say, a religious fanatic or a hedonist.)
THUS, we would accept Grünbaum's claim that the NCT is false and that truthfulness is not *indispensable* for cure after all.

BUT we would also be forced to accept that truthfulness is *relevant* for cure. Well, if empirical research showed that a truthful dynamic interpretation is more therapeutic than a bogus dynamic interpretation, and if there exist non-alethic therapeutic factors – then a truthful dynamic interpretation could with all justice be called an INUS-cause!
The COROLLARY would be that the efficacy of a psychotherapy can never be evidence of its respective theories' truthfulness. For one cannot depart from the effects of an INUS-cause to find it; one does not inevitably find there were more short-circuits in a group of burned houses than in a group of unburned houses (see pp. 102-105).

7.4 THE PLACEBO IN PSYCHOTHERAPY: IS THERE REALLY A PROBLEM HERE?

We have seen there have been studies comparing the outcomes of supportive and interpretive PDTs (Lacewing, 2014b), but, to my knowledge, there have not been experimental attempts to compare truthful-interpretive PDTs to "bogus-interpretive" PDTs. If we consider the promotion of *truthful* insights the characteristic factor of psychoanalytic therapy, thus, Grünbaum's placebo hypothesis is still standing – it is possible that what counts in the therapy is not the content of the interpretations presented, but the interpretation *qua* performance, which bears the risk of ensuing deceptive insights. A good philosopher, though, has the duty to question the apparently indisputable, nearly imperceptible, presuppositions: why is a placebo considered noxious in the context of psychotherapy, again?

Apparently, if a psychotherapy relies only on incidental factors, it cannot be considered, as it should, a secular, scientifically-informed practice, but rather a dishonest and muddled one. In this case, the claims about how it works would all be false. If it has no knowledge of what its competent factors consist in, it has no control over them; and, if it has no control over them, even if it works well in general, it should not work well all the time. A lack of knowledge, in this case, implies inconsistency of effectiveness.

A pharmacologist, at least, wants to know whether a specific molecule will work with every person, both the hopeful and the hopeless. No pharmacologist questions the power of the placebo; they all know a placebo-effect is real and measurable, that it is not a mirage. But they do not know what the psychosocial incidental factors in their experiments are, what establishes healing expectations, with their effective physiological mechanisms, and whether what triggers them in each subject is the white coat, the air conditioner, the smell of disinfectant, the colour and size of the pill, the prestige of the university, the profound hazel eyes of the pharmacologist or personal background. So, they randomly assign the subjects to experimental and control groups; if there are placebo-effects, they should happen in both groups, and the molecule, if it is really therapeutic, should add its action to the psychosocial incidental factors', producing in the experimental group a stronger therapeutic effect. Pharmacologists want to develop effective molecules, not effective hope[95], because they want to make sure their factors work all the time, not just sporadically. (If pharmacologists knew how to manipulate placebo factors, would the latter be manipulated? Maybe.)

But this picture is not applicable to psychotherapy. Hope, expectation, personal background are all psychosocial factors. Let us remember that a placebo-effect is one whose actual therapeutic factors are all incidental psychosocial factors, that is, psychosocial factors *unrecognized as therapeutic by the theory in question* (see pp. 196-197). So, *although in this positivistic/formalistic tone it makes sense to discuss placebo-effects in psychotherapy, it cannot not make sense to say they are alien or pernicious to psychotherapy*.

A pragmatic proof of that is the Dodo Bird Verdict: that every one of a large array of psychotherapies, each founded upon a distinctive cluster of theories, "has won and must have prizes" shows that *psychosocial incidental factors actually work well all the time in psychotherapy*. Would it not be because in any psychotherapy they are not that different or apart from characteristic factors, no matter how each theory elects the latter? After all, in any psychotherapy both kinds of factors, the incidental and the characteristic, are made of the same psychosocial stuff. Thus, the problem of a presumably secular, scientifically-informed, practice being confused or dishonest about how it works remains, but it becomes less dramatic. Even if what psychotherapists do is not what they claim they do, the Dodo Bird Verdict indicates they have *some* control (an unconscious control, say) over results. The situation is comparable to that of the artisans, who may come up with false theories about their operations but still be consistently able to make beautiful things. So, although in a formalistic dimension a notion of placebo in psychotherapy is important, *in a pragmatic dimension* it is nothing but *trivial*.

This argument is similar to the one given by Irving Kirsch (2005) in the paper "Placebo Psychotherapy: Synonym or Oxymoron?". A psychologist engaged with "placebo studies"[96], Kirsch (2005) there defends that the term "placebo psychotherapy" is both redundant and contradictory. It is redundant as far as

[95]Of course, "hope" also involves molecules, but molecules produced by our own body – not artificial ones.
[96]He is Associate Director of the "Program in Placebo Studies & Therapeutic Encounter (PiPS)" in Beth Israel Deaconess Medical Center/Harvard Medical School. The Director of this Program, by the way, is Ted Kaptchuk, cited in Chapter 5 (see OLP discussion on pp. 222-223) (PiPS, 2021).

> placebos in medical research are interventions that have all of the psychological characteristics of the investigational drug, lacking only the chemically active ingredient. Because psychotherapy has no chemically active ingredient, it can be considered a placebo by definition. Thus, in the context of psychotherapy research, the two terms can be considered synonymous. Note that the psychological ingredients of both placebos may be active, in the sense of having an effect on the target condition. In fact, were they not active, there would be no need to use them as controls in medical research (p. 800).

From the angle of the original instrumental use of the placebo notion, the conjunction "placebo psychotherapy", says Kirsch (2005), is also an oxymoron as far as "it is impossible to devise a meaningful placebo control for psychotherapy outcome research" (p. 800). As in this case the control-therapy and the official therapy share an exclusively psychosocial nature, an artificial difference between them is either not significant or so artificial that the control-therapy cannot resemble a psychotherapy at all[97]. The attentive reader would say that we *have* devised a meaningful placebo control for psychoanalytic therapy in the last section, thus that we cannot assent to this claim. We can assent to it, though, if we depart from Kirsch's non-formalistic definition of placebo-effect: "that portion of the treatment effect that was produced psychologically, rather than through physical means" (p. 791). Well, both truthful interpretations and whatever comes with bogus interpretations are psychologically active elements; thus, our imagined "placebo control" makes sense only through Grünbaum's definition of placebo and through the presumption that the Psychoanalytic Theory of Therapy states that a total lack of truthful interpretations cannot be remedial. In any case, Kirsch (2005) would agree with us in that, *pragmatically*, the notion of a placebo psychotherapy is trivial:

> if the control procedure is effective in treating the disorder, then it is a bona fide treatment. It must have psychologically active ingredients or else it would not work. And it does not matter what those active ingredients are. They may be hope or faith or response expectancy. In principle, these are no different from any other psychological factor than can alleviate distress (Kirsch, 2005, p. 800).

Now, to the extent that a placebo-effect is attained through psychosocial instruments, it is also no *epistemological* intruder in psychotherapy – a conclusion leading us back to one of the arguments of Chapter 5. A placebo-process is a kind of suggestion-process, and we have found most, if not all, of the targets of psychoanalytic theory inside the scope of suggestion, according to our understanding of this notion. If a placebo is also an unconscious process related to fantasies about what caregiver-prototypes expect or need of us, about how to confirm, keep or increase their love, then analysts are also interested in explaining placebos – indeed, Brakel

[97]"Among the procedures used as placebos in psychotherapy research are the following: listening to stories, reading books, attending language classes, viewing films, participating in "bull" sessions, playing with puzzles, sitting quietly with a silent therapist, and discussing current events […]. Indeed, simply being placed on a waiting list has been labeled a placebo […]" (Kirsch, 2005, p. 796).

(2007) has a paper where she discusses how the psychoanalytic theory, especially its hypotheses over the primary process, can help explaining them. Besides, evidence that a placebo took place can naturally be included among the mental products guiding and supporting the inference of a patient's unconscious motivation.

Let us suppose a patient gets better in her first encounter with an analyst. The effect would be a placebo according to any Psychoanalytic Therapeutical Theory the analyst is able to reconstruct. Should he be upset that the placebo-effect occurred? Should he encourage it? Should he ignore it? None of the three. He should silently[98] consider it a mental product requiring an explanation and silently consider the mental products contiguous and similar to it as evidence guiding and supporting its future explanation. The basic fantasy involved in a placebo-process would be the fantasy that the Other (the authority, the caregiver) wishes the subject's cure. By attentively observing the placebo-effect and its psychological surroundings, the analyst could try to answer three basic questions: which kind of cure the patient's Other wishes; who is the patient's Other ("transferred" to the analyst); why this Other wishes this kind of cure.

The last paragraph's example is one in which the placebo-effect seems obvious, by any reasonable standard. But we are still not sure about what the characteristic and the incidental factors of psychoanalytic therapy are, according to the latter's Theory. For we have indicated that the official, characteristic factors of psychoanalytic therapy are *only* truthful interpretations and at the same time concluded that, in any psychotherapy, official, characteristic factors are probably intermingled with clandestine, incidental ones. It seems questionable, thus, that truthful interpretations *in themselves* or *in isolation* should rightly be considered the therapeutic factor of the "talking cure".

Actually, it is contended that our Theory has never established truthful interpretations as the *only* remedial factor in its therapy. In spite of the orthodox defence that it stands opposed to other psychotherapies in that it would operate through insight rather than through the various forms of emotional support condensed under the label of "suggestion", the relevance of the *relationship* between doctor and patient has been recognized since Freud (Lacewing, 2014b).

> […] since at least the 1950s in the United Kingdom and the 1960s in the United States, psychoanalysts have come to recognize the therapeutic action of the relationship, as moderator and/or mediator, and there are few psychoanalysts who would accept the orthodox view now (Abend, 2001; Messer & Wolitzky, 2010; Wallerstein, 1995). The case in favor of the therapeutic action of the relationship is overwhelming and the debate closed (Diamond & Christian, 2011; Eagle, 2011; Gabbard & Westen, 2003) (Lacewing, 2014b, p. 158)

[98]Because it is the beginning of the treatment.

This idea shall be developed in the next section. For now, let us just remark that, if it is true that this debate is closed, Grünbaum's placebo argument fails. First, if all versions of psychoanalysis' Theory of Therapy comprehend that the affective and cognitive outcomes of an intense relationship is one indispensable condition for mental healing, his definition of this therapy's characteristic factor becomes one more strawman to his collection. Let us, moreover, admit that Lundh's MPS common-factors hypothesis is a good explanation for the Dodo Bird Verdict; that is, let us admit that some factors common to many psychotherapies are responsible for their similar success *and* are perfectly recognized under different designations by their respective theories. We may then conclude that one cannot use the Dodo Bird Verdict to argue, as Grünbaum did, that Freudian psychotherapy is probably a placebo; that what actually works in it is hardly what is claimed to work in it. In clearer terms: with the Dodo Bird, it may be the case that different psychotherapists are successful because they operate in a way that is similar to the way analysts avowedly operate.

The argument above challenges Grünbaum by his own standards: it keeps his definition of placebo. If we became indifferent to this theory-relative definition, though, the argument reveals itself as an incomplete response to the problem posed by the philosopher. We still have to prove why it is reasonable to say, as analysts do, that the outcomes of an intense relationships *involving truthfulness* are especially remedial. After all, analysts are not willing to accept the misfortunes of Lundh's "RPP model". And would not everybody be bothered by the possibility that things such as bogus interpretations be remedial?

7.5 THE REMEDIAL FACTORS IN PSYCHOANALYSIS

We have just seen the inflation and the deflation of the monster of placebo in clinical psychoanalysis. If we inflate it too much, we get a picture where the avowed cause of good outcomes in clinical psychoanalysis is eternally suspicious, where its interpretations may be revealed as blah-blahs making no difference at all to its therapeutic results. If we deflate it too much, on the other hand, we get a picture where the connection between genuineness and mental health ceases to be a preoccupation – and we may fall into a relativistic hole where *any* theory and procedure that makes us feel good, thrive, relate well to others, etc., should be praised by secular communities. It is now time for a synthesis of both operations, and for an answer to the question posed at the beginning of this chapter: could therapeutic outcomes in a patient stand for some kind of evidence for the genuineness of the discourse developed along his or her psychoanalytic process?

A question was cast above about which kind of life would clinical analysts consider good, and we have seen that any Therapeutical Theory depends on a concept of pathos; it is from this ethical dimension[99] that we should head toward our answer. What is the aim of a psychoanalytic therapy? Which mental states and processes should analysts counteract? Which values should analysts promote? We cannot linger too much in this complex cluster of questions – it would make us go far beyond the aims of this chapter – but it shall be rewarding to reflect on it for a while.

An answer to such a cluster cannot be too operational, nor too concrete in any other sense (like in the criteria of itemized manuals, such as the DSMs and the SWAP[100]). Among the greatest cultural accomplishments of the psychoanalytic traditions is the promotion of the ideas that the difference between the "unsane" and the "sane" is one of degree, not of substance, and that what is painful or crippling for the mind of one person at a certain moment may not be so for him or her at another moment or for another person[101]. However, most of the psychoanalytic traditions did not fall into moral relativism over the years: most analysts have always reserved their right to say that a person is not alright regardless of what the person says about him or herself and even regardless, I would risk to affirm, of the standards posed by his or her culture.

Allow me to skip many steps in this discussion. I would say that psychoanalytic therapy aims, not at specific changes in external and reported behaviour, but at a relative increase in the patients' *readiness to change something* of their own mental representations and schemas more or less *in keeping with their contacts with current persons and circumstances* – one could call this "mental flexibility", perhaps.

[99]I should elucidate what I mean by the term "ethical" when I apply it to psychoanalysis or to psychotherapy in general. In common-sense, the term usually qualifies altruistic behaviour, or at least a compromise between altruist and egoistic needs; in this case, it is a synonym of "righteous". However, here I use the term in a wider sense: it qualifies the parameters to define which conduct is "good", be it in the sense of "righteous", "enlightening", "interesting", "pleasant" or "sane", and which is not. There is a debate in philosophy about whether an "ethical life" is necessarily a "good life" and vice-versa (see Lacewing, 2019). I cannot enter this debate here; let me just state that in this chapter I take mental health, gratification and righteousness to be all ethical concepts, therefore concepts similar to one another.

[100]The items defining and operationalizing Mental Health in the Shedler-Westen Assessment Procedure (SWAP) (Shedler, 2010) do not seem the final answer to such questions, at least in their philosophical level, because it does not seem reasonable that the mental health of patients who at a specific point in their lives are not "energetic and outgoing" or not "able to find meaning and fulfillment in guiding, mentoring, or nurturing others", for example, should be always labelled relatively poor.

[101]These ideas have not been exclusive of the psychoanalytic traditions, of course.

In a phenomenological and etymological sense, "pathological" is that which brings suffering; on a theoretical level, some of the main things that bring mental suffering, it is said since Freud, are ideas that have remained intact for many ages, that are "stuck in time" or "mummified"[102]. Mental suffering, the psychoanalytic Illness Theory would tell us, is caused by certain ideas that have been repressed. But it would not be the ideas themselves, nor the fact that they have been repressed, that would cause the illness, but rather a *consequence of the fact that they have been repressed*. If unconscious representations and schemas do not rise up to consciousness, they cannot be properly shaken, amended or transformed and, because of this, they become less and less adequate to help us fulfil our ever-shifting intra and interpersonal needs. Mental illness owes thus its existence to a *negative factor*, for it is not the idea and its repression, but the fact that an important idea *has not been elaborated, and hence cannot fulfil present psychological needs*, that causes it. Prudent analysts could never deem pathological a specific behaviour, but only a stereotypical one that has been so for many stages in a person's life[103]. For such analysts, mental health should be measured, not within discrete life moments, but along life stages; it is a diachronic, not a synchronic, concept.

If, in our mummy metaphor, repression is the tombs' walls and the stiff coffins inside the tombs, healings would not be attained only by breaking into and lifting up repressions: the mummy-idea has to be decomposed and used as manure to permit other life forms to grow. To counteract the cause of a symptom, it is not enough that psychoanalytic therapy lift repression and make the patients behold and feel the truth about their unconscious motives; once this psychological truth is revealed, it has to be more or less destroyed. It follows that, if analysts want patients to get better, they must allow patients to "corrupt" their memories (representations and schemas), after the latter are revealed in great amount. For the end of an analysis to be therapeutically satisfactory, it must be epistemically messy.

In the second section of this chapter ("Fact and Value…"), we have posed the question of how analysts could indicate the right values through their interpretations, since we have concluded that the determination of the specific values that a crude utterance of specific facts will indicate depends on the circumstances the utterance is made. Well, analysts should be sensible to how a particular patient is hearing an interpretation at a particular moment, but they should not be at the mercy of such circumstances. Now that we know what the aim of

---

[102]"Mommyfied", a Lacanian would say.

[103]Many analysts would consider, for example, a couple of delirious or dissociative crises in a critical moment of a patient's life as non-pathological insofar as these crises would probably be the only means to attain some psychical homeostasis under that circumstance.

psychoanalytic therapy is, we can give a preliminary answer to that question: when presenting an interpretation, analysts could pick one of the many ways to indicate the value of *courage*. Courage to behold an obsolete and repulsive part of ourselves, and courage to change it in congruence with current needs, both personal and social.

This point was indirectly raised and gracefully presented by Lacewing (2014a):

> We often associate courage closely with strength of will and self-control, with being 'tough' and 'impervious' (Baehr 2011: 178), but in […] [the] context [of facing through intellectual courage the pain against which mental defences protect], that association can be misleading. For what is required [in this context] is a kind of 'letting go'. In contrast to directed introspection or reflection, this involves an openness to one's passions, allowing them to 'surface'. The aim of control is – at least temporarily – relinquished in favour of an approach of understanding and engagement with whatever is felt, however painful, inappropriate, or irrational it seems. The challenge is compounded by the knowledge that one's emotional life is not under one's control, even if it can, over time, be transformed (p. 206).

With a clearer reference to psychoanalysis, Lacewing (2014a) claims that, in its clinical search for truth, psychoanalysis should also promote the value of *compassion* for oneself. Compassion would complement courage because, while "courage enables the pain of deconstructing defences to be endured, compassion diminishes or contains the pain as it preserves a sense of self-worth in the midst of emotions that challenge that sense" (p. 208). In fact, therapists would promote self-compassion, and thus the further deconstruction of defences, just by retaining their "interest, curiosity and care" and remaining "non-moralizing and non-retaliatory in the face of the patient's revelations and passions" (p. 208).

Analysts could find a way to imply, not that the revealed unconscious motive is a cue for pride and bliss, nor that it is a cue for guilt and shame, but that its contradictions should remain conscious for some time so that some episode and/or meaning, from inside or outside therapy, can turn it into another motive. The analyst should not imply to the patient *what should change* in the motive, but more simply that, if the motive is embraced with courage and compassion, the contradictions it comprises will cause it to be changed *in some way that will meet the patient's present needs*[104].

[104]Is the fulfilment of intra- and interpersonal needs through the awareness and transformation of unconscious motives inherently ethical? Can any type of need be met, can motive transformation have any type of result? Can such a simple ethical system ensue intuitively recognized ethical abominations? Let us suppose that a patient in psychoanalytic treatment become aware that he can only reach sexual satisfaction by enacting a masochistic fantasy; he then starts to attend BDSM communities, finds that it is his "*need*" to radicalize such a fantasy in order to be "happy", and radicalizes it at the point of putting his physical integrity at risk. This is a transformation of an unconscious motive turned conscious which I believe most analysts would not consider "good". Which ethical principles would complement the analysts' stance and conduct in this case and in other such difficult cases? This footnote is an admission that the ethical account of clinical psychoanalysis presented in this chapter is incomplete. I fully agree

This is what Freud's formula about the goal and action of his psychotherapy seems to mean. Getting patients to overcome their resistances (through transference love, analysis and dissolution) means conducting them into the mental tombs and getting them to bear gazing and smelling the mummified idea long enough to realize that they would rather say a little prayer for it and then feed the poor garden upstairs with it than leave it there, as it could remain contaminating the groundwater or, worse, waking up and haunting them in their beds. Freud, though, struggles to render explicit that to overcome resistance means more than just to get genuine access to the pathogenic elements – that it also means to transmogrify them through the relationship with the analyst and with other people. An exception can be found in the same 1917 lecture wherefrom Grünbaum derives the Tally Argument. There Freud (1963a) elucidates that suggestion serves, not only to keep the patients' ego in the painful process of accessing their own unconscious contents, but also to *instruct* this ego to grant the libido that has been released in this process some *direct satisfaction* in the world outside and/or to divert it to *sublimation*:

> […] our therapeutic work falls into two phases. In the first, all the libido is forced from the symptoms into the transference and concentrated there; in the second, the struggle is waged around this new object and the libido is liberated from it. The change which is decisive for a favourable outcome is the elimination of repression in this renewed conflict, so that the libido cannot withdraw once more from the ego by flight into the unconscious. This is made possible by the alteration of the ego which is accomplished under the influence of the doctor's suggestion. By means of the work of interpretation, which transforms what is unconscious into what is conscious, the ego is enlarged at the cost of this unconscious; by means of instruction, it is made conciliatory toward the libido and inclined to grant it some satisfaction, and its repugnance to the claims of the libido is diminished by the possibility of disposing of a portion of it by sublimation. The more closely events in the treatment coincide with this ideal description, the greater will be the success of the psycho-analytic therapy (Freud, 1963a, p. 455).

Years after such Lectures, in a 1934 paper that would become classical, Strachey (1999) intends to deliver an illuminating extension of this formula through Freud's then recent notion of superego, Sandor Radó's then recent theory that the hypnotist becomes the hypnotized subject's "parasitic superego" and Melanie Klein's then recent theory of superego formation. With his famous argument that the analyst promotes change by incorporating and deflating, through accurate and tolerant interpretations, the patient's superego, Strachey (1999) argues that truthful information and sensible relationship can interact to heal the patient. According to

---

with the following excerpt from Lacewing's paper "A Truthful Way to Live? Objectivity, Ethics and Psychoanalysis":

> We may say […] that the demands of truthfulness condemn certain ethical values and practices, and place constraints on which forms of ethical life are best. But this is not a complete defence of […] [the claim that a specific ethical life is best in general], as these constraints will not yield a single determinate way to live that is best (Lacewing, 2019, pp. 192-193).

Glenn Gabbard (1999), much of the controversy over whether change in the psychoanalytic process is ascribable to transference interpretation or to transference itself began with this paper, and the model it presents has served as the starting point for all others.

As Strachey (1999) reads in Freud that "the superego should play an important part, direct or indirect, in the setting-up and maintaining of [...] repressions and resistances" (p. 69) and that "the analyst brings about his changes in the patient's attitude by means of suggestion", he concludes that "the analyst owes his effectiveness, at all events in some respects, to his having stepped into the place of the patient's super-ego" and changed it by means of suggestion (p. 69). The superego would be a part of the patient's mind "in which a favorable alteration would be likely to lead to general improvement" and "which is especially subject to the analyst's influence" (p. 69). An analyst would be therapeutic in incorporating the patient's superego – due to the spontaneous process of transference, which is boosted by the statement of the fundamental rule ("say anything that comes to you mind, even if unpleasant or nonsensical, etc.") and by tallying interpretations – and, then, in deflating it, so to speak, from inside – through averting "any real behavior that is likely to confirm the patient's view of him [the analyst] as a 'bad' or a 'good' fantasy object" (p. 75). He continues:

> This is perhaps more obvious as regards the "bad" object. If, for instance, the analyst were to show that he was really shocked or frightened by one of the patient's id-impulses, the patient would immediately treat him in that respect as a dangerous object and introject him into his archaic severe super-ego (Strachey, 1999, p. 75).

To be honest, Strachey (1999) does not talk about the therapeutic action upon the superego as a sort of implosion, or deformation from the inside. He rather defends that, with the analyst becoming the patient's "auxiliary superego", there runs a first phase where "a portion of the patient's id-relation to the analyst is made conscious" (p. 73) and a second phase where "the patient becomes aware that his impulse is directed toward an archaic fantasy object and not toward a real one" (p. 74). Regarding this second phase, Strachey (1999) counts on the patient's inner ability to refine a "sense of reality" and his recommendations above for the analyst not to confirm the patient's projections would serve "not to submit [...] [this sense of reality] to any unnecessary strain" (p. 75). Both phases compose what Strachey (1999) calls a "mutative interpretation". In any case, however, the analyst would have a role in updating the patient's old-age complex derived from projected-introjected fantasies of aggression and real admonitions, the patient's superego, by taking its place. To mix Strachey's model with our metaphor (now slightly different, to give it an oneiric face): the analyst takes the patient into the tomb, is bitten by the mummy-vampire, but, with the help of the patient (or is the patient also the vampire?), pulls everyone to daylight, where the original vampire becomes ash and fertilizes the orchard[105]; the analyst then comes back to being a person sitting on the armchair. Of course, the superego is composed of a *host* of underground vampires, and the work goes on.

The Kleinian influence is strong in this paper. The same mechanism that has brought about the illness – the original, oppressive superego of the suffering mind – would be responsible for cure – its own deflation. For the "mother" of English psychoanalysis, the formation of the superego involves a projection of aggressive impulses onto the perception of external objects and a subsequent introjection into the psyche of the resulting aggressive figure "outside". This would be a vicious circle where the analyst – a hyper-vigilant and yet calm, tolerant and realistic figure – would seek to open a breach.

Strachey's model has become classical because it has, perhaps for the first time, given suggestion and the analyst's attitudinal or emotional acts – as opposed to the analyst's investigative and informative acts – clear and strong roles in the healing process. Most importantly for our purposes here, it has theoretically demonstrated that both kinds of acts can be connected: analysts would not only show patients what lies in their unconscious; *exactly because they would show patients what lies in their unconscious*, they would be merged with this unconscious and reshape it through their non-neurotic acts; the analyst would become an "underground vampiric" idea in the patient's mind but would not act like one. Gabbard (1999) comments that, is spite of Strachey's paper, there has been "a delay in fully recognizing the value of the noninterpretive mechanisms of change in psychoanalytic therapy" due to a "stigmatization of supportive therapy and of the nonexpressive elements in treatment" (p. 65). Analysts of today, though, would be "more inclined to think of psychoanalytic therapy as occurring on an expressive-supportive continuum that balances interpretive and noninterpretive interventions" (Gabbard, 1999, p. 65).

Indeed, very recent accounts of the mechanism of change in psychoanalytic therapy – and in psychotherapy in general – are consistent with the representation- and schema-bending account we have defended so far. Richard Lane and colleagues have been developing a neuroscientific theory to explain enduring change in all psychotherapy modalities (Lane, Ryan,

[105]Some analysts would say, perhaps, that, if the vampire can tolerate sunlight, then everyone should befriend it and feed it with the blood of some animal once in a while. If translated back to theoretical terms, this could also be considered a desirable outcome of psychoanalytic therapy. In Lacewing (2014a) we see one such recent translation:

> But should both the passions and the judgements of their inappropriateness remain unchanged [with the deconstruction of defence mechanisms], as frequently happens, then having granted them a place in one's conscious psychology, one may at least become aware of such influence as they may have on one's sense of oneself and the situations one faces, and correct for it as best one can (p. 209).

Nadel & Greenberg, 2015), and their dialogues with the neuropsychoanalytic movement is already significant (e.g., Lane, 2020). From the knowledge that "memories become labile or malleable whenever they are recalled", that "information made available when the memories are in the labile state can be incorporated into the original memory"[106] (Lane, 2020, p. 188), and that "emotion is a particularly potent way to update memories because synaptic plasticity, which is the molecular basis for encoding memories, is enhanced by the neurotransmitters and hormones (e.g., norepinephrine, cortisol) that are activated by emotional arousal"[107] (p. 189), Lane and colleagues propose a theory where psychotherapy brings about enduring change "through reconsolidation of emotional memories" (Lane, 2020, p. 188).

Lane (2020) asserts that this theory is consistent with Freud's claim that patients suffered from reminiscences and with Freud and Breuer's "Studies on Hysteria":

> [According to the book,] if recall could be accompanied by the experience and expression of the affect associated with the trauma, the memory would go through a process of what Freud called "retranscription." This process of retranscription would change and update the memory and the symptoms would be resolved (Freud, 1896/1966). It is therefore remarkable to note that, to our knowledge, Freud's concept of "retranscription" was the first reference to what has now come to be known as memory reconsolidation (p. 198).

Not only could his theory give a modern explanation of why psychoanalytic therapy works, Lane (2020) holds, but also of how psychoanalytic therapy would relate to the wide range of approaches available and of what would distinguish it from these approaches.

The theory includes the "integrated memory model (IMM model)", which tells that episodic memory, semantic memory and emotion are so linked to each other that, whenever one of these mental domains is activated, the other two are activated as well. One psychotherapy would diverge from another by the "doorway" through which it accesses the "integrated memory model". For example, "PDT preferentially enters this interactive matrix through episodic memories, whereas cognitive-behavioral therapy […] preferentially enters it through semantic memories and associated thoughts" (p. 189).

The central point of the theory is the "LRNG Model" (after the initials of Lane, Ryan, Nadel, and Greenberg, the last names of the model's creators), according to which enduring change in psychotherapy relies on three factors (Lane et al, 2015; Lane, 2020). The first factor is the reactivation of old memories and their associated painful affects through explicit recall or reminders; the second is the engagement in "new emotional experiences during treatment that are incorporated into those reactivated memories via the process of reconsolidation" (Lane,

[106]This would come from the works of Nadel, Samsonovich, Ryan and Moscovitch (2000) and Elsey, Van Ast and Kindt (2018) (Lane, 2020).
[107]This would come from the work of Schwabe, Joëls, Roozendaal, Wolf and Oitzl (2012) (Lane, 2020).

2020, p. 189); the third and last is the reinforcement of the updated memory by exercising new manners of exchanging with the world in a miscellany of contexts. Lane (2020) argues that enduring change in psychotherapy results from a sort of learning process that consists in the interaction between emotions and different types of memory: episodic and semantic memory that have risen to consciousness accompanied by an intense emotion, and are due to this fact in a labile state, would be transformed by the unprecedented emotional experiences that psychotherapy enables and promotes.

Semantic memory would correspond to what since Chapter 3 we are calling "mental schemas"; the therapist who can change semantic memory can change the "recurrent pattern that is the focus of treatment" (Lane, 2020, p. 189). In PDT, such corrective emotional experiences would be possible in the transference relationship with the analyst: the reactions of the analyst to the patient's most intimate motives would contribute to the patient's learning of more trusting and less defensive ways to relate to others (Lane, 2020). The integrated memory and the LRNG models would present a framework that unifies psychotherapy modalities and that as such is not limited to PDT; it may be the case, however, that "the frequency of sessions and the intensity of the relationship with the therapist may provide a learning context for the transformation of recurrent patterns through reconsolidation that may differentiate PDT from other modalities and provide unique advantages" (Lane, 2020, p. 199).

Even though it has been inspired by consolidated knowledge about the brain, Lane and colleagues' model to explain enduring change in psychotherapy is presented above at a psychological level of description. However, the team has also been attempting to "provide greater specificity with respect to its neural basis" (Lane & Nadel, 2020, p. 434). In the book "Neuroscience of Enduring Change: Implications for Psychotherapy", they have given a first translation of the IMM and the LRNG models into neural circuitry models with the aim of making them more specific and amenable to revision by empirical research; it should help the team "identify where gaps exist and where research is most urgently needed" (p. 461).

From all the discussion above over the goal and action of psychoanalytic therapy, we see a well-supported principle that psychopathological processes become more susceptible to beneficial alterations when *integrally accessed*, that is, when *all of their dimensions* become *conscious*. It follows that, if the psychological processes felt, thought and talked about during session are not authentic – are too partial or superficial, misunderstood, deceptive, in sum, if they do not illuminate the patient's psychological reality – the analyst's attempts at relational correction will have little effect, if no effect at all. One has to get to the mummy to change the mummy (scaring a camel away will not prevent the mummy from chasing us). In other words, it follows that *psychological truthfulness may play a role in psychological cure*; that *truthfulness is relevant in any psychotherapeutic endeavour*. Why "may"? Why "relevant" rather than "necessary", as Freud's NCT would dictate? Because, as already observed, symptom remission can be reached independently of truthful perceptions; it would be thus more appropriate to say that truthful insights are INUS conditions (rather than INNS conditions; see pp. 102-105) conditions for symptom remission.

Truthful insights are a necessary but insufficient condition for symptom remission because only bringing a pathogenic idea to a person's consciousness without also transforming it to serve, or at least not to hinder, the satisfaction of that person's current needs does nothing for the mental health in question. As Lane (2020) argues:

> Recall of past traumas or adverse experiences without competing emotional experiences will lead to a memory that is further reconsolidated and thus more likely to be retrieved during similar situations in the future. As the memory itself is strengthened, so too is the emotional response and the semantic structures that result in novel situations being interpreted in maladaptive ways. Recollection alone only serves to reinforce and further ingrain the patient's original version of the traumatic or adverse memories, and is insufficient to bring about clinical change (p. 200).

In the second place, consciousness of motives plus correction of motives is a set of sufficient but unnecessary conditions for mental healing because conditions such as psychotropic ingestion, art fruition, religious experiences and meaningful episodes in general (new partner, new friend, new work, an apology, a speech, etc.), among others, may be responsible for symptom alleviation without involving any truthful insight. But if we said that psychoanalytic Illness Theory is the one that best accounts for the concept and the cause of mental illness – that mental illness consists in the vacuum-conserved character of a relevant idea, in its unsusceptibility to change in the face of new inputs – we would have to conclude that such non-alethic factors counteract this unsusceptibility as well as truthful-insight-plus-idea-revamping but, of course, they would counteract it mostly without making the pathogenic idea integrally conscious[108]. We could fit this consequence into our model with the hypothesis that such non-alethic factors get to the unconscious idea and affect it *indirectly* (whatever it may mean in neurological terms). Back to our metaphor, it is as if an artistic or religious experience, a psychotropic ingestion, etc., were a long-rooted tree, an army of earthworms, an infinity of dirt microbes, in fact anything that could reach and decompose the vampiric creature in the tomb without one having to head to the tomb and open it in person ("in ego"). For sure, all that certain psychotropics, religions, etc., do is to lock the tomb door with more padlocks

[108]Here I refer to cases where the non-alethic factors cited above are not conjugated with typically psychotherapeutic interventions, as there exist psychedelic therapies and art therapies, for example.

and just depollute the groundwater; but anyone would agree that in this case the good outcome was not so good after all, for then, although the discomfort or misery might have been attenuated, the mind is not more flexible, adaptive, creative, free, etc., than before. In any case, though, our model implies that formal psychotherapies promoting truthful insights, especially if these are emotional and semantic insights, have a much greater chance of affecting a pathogenic idea than religious experiences, art fruition, life episodes, etc., and a smaller chance than such expedients of causing deleterious side-effects.

This defence that there may be sufficient conditions for healing that do not involve truthfulness is consistent with Bucci's Tripartite Probabilistic Tally Argument, presented in Chapter 2 (see pp. 46-48). The TPTA would be "compatible with 'cure' being achieved by means other than change in the PME schemata" (Bucci, 1989, p. 262) – in our metaphor above, this "cure" would correspond to the padlocks and the depollution. It would also be compatible "with change in the PME schemata occurring through means other than activation of referential connections" (p. 262) – the dirt microbes metaphorically cited above. Finally, the TPTA would be compatible "with activation of the referential connections occurring by means other than through an accurate intervention" (p. 262) – the tree of our metaphor. But the TPTA also "holds that alternate means are less likely in each case" (p. 262). To be fair, this *whole section* is consistent with Bucci's 1989 response to Grünbaum, and her model in this response is very similar to Lane's.

There comes finally the moment to answer the question of this chapter: how is a psychoanalytic discourse that tallies with the mental reality of a patient related to this patient's healing, and vice-versa? Again, we must evoke the evidential pattern of INUS-causation (see p. 105). If a truthful insight is an INUS condition for healing, then healings cannot be unequivocal indexes of previous interpretation/report accuracy and experience authenticity – but truthful insights can enhance the chances that the patient will get better. At the same time, we have just concluded that, if our model is right, then formal psychotherapies promoting truthful insights have a much greater chance of getting the patient better than alternative therapies and expedients. So, even though healings cannot be considered *unequivocal* evidence for antecedent accuracy and authenticity, we would be inductively correct if we considered them *complementary* evidence for that hypothesis – complementary to the kind of evidence we have discussed in Chapters 4 and 5. This means that the patient getting better favours an interpretation if and only if the analyst has relied on control-instances, repetitive specificities, smoking guns, cognitive baselines, etc., to interpret. However, it also means that the truthfulness of an interpretation would be rigorously *undetermined* in case the interpretation were guessed rather than inferred and therapeutic evidence were all the analyst had. Finally, it means that lack of therapeutic effects is no unequivocal index of lack of truthfulness – it may be just a consequence of an exaggerated and prolonged abstinence on the analyst's part.

7.6 CONCLUSION: A RESPONSE TO GRÜNBAUM

Freud (1958) once commented that "psycho-analytic treatment is founded on truthfulness" (p. 164). We have seen it is also founded on something else – one may modernly call it memory reconsolidation or perhaps reorganization of perceptual-motoric-emotional schemata, but not even Freud was afraid to call it "suggestion". According to most psychoanalytic communities today, psychoanalytic healing involves not only the dramatization, revelation and verbalization of the truth about the patient's mind, but also – to use Grünbaum's term – the "contamination" of this truth through the affective dimension of a close relationship. Through Lundh and Lane, we can sustain that the two factors – truthfulness and its "contamination" or reconsolidation – are probably the main remedial factors of most psychotherapies, even though the theory of each approach give them distinct names. It follows that the Dodo Bird Verdict cannot imply that psychoanalytic therapy is probably a placebo; Grünbaum cannot be correct here, and this in keep with his own definition of placebo-therapy. It follows, besides, that truthfulness can still be considered *relevant* for mental healing, for reconsolidation (or reorganization, suggestion, etc.) is *better* operated at the moments when the pathogenic ideas are genuinely and intensely accessed.

In the last phrase, I have used "better" instead of "only", and "relevant" instead of "necessary". This is because I endorse Grünbaum's point about the unsoundness of Freud's 1917 Tally Argument. The NCT is indeed false: a tallying interpretation is not a necessary condition (in a set of necessary and sufficient conditions) for symptom alleviation insofar as non-alethic factors can equally ensue symptom alleviation. Although Grünbaum's attack on the Tally Argument has been inoffensive for psychoanalytic therapy *per se*, it is not without consequences for the epistemology of the research than can come from this therapy. If truthfulness is relevant but unnecessary for salubrious effects, the latter's variations indeed cannot serve as unequivocal evidence in clinical-psychoanalytic research. Not all is lost, though: they can serve as complementary evidence. A hypothesis that is deemed truthful because there is plenty of outcome-independent evidence for it is expected to be remedial in case it is expressed and worked-through for and by the patient.

However – it should have become clear – while I agree with Grünbaum's debunking of the Tally Argument, I do not agree that cogent research in clinical psychoanalysis is impossible. First because I do not consider the soundness of the Argument the one thing that can save this kind of research. Second because salubrious effects, we have just said, can be taken as complementary evidence. This claim presupposes that "contamination" favourable to salubrious effects can only take place *after* the main evidence is collected – the evidence from the "baseline" phase, discussed in Chapter 5 – and strong, abductive inferences can be made out of it and communicated. Does it make sense to say that, in psychoanalysis, what is advantageous to the therapy is damaging to the research? Yes and no. Healing depends on the "contamination" of the patient's motives as they stand; but this "contamination" is only prudent (only *possible*, sometimes) after the main data suitable for use in research are collected.

8 CONCLUSION

In this doctoral dissertation, we have walked some of the logical paths to which the three great Grünbaumian charges against the epistemology of clinical psychoanalysis have conducted us and, with our generous models of psychoanalytic theory, inquiry, inference and therapeutic intervention as well as with some measures that could probably make this work more recognizable as a scientific endeavour, we hope to have evened those bumpy roads out, or at least to have honestly diverted from them. We have chosen and settled some deviations and maintenance initiatives many authors had already started in order to explore the roads without tumbling down. Should the reader consider it unique or not, the whole path we have chosen and settled should at least call the wanderers' attentions to some forgotten or poorly discussed resolutions, and perhaps convince them that such resolutions are significantly promissory.

In Chapter 4, we have argued that, as long as inferences in clinical psychoanalysis are almost identical to Explanationist-Bayesian inferences in the historical sciences, they are essentially Millian – or at least that they could become essentially Millian without ceasing to be essentially Freudian. We can surmise explicit and implicit comparisons of target-instances to control-instances in the inferences of the three kinds of hypotheses we have identified in clinical psychoanalysis, namely, the particular, the categoric and the core hypotheses. It was our suggestion, thus, that a guardian of the Method of Difference such as Grünbaum could be content with clinical psychoanalysis, and that he specifically have not been so for the reason that his range of psychoanalytic sources has been too narrow and his confidence in the analysts' accounts of their own inferences, too high. More notably, we have demonstrated that the facts that analysts entertain only repetitive and specific thematic affinities and that it is perfectly possible to devise control-groups for the inference of categoric and core hypotheses move clinical psychoanalysis away from the thematic affinity and the *post hoc ergo propter hoc* fallacies. We have also seen that, if clinical psychoanalysis can sustain a certain inferential relationship between those three kinds of hypotheses, it can also be exempt from the both pre- and post-Grünbaumian charges of confirmation bias and *petitio principii*.

In Chapter 5, we have pointed to a conceptual and theoretical gap in Grünbaum's contamination argument. We have been at pains to fill this gap and we have conquered a definition of suggestion and a telling fact about the conditions to its advent. Suggestion, we have concluded, is a partially unconscious process in which certain intra- and intersubjective circumstances cause a subject's mental state to conform itself to his or her representations of an authority's wish. We have presented empirical evidence that the intrasubjective circumstances are more determinative of the suggestion response than the intersubjective ones. The whole discussion has led us to confound the domain of clinical psychoanalysis with the domain of suggestion: symptoms, dreams, parapraxes, jokes, etc., have been revealed as nothing but pure autosuggestions, and the most famous version of the phenomenon – the hetero- or induced suggestion of hypnosis and persuasion, for example – has also been revealed to a certain degree as autosuggestions, that is, as subject matters to clinical psychoanalysis as much as any one of the classical unconscious formations. Having taken heterosuggestion responses to be valuable evidence for the patients' motivations, we have witnessed the deflation of Grünbaum's contamination bomb. But we have also argued that this deflation does not imply that analysts are freed from dealing with the problem of induced suggestion but, on the contrary, it implies that they should devote a great deal of attention to it. To include induced suggestion in their database, analysts should be able to reduce to a minimum the ones they foster (in order to avert superfluous complexity) and to notice when, where and how induced suggestion has occurred. Finally, we have proposed a "cognitive baseline" as a measure to overcome such challenges: the installation of an initial phase in the treatment where the interventions are less directive (according to both common-sense and empirical research on suggestibility) and where the patients' responses can serve as controls to their responses in a late, more directive phase.

In Chapter 6, inspired by Grünbaum's debunking of the Tally Argument, we have faced the problem of how truthful interpretations could be related or unrelated to therapeutic effects in the patient. We have first taken a long detour to visit the classical debate about the relationship between fact and value and apply it to clinical psychoanalysis, concluding that truthful psychological utterances could promote in patients both "good" and "bad" responses, and that analysts should clarify their notion of mental health if they wish to defend a treatment based on truthful interpretations. Next, we have discussed the Dodo Bird Verdict and explored an inflation of Grünbaum's placebo argument, confirming that good or better-than outcomes in a psychotherapy cannot prove the psychological theory on which it is based and concluding that the only way to prove that truthful psychoanalytic insights are therapeutic would be by comparing (equipped with non-therapeutic indexes of truthfulness) the outcomes of psychoanalytic therapies with truthful interpretations to the outcomes of psychoanalytic therapies with bogus interpretations. We have also explored a deflation of Grünbaum's placebo argument, much in the manner of what we have done in Chapter 5 with the contamination argument: if placebo factors have a psychosocial nature, they would not pose an inherent threat for psychoanalysis; neither a therapeutic threat, for the Dodo Bird Verdict would prove that placebo factors have consistent effects, nor an epistemic threat, for placebo responses would be relevant data as much as any symptom. We have seen that, anyway, Grünbaum is wrong by his own standards in his placebo argument if we consider Lundh's MPS model and the fact that the analyst-patient intense relationship is included among the characteristic factors of the psychoanalytic Theory of Therapy. Finally, we have discussed the place of truthfulness among such factors through the wits of Strachey, with his classical account on the therapeutic mechanism of psychoanalysis, and Lane, with his recent neuroscientific account on the same mechanism, theorizing with them that a *genuine access to a psychopathogenic element* (a superegoic structure, an episodic memory, a semantic memory, an emotional reflex) makes it more liable to change, and that this change is not automatic: the element must be "contaminated" to fit the patient's present needs and the values both analyst and patient consider of worth, a process for which a *loving and honest therapeutic relationship* would also be crucial. The final conclusion of this intricate chapter was, therefore, that truthful interpretations can be at theoretical and empirical levels considered causes of mental improvements – not INNS causes, as Freud supposed with his NCT, but rather INUS causes. This fact would imply that such improvements cannot be considered *unequivocal* evidence for antecedent truthfulness, but only *complementary* evidence for it. Grünbaum would be right, but shallow, about the Tally Argument. The therapeutic and the scientific goals of psychoanalysis are, indeed, harmonious.

Let us localize the responses to Grünbaum given in this doctoral thesis in a couple of the groups we have presented in Chapter 2. The responses in Chapter 4 disagree with Grünbaum's conclusion – that it is impossible to make retrospective causal inferences in the clinical context – and with his premises – but only with his premises about how clinical psychoanalysis works, *not* with his scientific standards for causal inferences. Likewise, the ones in Chapter 5 disagree with his conclusion that much of the evidence in a clinical-psychoanalytic context is probably and blindly deceptive, as well as with his premises that suggestion is always a disturbing element in psychological research and that one cannot control for it in the clinical context – but not with his premise that suggestion is ubiquitous in the clinical context. Again, the responses in Chapter 6 contests his conclusion that the unsoundness of the Tally Argument implies a dead-end for clinical-psychoanalytic research, while partially contesting his premises. In Chapter 6, the unsoundness of the Tally Argument and its impotence to guarantee the truthfulness of psychoanalytic interpretations is confirmed, but the irrelevance of truthfulness for mental healing is not, nor is the premise that psychoanalytic therapy is probably a placebo.

The responses as a whole, thus, are definitively inconsistent with Grünbaum's conclusion that it is impossible to do cogent clinical-psychoanalytic research on the causes of human behaviour. We have seen in Chapter 3 that there is a kind of knowledge only this kind of research is able to produce. It is an irreplaceable and extremely relevant kind of research, and the most encompassing conclusion of this thesis is that it is already cogent in many respects and can become more and more cogent through the stabilization or implementation of certain principles and guidelines - without thereby vandalizing its psychoanalytic face. This conclusion stands upon strictly philosophical resources and is thus limited by this fact. For the project of making clinical-psychoanalytic research widely recognized as scientific research, what is needed is, of course, a combination of philosophical, technological and institutional efforts. The recent psychological research with clinical data and the fact that a collaborative, global and unified database to support it is thriving are first signs of technological and institutional efforts in this direction. Should epistemologists interested in this movement take a closer look at it (I include myself in this imperative) and should therapists and empirical researchers take more time to read their insights (the armchair is not the lab, but it has its advantages) – in sum, should a rich dialogue between philosophers, therapists and empirical researchers in favour of case-based psychological research become more and more intense, I believe a rich source of psychological knowledge will have been set up for our challenging 21st century. It is moreover great news in this regard that one of the aims of the currently thriving neuropsychoanalytic program is to perform experimental research on core claims of the Freudian legacy.

Much was here left unresolved, in particular the problem of the reliability or consistency of the inferential method presented in Chapter 4. As already noted, it needs to gain a greater level of precision so that most (if not all) persons who are educated to operate it and who are faced with the same pieces of clinical evidence can arrive at similar (if not identical) conclusions. How could every analyst agree about the high specificity of a mental product, or about its *higher* specificity in relation to another mental product? What would be the procedures and the criteria to determine the statistical abnormality of a mental product in relation to the patient's culture or overall mental life? Which degree of repetitiveness in a behavioural affinity, with and without a high specificity of its theme, should be considered high enough to prove that it consists in a causal network? We would need a value for Pr(H|E&B) in the face of which every clinical-psychoanalytic researcher would agree that the hypothesis in question has been confirmed – but first we would need to calculate it, and for such we would need a precise way to estimate the values of Pr(E|H&B), Pr(H|B) and Pr(E|B). We have seen Lipton's argument that Explanationism can help us estimate such values, but even if he has convinced us here there remains the question of how clinical-psychoanalytic researchers would agree about a good or a best explanation – about the explanatory virtues of a hypothesis. By the way, we have also seen Lipton's argument that Explanationism is the best start to account for the mystery, not only of how we test our hypotheses, but also of how we *generate* them. Explanationism may be the best start to account for this mystery, but even with it in hands one may still wonder whether there could be a sort of algorithm to *generate* candidate hypotheses in clinical-psychoanalytic research – a procedure that, for example, would lead every researcher in contact with a collection of transcribed sessions of a certain patient to the same list of candidate hypotheses about the overarching mental schema in action in that patient's unconscious. From Chapters 5 and 6, too, we can point to similar questions left unanswered – for example, how long should a "cognitive baseline" last, and under which operational conditions would it be reasonable to use therapeutic effects as complementary confirmation of an interpretation? In respectable programs of scientific research, some reliability problems similar to these have been resolved and some have not. In clinical-psychoanalytic research, though, it seems that problems of reliability have not even been properly recognized yet. The models I propose in this thesis fall short in overcoming this fact, and I should be more attentive to it if I am to discuss such models in the years to come.

Regarding the discussions I was able to substantially develop, on the other side, I should also cite some of its points which I think deserve further thought, on my part or on the part of whoever deems them interesting. First, the hypothesis that something like the Method of Difference underlies clinical-psychoanalytic inferences since Freud deserves to be historically and epistemically inspected abreast debates on the intellectual influence of Mill on Freud, on the psychoanalytic instantiation of IBE and on the relations between that Method and IBE. Needless to say, the hypothesis that notions of probability underlie clinical-psychoanalytic inferences since Freud also calls for deeper inspection. Third, as indicated in Chapter 4's conclusion, a more detailed *justification* of Millian, Bayesian and Explanationist models as they may be drawn from psychoanalytic inferences would be helpful for the debate on how scientific such inferences actually or potentially are. The possibility of applying empirical research on suggestion and suggestibility to the clinical-psychoanalytic method also deserves further thought. Finally, and I hope this has become clear along Chapter 6, the issue of which ethical systems underlie and justify the various kinds of psychotherapy is not only interesting but also ridiculously primary, and I sustain that all philosophers of psychology should be concerned with it; for the facts about our minds, by themselves, will never be able to tell us what is good for our minds.

The suggestion that engaging with Grünbaum's problems was like walking on a dirt road with long-carved, rough paths – the logical consequences of the problems –, but equipped with tools to open new roads ahead or to even the bumpy ones out – with the power to make syntheses and propositions – this suggestion looks familiar. The metaphor could also be applied to the psychoanalytic treatment. Psychoanalytic patients should deal with the given, with the rough paths that life has carved in their minds, but they should also find the means to change the given. We can only make our mental paths good if we depart from the ones that already exist: apart from the fact that it is harder to open up a new road than to even out an already carved one, it is much harder to open up a new road when standing upon wild vegetation. With satisfaction I realize that this work's propositional-synthetical method (as well as its metaphor-saturated rhetoric) tallies with its object, namely, the psychoanalytic method and theory. Is philosophizing just telling arguments to lay on the couch?

REFERENCES

Abend, S. M. (2001). Expanding psychological possibilities. *Psychoanalytic Quarterly*, *70*, 3–14.

Abrahamsen, D. (1973). *The murdering mind*. New York, NY: Harper & Row.

Alexopoulos, J. S., Grieve, R. A. F. & Robertson, P. B. (1988). Microscopic lamellar. Deformation features in quartz: Discriminative characteristics of shock-generated varieties. *Geology*, *16*, pp. 796–799.

Allport, G. W. (1961). *Pattern and growth in personality*. New York, NY: Holt, Rinehart, & Winston.

Associação Universitária de Pesquisa em Psicopatologia Fundamental. (2020, February 28). *Produção científica: Biblioteca de psicopatologia fundamental* http://www.fundamentalpsychopathology.org.br/producao-cientifica/producao-cientifica-biblioteca-de-psicopatologia-fundamental/

Azcona, M. (2016). *Las críticas de popper y grünbaum al psicoanálisis: Un abordaje epistemológico de la racionalidad freudiana* [Doctoral Dissertation, Universidad Nacional de La Plata]. UNLP Repository. http://sedici.unlp.edu.ar/handle/10915/59340

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. Oxford, England: Oxford University Press.

Baghramian, M., & Carter, J. A. (2020, September 15). *Relativism*. Stanford Encyclopedia of Philosophy. Retrieved February 29, 2020, from https://plato.stanford.edu/entries/relativism/

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Basch, M. (1980). *Doing psychotherapy*. New York, NY: Basic Books.

Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization-based treatment*. New York, NY: Oxford University Press.

Baudouin, C. (1921). *Suggestion and autosuggestion: A psychological and pedagogical study based upon the investigations made by the new Nancy School*. New York, NY: Dodd, Mead and Company.

Baudouin, C. (2015). *Studies in psychoanalysis: An account of twenty-seven concrete cases preceded by a theoretical exposition*. London, England and New York, NY: Routledge.

Beer, P. A. C. (2015). *Questões e tensões entre psicanálise e ciência: Considerações sobre validação* [Master's Dissertation, Universidade de São Paulo]. USP Repository. https://www.teses.usp.br/teses/disponiveis/47/47134/tde-04042016-122531/en.php

Bell, D. (2009). Is Truth an illusion? Psychoanalysis and postmodernism. *The International Journal of Psychoanalysis*, *90*, 331–345.

Bering, J. (2010, August 17). *Oedipus Complex 2.0: Like it or not, parents shape their children's sexual preferences*. Scientific American. https://blogs.scientificamerican.com/bering-in-mind/oedipus-complex-2-0-like-it-or-not-parents-shape-their-childrens-sexual-preferences/

Bernheim, H. (1886). *De la suggestion et de ses applications à la thérapeutique*. Paris, France: Octave Doin.

Bernheim, H. (1891). *Hypnotisme, suggestion, psychothérapie: Études nouvelles*. Paris, France: Octave Doin.

Bernheim, H. (1917). *De la suggestion*. Paris, France: Albin Michel.

Binet, A. (1900). *La suggestibilité*. Paris, France: Schleicher Frères.

Bisiach, E., & Rusconi M. L. (1990). Break-down of perceptual awareness in unilateral neglect. *Cortex*, *26*, 643–649.

Blake, R. & Wilson, H. (2011). Binocular vision. *Vision Res.*, *51*, 754–770.

Blanck, G., & Blanck, R. (1974). *Ego psychology: Theory and practice*. New York, NY: Columbia University Press.

Blatt, S. J. (1992). The differential effect of psychotherapy and psychoanalysis with anaclitic and introjective patients: The Menninger Psychotherapy Research Project revisited. *Journal of the American Psychoanalytic Association*, *40*, 691–724.

Blatt, S. J., Ford, R. Q., Berman, W., Cook, B., & Myers, R. (1988). The assessment of therapeutic change in schizophrenic and borderline young adults. *Psychoanalytic Psychology*, *5*, 127–158.

Bleuler, E. (1921). *Das autistisch-undisziplinierte Denken in der Medizin und seine Überwindung*. Berlin, Germany: Verlag von Julius Springer.

Boag, S. (2017). *Metapsychology and the foundations of psychoanalysis: Attachment, neuropsychoanalysis and integration*. London, England and New York, NY: Routledge.

Bohor, B. F., Foord, E. E., Modreski, P. J. & Triplehorn, D. M. (1984). Mineralogic evidence for an impact event at the Cretaceous-Tertiary Boundary. *Science*, *224*, 867–869.

Boisen, A. (1936). *The exploration of the inner world*. New York, NY: Harper.

Brakel, L. A. W. (2007). The placebo effect: Can psychoanalytic theory help explain the phenomenon? *American Imago*, *64*(2), 273–281.

Brakel, L. A. W. (2009). *Philosophy, psychoanalysis, and the a-rational mind*. Oxford, England: Oxford University Press.

Brakel, L. A. W. (2015). Critique of Grünbaum's "Critique of psychoanalysis". In S. Boag, L. A. W. Brakel & V. Talvitie (Eds.), *Philosophy, science, and psychoanalysis: A critical meeting* (pp. 59-72). London, England: Karnac.

Brakel, L. A. W., Kleinsorge, S., Snodgrass, M., Shevrin, H. & Arbor, A. (2000). The primary process and the unconscious: Experimental evidence supporting two psychoanalytic presuppositions. *International Journal of Psychoanalysis*, *81*(3), 553-569.

Breuer, J., & Freud, S. (1895). *Studies on hysteria (The standard edition of the complete psychological works of Sigmund Freud, Vol. III)*. London, England: The Hogarth Press and The Institute of Psycho-analysis.

Brown, T. (1820). *Lectures on the philosophy of the human mind*. Edinburgh, Scotland: James Ballantyne and Co.

Bucci, W. (1988). Converging evidence for emotional structures: Theory and method. In H. Dahl, H. Kächele, & H. Thomä (Eds.), *Psycho-Analytic process research strategies* (pp. 29-49). Heidelberg, Germany: Springer-Verlag.

Bucci, W. (1989). A reconstruction of Freud's Tally Argument: A program for psychoanalytic Research. *Psychoanalytic Inquiry: A topical journal for mental health professionals*, *9*(2), 249–281.

Bucci, W. (1997). *Psychoanalysis and cognitive science: A Multiple Code Theory*. New York, NY: The Guilford Press.

Bucci, W. (2021). *Emotional communication and therapeutic change: Understanding psychotherapy through Multiple Code Theory*. (W. F. Cornell, Ed.). London, England and New York, NY: Routledge Taylor & Francis Group.

Burian, R. M. (1977). More than a marriage of convenience: On the inextricability of history and philosophy of science. *Philosophy of Science*, *44*, 1–42.

Caplan, A. (1986). Open peer commentary: With a friend like Professor Grünbaum, does psychoanalysis need any enemies? *The Behavioral and Brain Sciences*, *9*, 228–229.

Cardeña, E., & Terhune, D. B. (2019). The roles of response expectancies, baseline experiences, and hypnotizability in spontaneous hypnotic experiences. *International Journal of Clinical and Experimental Hypnosis*, *67*(1), 1–27.

Caropreso, F. & Simanke, R. T. (2011). *Entre o corpo e a consciência: Ensaios de interpretação da metapsicologia freudiana*. São Carlos, Brazil: EdUFSCar.

Carrol, L. (2014). *Alice's adventures in Wonderland*. London, England: Macmillan Children's Books.

Caws, P. (1986). Open peer commentary: The scaffolding of psychoanalysis. *The Behavioral and Brain Sciences*, *9*, 229–230.

Chapman, L. J. & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, *72*, 193-204.

Chapman, L. J. & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*, 271-280.

Cioffi, F. (1970). Freud and the idea of a pseudo-science. In R. Borger & F. Cioffi (Eds.), *Explanation in the behavioural sciences*. Cambridge University Press.

Cioffi, F. (1986). Did freud rely on the Tally Argument to meet the argument from suggestibility? *The Behavioral and Brain Sciences*, *9*, 230–231.

Clarke, S. (2015). *Jonathan Strange & Mr. Norrel*. London, England: Bloomsbury.

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for Borderline Personality Disorder: A multiwave study. *American Journal of Psychiatry*, *164*, 922–928.

Cleland, C. E. (2001). Historical science, experimental science, and the scientific method. *Geology*, *29*(11), 987–990.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science*, *69*(3), 447–451.

Cleland, C. E. (2011). Prediction and explanation in historical natural science. *British Journal for the Philosophy of Science*, *62*, 551–582.

Coffin, T. E. (1941). *Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion*. (J. F. Dashiell, Ed.). Evanston, IL: The American Psychological Association.

Collingwood, R. (1946). *The idea of history*. Oxford, England: Oxford University Press.

Copi, I., Cohen, C., & McMahon, K. (2014). *Introduction to logic*. Essex, England: Pearson.

Cornelis, S., Desmet, M., Meganck, R., Cauwe, J., Inslegers, R., Willemsen, J., ... & Vandenbergen, J. (2017). Interactions between obsessional symptoms and interpersonal dynamics: An empirical single case study. *Psychoanalytic Psychology*, *34*(4), 446.

Coué, E. (1922). *Self mastery through conscious autosuggestion*. New York, NY: Malkan.

Dalbiez, R. (1941). *Psychoanalytical method and the doctrine of Freud (Vol. II: Discussion)*. London, England: Longmans, Green and Co..

Danner, H. G. (2014). *A thesaurus of English word roots*. Lanham, MD: Rowman & Littlefield.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. New York, NY: Cambridge University Press.

Del Toro, G. & Funke, C. (2019). *Pan's Labyrinth*. London, England: Bloomsbury.

De Pascalis, V., Chiaradia, C., & Carotenuto, E. (2002). The contribution of suggestibility and expectation to placebo analgesia phenomenon in an experimental setting. *Pain*, *96*(3), 393-402.

De Saussure, R. (1925). Remarque sur la technique de la psychanalyse freudienne. *L'Evolution Psychiatrique*, *1*, 37-54.

Derksen, A. A. (1992). Does the Tally Argument make Freud a sophisticated methodologist? *Philosophy of Science*, *59*(1), 75–101.

Desmet, M. (2013). Experimental versus naturalistic psychotherapy research: Consequences for researchers, clinicians, policy makers and patients. *Psychoanalytische Perspectieven*, *31*(1), 59-78.

Diamond, M., & Christian, C. (2011). A brief history of therapeutic action: Convergence, divergence, and integrative bridges. In M. Diamond & C. Christian (Eds.), *The second century of psychoanalysis: Evolving perspectives on therapeutic action* (pp. 3–20). London, England: Karnac.

Eagle, M. N. (2011). *From classical to contemporary psychoanalysis: A critique and integration*. New York, NY: Taylor & Francis.

Eagle, M. N. (2019). Complexities in the evaluation of the scientific status of psychoanalysis. In R Gipps & M Lacewing (Eds.), *The Oxford handbook of philosophy and psychoanalysis* (pp. 354–376). Oxford, England: Oxford University Press.

Edelson, M. (1977). Psychoanalysis as science. *Journal of Nervous and Mental Disease*, *165*, 1-28.

Edelson, M. (1984). *Hypothesis and evidence in psychoanalysis*. Chicago, IL: The University of Chicago Press.

Edelson, M. (1986). Open peer commentary: The evidential value of the psychoanalyst's clinical data. *The Behavioral and Brain Sciences*, *9*, 232–234.

Edelson, M. (1988). *Psychoanalysis: A theory in crisis*. Chicago, IL and London, England: The University of Chicago Press.

Ellenberger, H. F. (1970). *The discovery of the unconscious: The history and evolution of dynamic psychiatry*. London, England: Fontana Press/Harper Collins Publishers.

Elsey, J. W., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, *144*(8), 797–848.

Emde, R. N. (1980). Toward a psychoanalytic theory of affect. In S. E. Greenspan & G. H. Pollock (Eds.), *The course of life: Psychoanalytic contributions toward understanding personality development (Vol. 1: Infancy and early childhood)*. NIMH

Ende, M. (2008). *Die Zauberschule und andere Geschichten*. Stuttgart, Germany: Thienemann-Esslinger Verlag GmbH.

Erdelyi, M. H. (1986). Open peer commentary: Psychoanalysis has a wider scope than the retrospective discovery of etiologies. *The Behavioral and Brain Sciences*, *9*, 234–235.

Erwin, E. (1986). Open peer commentary: Defending Freudianism. *The Behavioral and Brain Sciences*, *9*, 235–236.

Erwin, E. (1996). *A final accounting: Philosophical and empirical issues in Freudian psychology*. Cambridge, MA: The MIT Press.

Etchegoyen, R. H. (2005). *Fundamentals of psychoanalytic technique: Revised edition*. London, England: Karnac.

Evans, F. J. (1989). The independence of suggestibility, placebo response, and hypnotizability. In V. A. Gheorghiu, P. Netter, H. J. Eysenck and R. Rosenthal (Eds.), *Suggestion and suggestibility: Theory and research* (pp. 145-154). Berlin, Germany: Springer-Verlag.

Eysenck, H. J. (1947). *Dimensions of personality*. London, England: Routledge & Kegan Paul.

Eysenck, H. J. (1986). Open peer commentary: Failure of treatment – failure of theory? *The Behavioral and Brain Sciences*, *9*, 236.

Eysenck, H. J., & Furneaux, W. D. (1945). Primary and secondary suggestibility: An experimental and statistical study. *Journal of Experimental Psychology*, *35*, 483-503.

Eysenck, H.J., Arnold, W.J., & Meili, R. (1975). *Encyclopedia of psychology* (Vol. 2). Bungay, UK: Fontana.

Farrel, B. A. (1986). Open peer commentary: The probative value of the clinical data of psychoanalysis. *The Behavioral and Brain Sciences*, *9*, 236-237.

Ferretti, M. G. (2014). *Ontogênese e filogênese em Freud: Uma visão de conjunto* [Doctoral Dissertation, Universidade Estadual de Campinas]. FAPESP Repository https://bv.fapesp.br/en/publicacao/135476/ontogeny-and-phylogeny-in-freud-an-overview/

Fillion, T. J. & Blass, E. M. (1986). Infantile experience with suckling odors determines adult sexual behaviour in male rats. *Science*, *231*(4739), 729-731.

Fine, A. & Forbes, M. (1986). Open peer commentary: Grünbaum on Freud: Three grounds for dissent. *The Behavioral and Brain Sciences*, *9*, 237-238.

Fisher, S. & Greenberg, R. P. (1985) *The scientific credibility of Freud's theories and therapy (paperback edition)*. Columbia University Press.

Fisher, S. & Greenberg, R. P. (Eds.) (1978). *The scientific evaluation of Freud's theories and therapy*. Basic Books.

Fisher, S., & Greenberg, R. R. (1977). *The scientific credibility of Freud's theory and therapy*. New York, NY: Basic Books.

Fiske, S. T. & Taylor, S. E. (1991). *Social cognition*. New York, NY: McGraw-Hill.

Flavell, L. & Flavell, R. (1995). *Dictionary of word origins*. London, England: Kyle Books.

Flax, J. (1981). Psychoanalysis and the philosophy of science: Critique or resistance? *The Journal of Philosophy*, *78*(10), 561–569.

Fletcher, G. (1995). *The scientific credibility of folk psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fonagy, P. (2015). The effectiveness of psychodynamic psychotherapies: An update. *World Psychiatry*, *14*(2), 137-150.

Fonagy, P., & Target, M. (2000). The place of psychodynamic theory in developmental psychopathology. *Development and Psychopathology*, *12*, 407–425.

Forel, A. (1906). *Hypnotism, or Suggestion and psychotherapy: A study of the psychological, psycho-physiological and therapeutic aspects of hypnotism*. New York, NY; London, England: Rebman.

Fotopoulou, A. (2012). The history and progress of neuropsychoanalysis. In A. Fotopoulou, D. Pfaff and M. A. Conway (Eds.), *From the couch to the lab: Trends in psychodynamic neuroscience* (pp. 12-24). Oxford, England: Oxford University Press.

Fraley, R. C., & Marks, M. J. (2010). Westermarck, Freud, and the incest taboo: Does familial resemblance activate sexual attraction? *Personality and Social Psychology Bulletin*, *36*(9), 1202–1212.

Frank, G. (2000). The status of psychoanalytic theory today: There is an elephant there. *Psychoanalytic Psychology*, *17*, 174–179.

Freud, S. (1896/1966). Letter 52 from extracts from the Fliess papers. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud,* *Volume I (1886–1899): Pre-psycho-analytic publications and unpublished drafts* (pp. 233–239). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1915-1917). *Introductory lectures on psycho-analysis (standard edition, Vols. 15 & 16)*. London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1925). *An autobiographical study (standard edition, Vol. 20)*. London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1953a). On psychotherapy. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume VII (1901-1905): A case of hysteria, three essays on sexuality and other works* (pp. 257-270). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1953b). Three essays on the theory of sexuality. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume VII (1901-1905): A case of hysteria, three essays on sexuality and other works* (pp. 123-245). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1955a). Beyond the pleasure principle. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XVIII (1920-1922): Beyond the pleasure principle, group psychology and other works* (pp. 3-64). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1955b). From the history of an infantile neurosis. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XVII (1917-1919): An infantile neurosis and other works* (pp. 3-123). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1955c). Group psychology and the analysis of the ego. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XVIII (1920-1922): Beyond the pleasure principle, group psychology and other works* (pp. 67-143). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1955d). Notes upon a case of obsessional neurosis. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume X (1909): Two case histories ('Little Hans' and the 'Rat Man')* (pp. 151-318). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1955e). Two encyclopaedia articles. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XVIII (1920-1922):*

Beyond the pleasure principle, group psychology and other works (pp. 235-262). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1958). Observations on transference love (Further recommendations on the technique of psycho-analysis III). In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XII (1911-1913): The case of Schreber, papers on technique and other works* (pp. 179-173). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1960). *The standard edition of the complete psychological works of Sigmund Freud, Volume VI (1901): The psychopathology of everyday life*. (J. Strachey, Ed.). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1961a). The dissolution of the Oedipus Complex. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XIX (1923-1925): The ego and the id and other works* (pp. 171-179). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1961b). The infantile genital organization: An interpolation into the theory of sexuality. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XIX (1923-1925): The ego and the id and other works* (pp. 139-145). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1961c). Some psychical consequences of the anatomical distinction between the sexes. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XIX (1923-1925): The ego and the id and other works* (pp. 241-258). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1962a). A reply to criticisms of my paper on anxiety neurosis. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume III (1893-1899): Early psycho-analytic publications* (pp. 123-139). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1962b). The aetiology of hysteria. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume III (1893-1899): Early psycho-analytic publications* (pp. 187-221). London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. (1963a). Analytic therapy. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XVI (1916-1917):*

Introductory lectures on psycho-analysis (Part III) (pp. 448-463). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1963b). The sexual life of human beings. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XVI (1916-1917): Introductory lectures on psycho-analysis (Part III)* (pp. 303-319). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Freud, S. (1966). Preface to the translation of Bernheim's Suggestion. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume I (1886-1899): Pre-psycho-analytic publication and unpublished drafts* (pp. 73–87). London, England: The Hogarth Press and The Institute of Psycho-analysis.

Freud, S. (1975). Fetishism. In J. Strachey (Trans., Ed.), *The standard edition of the complete psychological works of Sigmund Freud, Volume XXI (1927-1931): The future of an illusion, civilization and its discontents and other works* (pp. 149-158). London, England: The Hogarth Press and The Institute of Psycho-Analysis.

Frosh, S. (2006). *For and against psychoanalysis* (2nd ed.). London, England and New York, NY: Routledge.

Frosh, S., & Baraitser, L. (2008). Psychoanalysis and psychosocial studies. *Psychoanalysis, Culture & Society*, *13*, 346–365.

Fulgencio, L. (2003). Kant e as especulações metapsicológicas em Freud. *Natureza Humana*, *5*(1), 129-173.

Fulgencio, L. (2013). Pode haver uma ciência psicanalítica sem uma metapsicologia especulativa? *Scientiae Studia*, *11*(3), 491-510.

Gabbard, G. O. (1999). Classic article: The nature of the therapeutic action of psycho-analysis (Introduction). *Journal of Psychotherapy Practice and Research*, *8*(1), 64–65.

Gabbard, G. O., & Westen, D. (2003). Rethinking therapeutic action. *International Journal of Psychoanalysis*, *84*, 823–841.

Gardiner, P. (1961). *The nature of historical explanation*. Oxford, England: Oxford University Press.

Garfinkel, A. (1981). *Forms of explanation*. New Haven, CT: Yale University Press.

Gauld, A. & Shotter, J. (1986). Open peer commentary: Warranting interpretations. *The Behavioral and Brain Sciences*, *9*, 239-240.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, *15*, 183–217.

Gheorghiu, V. (1989a). The development of research on suggestibility: Critical considerations. In V. A. Gheorghiu, P. Netter, H. J. Eysenck, & R. Rosenthal (Eds.), *Suggestion and suggestibility* (pp. 3-56). Berlin, Germany: Springer-Verlag.

Gheorghiu, V. (1989b). The difficulty in explaining suggestion: Some conceivable solutions. In V. A. Gheorghiu, P. Netter, H. J. Eysenck, & R. Rosenthal (Eds.), *Suggestion and suggestibility* (pp. 99–112). Berlin, Germany: Springer-Verlag.

Gheorghiu, V. A., Netter, P., Eysenck, H. J., & Rosenthal, R. (Eds.). (1989). *Suggestion and suggestibility: Theory and research*. Berlin, Germany: Springer-Verlag.

Gibbons, F. X., & McCoy, S. B. (1991). Self-perception and self-deception. In J. Schumaker (Ed.), *Human suggestibility* (pp. 185–199). New York, NY: Routledge.

Giere, R. N. (1973). History and philosophy of science: Intimate relationship or marriage of convenience? *The British Journal for the Philosophy of Science*, *24*(3), 282–297.

Giere, R. N. (2012). History and philosophy of science: Thirty-five years later. In S. Mauskopf & T. Schmaltz (Eds.), *Integrating history and philosophy of science: Problems and prospects* (pp. 59–65). Dordrecht, The Netherlands: Springer.

Glover, E. (1952). Research methods in psycho-analysis. *The International Journal of Psychoanalysis*, *33*, 403–409.

Glymour, C. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Glymour, C. (1982). Freud, Kepler, and the clinical evidence. In R. Wollheim & J. Hopkins (Eds.), *Philosophical essays on Freud* (pp. 12–31). Cambridge, England: Cambridge University Press.

Glymour, C. (1983). The theory of your dreams. In R. S. Cohen, & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis (Boston studies in the philosophy of science, 76)* (pp. 57–71). Dordrecht, The Netherlands: D. Reidel Publishing Company.

Green, A. (2005). The illusion of common ground and mythical pluralism. *International Journal of Psychoanalysis*, *86*, 627–632.

Greenberg, R. P. (1986). Open peer commentary: The case against Freud's cases. *The Behavioral and Brain Sciences*, *9*, 240-241.

Grünbaum, A. (1959). Remarks on Dr. Kubie's views. In: S. Hook (Ed.), *Psychoanalysis, scientific method, and philosophy*. New York, NY: New York University Press.

Grünbaum, A. (1980). Epistemological liabilities of the clinical appraisal of psychoanalytic theory. *Noûs*, *14(3)*, 307-385.

Grünbaum, A. (1981). The placebo concept. *Behaviour Research and Therapy*, *19*(2), 157–167.

Grünbaum, A. (1983). Retrospective vs. prospective testing of aetiological hypothesis in Freudian theory. In J. Earman (Ed.), *Testing scientific theories* (pp. 315–347). Minneapolis, MN: University of Minnesota Press.

Grünbaum, A. (1984). *The foundations of psychoanalysis: A philosophical critique*. Berkeley, CA: University of California Press.

Grünbaum, A. (1986a). Précis of "The Foundations of Psychoanalysis: A Philosophical Critique"; Author's response: Is Freud's Theory well-founded? *The Behavioral and Brain Sciences*, *9*, 217–284.

Grünbaum, A. (1986b). The placebo concept in medicine and psychiatry. *Psychological Medicine*, *16*, 19–38.

Grünbaum, A. (1989). Why thematic kinships between events do not attest their causal linkage. In R. S. Cohen (Ed.), *An intimate relation: Studies in the history and philosophy of science* (pp. 477–494). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Grünbaum, A. (1993). *Validation in the clinical theory of psychoanalysis: A study in the philosophy of psychoanalysis*. Madison, CT: International Universities Press.

Grünbaum, A. (2007). The reception of my Freud-critique in the psychoanalytic literature. *Psychoanalytic Psychology*, *24*(3), 545-576.

Grünbaum, A. (2008). Popper's fundamental misdiagnosis of the scientific defects of Freudian psychoanalysis, and of their bearing on the theory of demarcation. *Psychoanalytic Psychology*, *25(4)*, 574-589.

Grünbaum, A. (2015). Critique of psychoanalysis. In S. Boag, L. A. W. Brakel, & V. Talvitie (Eds.), *Philosophy, science, and psychoanalysis: A critical meeting* (pp. 1–36). London, England: Karnac.

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, *5*(3), 303–314.

Gudjonsson, G. H. (1987). Historical background to suggestibility: How interrogative suggestibility differs from other types of suggestibility. *Pers. Individ. Diff.*, *8*(3), 347-355.

Gudjonsson, G. H. (1989). Theoretical and empirical aspects of interrogative suggestibility. In V. A. Gheorghiu, P. Netter, H. J. Eysenck, & R. Rosenthal (Eds.), *Suggestion and suggestibility: Theory and research* (pp. 135–144). Berlin, Germany: Springer-Verlag.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. New York, NY: John Wiley & Sons.

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1*(2), 83–104.

Habermas, J. (1968). *Knowledge and human interests*. Cambridge, England: Polity Press.

Hacking, I. (1975). *The emergence of probability*. New York, NY: Cambridge University Press.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. New York, NY: Cambridge University Press.

Hahn, R. A. (1997). The nocebo phenomenon: Concept, evidence and implications for public health. *Preventive Medicine*, *26*, 607–611.

Halligan, P. W., & Oakley, D. A. (2014). Hypnosis and beyond: Exploring the broader domain of suggestion. *Psychology of Consciousness: Theory, Research and Practice*, *1*(2), 105–122.

Hanly, C. M. T. (2013). Foreword. In R. D. Hinshelwood, *Research on the couch: Single-case studies, subjectivity and psychoanalytic knowledge* (pp. xi-xiii). Hove, England: Routledge.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge, England: Cambridge University Press.

Hanson, N. R. (1971). *Observation and explanation: A guide to philosophy of science*. New York, NY: Harper Torchbook.

Hempel, C. (1965). *Aspects of scientific explanation*. New York, NY: Free Press.

Hinshelwood, R. D. (2013). *Research on the couch: Single-Case studies, subjectivity and psychoanalytic knowledge*. Hove, England: Routledge.

Hitchcock, C. (1998). The common cause principle in historical linguistics. *Philosophy of Science*, *65*(3), 425–447.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*, 687–701.

Holmes, J., & Nolte, T. (2019). "Surprise" and the Bayesian brain: Implications for psychotherapy theory and practice. *Hypothesis and Theory*, *10*, 1-13

Holt, R. R. (1986). Open peer commentary: Some reflections on testing psychoanalytic hypotheses. *The Behavioral and Brain Sciences*, *9*, 242-244.

Hook, S. (Ed.). (1959). *Psychoanalysis, scientific method, and philosophy*. New York, NY: New York University Press.

Hoover, K. D. (1990). The logic of causal inference. *Economics and Philosophy*, *6*, 207–234.

Hopkins, J. (1988). Epistemology and depth psychology: Critical notes on "The foundations of psychoanalysis." In P. Clark & C. Wright (Eds.), *Mind, psychoanalysis and science* (pp. 33–60). Oxford, England: Blackwell.

Huemer, M. (2015). The failure of analysis and the nature of concepts. In C. Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 51–76). London, England: Palgrave Macmillan.

Hughes, S. (1964). *History as art and as science*. New York, NY: Harper and Row.

Janet, P. (1919). *Les médications psychologiques: Études historiques, psychologiques et cliniques sur les méthodes de la psychotherapie. I: L'action morale, l'utilisation de l'automatisme*. Paris, France: Librairie Félix Alcan.

Jones, E., & V. Harris. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24

Kächele, H. (1986). Open peer commentary: Validating psychoanalysis: What methods for what task? *The Behavioral and Brain Sciences*, *9*, 244-245.

Kächele, H., Schachte, J., & Thomä, H. (2009). *From psychoanalytic narrative to empirical single case research: Implications for psychoanalytic method*. New York, NY: Routledge.

Kaptchuk, T. J. (2018). Open-label placebo: Reflections on a research agenda. *Perspectives in Biology and Medicine*, *61*(3), 311–334.

Kaszubowski, E. (2016). *Modelo de tópicos para associações livres* [Doctoral dissertation, Universidade Federal de Santa Catarina]. UFSC Repository.

Keat, R. (1981). *The politics of social theory: Habermas, Freud and the critique of positivism*. Chicago, IL: The University of Chicago Press.

Kelley, H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on motivation* (pp. 192–238). Lincoln, NE: University of Nebraska Press.

Kelley, H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128.

Kertscher, T. (2020, April 10). *Fact-check: Has a pandemic occurred every 100 years?* Statesmen. https://www.statesman.com/news/20200410/fact-check-has-pandemic-occurred-every-100-years

Kihlstrom, J. F. (2008). The domain of hypnosis, revisited. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis* (pp. 21–52). Oxford, England: Oxford University Press.

Kirsch, I. (2005). Placebo psychotherapy: Synonym or oxymoron? *Journal of Clinical Psychology*, *61*(7), 791–803.

Kirsch, I., Cardeña, E., Derbyshire, S., Dienes, Z., Heap, M., Kallio, S.,... Whalley, M. (2011). Definitions of hypnosis and hypnotizability and their relation to suggestion and suggestibility: A consensus statement. *Contemporary Hypnosis and Integrative Therapy*, *28*, 107–115.

Klerman, G. L. (1986). Open peer commentary: The scientific tasks confronting psychoanalysis. *The Behavioral and Brain Sciences*, *9*, 245.

Kline, P. (1972). *Fact and fantasy in Freudian theory* (1st ed.). Methuen.

Kline, P. (1981). *Fact and fantasy in Freudian theory* (2st ed.). Methuen.

Krech, D., & Crutchfield, R. S. (1948). *Theory and problems of social psychology*. New York, NY: McGraw-Hill.

Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th ed.). Chicago, IL and London, England: The University of Chicago Press.

Kukla, A. (2001). *Methods of theoretical psychology*. Cambridge, MA: The MIT Press.

Lacan, J. (2007). The direction of the treatment and the principles of its power. In B. Fink, H. Fink, & R. Grigg (Eds.), *Écrits: The first complete edition in English* (pp. 489–542). New York, NY and London, England: W. W. Norton & Company

Lacewing, M. (2012a). Inferring motives in psychology and psychoanalysis. *Philosophy, Psychiatry and Psychology*, *19*(3), 197–212.

Lacewing, M. (2012b). Statistics, desire, and interdisciplinarity. *Philosophy, Psychiatry and Psychology*, *19*(3), 221–225.

Lacewing, M. (2013a). Could psychoanalysis be a science? In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (pp. 1103–1127). Oxford, England: Oxford University Press.

Lacewing, M. (2013b). The problem of suggestion in psychoanalysis: An analysis and solution. *Philosophical Psychology*, *26*(5), 718-743.

Lacewing, M. (2014a). Emotions and the virtue of self-understanding. In S. Roeser & C. Todd (Eds.), *Emotion and value*. Oxford, England: Oxford University Press.

Lacewing, M. (2014b). Psychodynamic psychotherapy, insight, and therapeutic action. *Clinical Psychology: Science and Practice*, *21*(2), 154–171.

Lacewing, M. (2018). The science of psychoanalysis. *Philosophy, Psychiatry and Psychology*, *25*(2), 95–111.

Lacewing, M. (2019). A truthful way to live? Objectivity, ethics and psychoanalysis. *Royal Institute of Philosophy Supplement*, *85*, 175–193.

Lakatos, I., & Feyerabend, P. (1999). *For and against method: Including Lakato's lectures on scientific method and the Lakatos-Feyerabend correspondence*. (M. Motterlini, Ed.). Chicago, IL and London, England: The University of Chicago Press.

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 169–218). New York, NY: John Wiley & Sons.

Lane, R. D. (2020). Memory reconsolidation, emotional arousal and the process of change in psychoanalysis. In M. Leuzinger-Bohleber, M. Solms, & S. E. Arnold (Eds.), *Outcome research and the future of psychoanalysis: Clinicians and researchers in dialogue* (pp. 188–205). London, England and New York, NY: Routledge Taylor & Francis Group.

Lane, R. D., & Nadel, L. (Eds.). (2020). *Neuroscience of enduring change: Implications for psychotherapy*. New York, NY: Oxford University Press.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, 38, 1–19.

Lane, R. D., Subic-Wrana, C., Greenberg, L., & Yovel, I. (2020, September 17). The role of enhanced emotional awareness in promoting change across psychotherapy modalities. *Journal of Psychotherapy Integration*. Advance online publication. http://dx.doi.org/ 10.1037/int0000244

Langer, W. (1958). The next assignment. *The American Historical Review*, *63*, 283-304.

Laplanche, J., & Pontalis, J.-B. (1988). *The language of psychoanalysis*. London, England: Karnac.

Larède, J. (1980). *Qu'est-ce que la suggestologie?* Toulouse, France: Privat.

Latour, B. & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.

Laudan, L. (1978). *Progress and its problems: Toward a theory of scientific growth*. Berkeley, CA: University of California Press.

Laudan, R. (1992). What's so special about the past? In M. H. Nitecki & D. V. Nitecki (Eds.), *History and evolution* (pp. 55–67). Albany, NY: State University of New York Press.

Leavy, S. (1980). *The psychoanalytic dialogue*. New Haven, CT: Yale University Press.

Lebrun, G. (1977/2006). A ideia de epistemologia. In: G. Lebrun, *A filosofia e sua história*. São Paulo, Brazil: Cosacnaify.

Lee, S. H., Blake, R. & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nat. Neurosci.*, *8*, 22–23.

Leeds, J. (1988). *Repetitive structures — Theoretical considerations and an empirical study* [Unpublished Doctoral Dissertation]. Adelphi University.

Leichsenring, F., & Klein, S. (2020). Evidence for psychodynamic psychotherapy in specific mental disorders. In M. Leuzinger-Bohleber, M. Solms, & S. E. Arnold (Eds.), *Outcome research and the future of psychoanalysis: Clinicians and researchers in dialogue* (pp. 99–127). London, England and New York, NY: Routledge Taylor & Francis Group.

Leichsenring, F., Klein, S., & Salzer, S. (2014). The efficacy of psychodynamic psychotherapy in specific mental disorders: A 2013 update of empirical evidence. *Contemporary Psychoanalysis*, *50*(1–2), 89–130.

Levy, S., & Inderbitzin, L. (2000). Suggestion and psychoanalytic technique. *Journal of the American Psychoanalytic Association*, *48*, 739–758.

Lewin, K. (1946). Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of Child Psychology*. New York, NY: Wiley.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*(4), 455–476.

Lewis, D. (1986). Causal explanation. In D. Lewis, *Philosophical Papers (Vol. II)* (pp. 214-240). New York, NY: Oxford University Press.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London, England and New York, NY: Routledge Taylor & Francis Group.

Loewald, H. (1977). *Psychoanalysis and the history of the individual*. New Haven, CT: Yale University Press.

Lohmeier, J. H. (2010). Nonexperimental designs. In N. J. Salkind (ed.), *Encyclopedia of research designs*. Los Angeles, CA: SAGE Publications.

Lovecraft, H. P. (2016). The colour out of space. In H. P. Lovecraft, *The complete fiction of H. P. Lovecraft* (pp. 637-661). New York, NY: Chartwell.

Luborsky, L. (1986). Open peer commentary: Evidence to lessen professor Grunbaum's concern about Freud's clinical inference method. *The Behavioral and Brain Sciences*, *9*, 247-249.

Luborsky, L., Singer, B., & Luborsky, E. (1975). Comparative studies of psychotherapies: Is it true that "everybody has won and all must have prizes"? *Archives of General Psychiatry*, *32*, 995–1008.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT press.

Lundh, L.-G. (1998). Normal suggestion. An analysis of the phenomenon and its role in psychotherapy. *Clinical Psychology and Psychotherapy*, *5*, 24–38.

Lundh, L.-G. (2014). The search for common factors in psychotherapy: Two theoretical models with different empirical implications. *Psychology and Behavioral Sciences*, *3*(5), 131–150.

Luyten, P., Blatt, S. J., & Corveleyn, J. (2006). Minding the gap between positivism and hermeneutics in psychoanalytic research. *Journal of the American Psychoanalytic Association*, *54*(2), 571–610.

Lynch, K. (2014). The vagaries of psychoanalytic interpretation: An investigation into the causes of the consensus problem in psychoanalysis. *Philosophia (United States)*, *42*, 779–799.

Mackay, N. (1989). *Motivation and explanation*. Madison, CT: International Universities Press.

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quaterly*, *2*(4), 245–264.

Macnab, G. (2011, October 23). *Lars Von Trier: 'If I am an idiot in the eyes of the world, so be it'*. Independent. https://www.independent.co.uk/news/people/profiles/lars-von-trier-if-i-am-idiot-eyes-world-so-be-it-2346972.html

Marinho, N. C. (2006). *Razão e psicanálise: "O caso Schreber (Freud, 1911)", revisitado a partir das contribuições de Marcia Cavell e Ludwig Wittgenstein* [Doctoral dissertation, Pontifícia Universidade Católica do Rio de Janeiro]. BDTD. https://bdtd.ibict.br/vufind/Record/PUC_RIO-1_8580e0304dfa8f25305a90c320728097

Marmor, J. (1970). Limitations of free association. *Archives of General Psychiatry*, *22*, 160-165.

Marmor, J. (1974b). Psychoanalytic therapy as an educational process. In *Psychiatry in transition: Selected papers of Judd Marmor*. Brunner/Mazel.

Marmor, J. (1986). Open peer commentary: The question of causality. *The Behavioral and Brain Sciences*, *9*, 249.

Martins, E. de C. (2012). *Freud e os modelos biológicos de explicação* [Doctoral dissertation, Universidade Federal de São Carlos]. UFSCar Repository. https://repositorio.ufscar.br/handle/ufscar/4790

Masling, J. & Schwartz, M. (1979). A critique of research in psychoanalytic theory. *Cenetic Psychology Monographs*, *100*, 257-307.

Masling, J. (1983). *Empirical studies of psychoanalytical theories (Vol. 1)*. Analytic Press.

Masling, J. (1986). Open peer commentary: Psychoanalysis, case histories, and experimental data. *The Behavioral and Brain Sciences*, *9*, 249-250.

McDougall, W. (1908). *Introduction to social psychology*. London, England: Methuen.

McDougall, W. (1920). A note on suggestion. *The Journal of Neurology and Psychopathology*, *1*(1), 1–10.

McDougall, W. (2001). *An introduction to social psychology* (14th ed.). Kitchener, Canada: Batoche.

McMullin, E. (1970). The history and philosophy of science: A taxonomy. In *Minnesota studies in the philosophy of science (Vol. V)* (pp. 12–67). Minneapolis, MN: University of Minnesota Press.

Meganck, R., Inslegers, R., Krivzov, J., & Notaerts, L. (2017). Beyond clinical case studies in psychoanalysis: A review of psychoanalytic empirical single case studies published in ISI-ranked journals. *Frontiers in Psychology*, *8*(1749).

Melden, A. I. (1969). Historical objectivity. In R. Nash (Ed.), *Ideas in history (Vol. 2)*. New York, NY: E. P. Dutton.

Messer, S. B., & Wolitzky, D. L. (2010). A psychodynamic perspective on the therapeutic alliance: Theory, research, and practice. In J. C. Muran & J. P. Barber (Eds.), *Therapeutic alliance: An evidence-based guide to practice* (pp. 97–122). New York, NY: Guilford Press.

Meyerhoff, H. (1959). Introduction. In H. Meyerhoff (Ed.), *The philosophy of history in our time*. New York, NY: Anchor.

Meyerhoff, H. (1962). On psychoanalysis as history. *Psychoanalytic Review*, *49*, 3-20.

Mezan, R. (2006). Pesquisa em psicanálise: Algumas reflexões. *Jornal de Psicanálise*, *39*(70), 227-241.

Michael, M. T. (2008). On the validity of Freud's dream interpretations. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *39*, 52–64.

Michael, M. T. (2015). *Freud's theory of dreams: A philosophico-scientific perspective*. Lanham, MD: Rowman & Littlefield.

Michael, M. T. (2018). Why aren't more philosophers interested in Freud? Re-evaluating philosophical arguments against psychoanalysis. *Philosophia*, *47*, 959-976.

Michael, M. T. (2019). The case for the Freud–Breuer theory of hysteria: A response to Grünbaum's foundational objection to psychoanalysis. *The International Journal of Psychoanalysis*, *100*(1), 32–51.

Michael, R. B., Garry, M., & Kirsch, I. (2012). Suggestion, cognition and behavior. *Current Directions in Psychological Science*, *21*, 151–156.

Milrod, B., Leon, A.C., Busch, F., Rudden, M., Schwalberg, M., Clarkin, J., … Shear, M.K. (2007). A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. *American Journal of Psychiatry*, *164*, 265–272.

Mitchell, S. A. (1997). *Influence and autonomy in psychoanalysis*. London and New York: Routledge.

Moernaut, N., Vanheule, S., & Feyaerts, J. (2018). Content matters, a qualitative analysis of verbal hallucinations. *Frontiers in Psychology*, *9*(1958).

Munby, M., & Johnston, D. W. (1980). Agoraphobia: The long-term follow-up of behavioural treatment. *The British Journal of Psychiatry*, *137*, 418-427.

Nadel, L., Samsonovich, A., Ryan L., and Moscovitch M. (2000). Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352–368.

Nagel, E. (1959). Methodological issues in psychoanalytic theory. In S. Hook (Ed.), *Psychoanalysis, scientific method and philosophy: A symposium* (pp. 38-56). New York, NY: New York University Press.

Nagel, E. (1961). *The structure of science*. New York, NY: Routledge & Kegan Paul.

Nagel, E. (1963). Relativism and some problems of working historians. In S. Hook (Ed.), *Philosophy and history*. New York, NY: New York University Press.

Novey, S. (1968). *The second look: The reconstruction of personal history in psychiatry and psychoanalysis*. Baltimore, MD: Johns Hopkins University Press.

Novick, K. K. & Novick, J. (2002). Reclaiming the land. *Psychoanalytic Psychology*, *19*, 348–377.

Oakley, D. A., & Halligan, P. W. (2009). Hypnotic suggestion and cognitive neuroscience. *Trends in Cognitive Sciences*, *13*(6), 264–270.

Osgood, C. E., May, W., & Miron, M. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Oxford. (n. d.). Suggest. In *Lexico*. Retrieved October 27, 2021 from https://www.lexico.com/definition/suggest

Padilha, J. (Director). (2007). *Tropa de Elite (Elite Squad)* [Film]. Zazen Produções; Universal Pictures.

Padovan, C. (2018). *Les origines de la méthode psychanalytique: Une étude d'histoire conceptuelle* [Doctoral dissertation, Université Paris Diderot (Paris 7)]. Theses. http://www.theses.fr/2018USPCC092

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart, & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Partridge, E. (2006). *Origins: A short etymological dictionary of modern English* (4th ed.). London, England and New York, NY: Routledge Taylor & Francis Group.

Passmore, J. (1967). Logical positivism. In P. Edwards (Ed.). *The encyclopedia of philosophy (Vol. 5)* (pp. 52- 57). New York, NY: Macmillan.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge, England: Cambridge University Press.

Peirce, C. S. (1957). The general theory of probable inference. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 190–217). New York, NY: Dover.

Pekala, R. J. (1991). *Quantifying consciousness: An empirical approach*. New York, NY: Plenum.

Pereira, M. E. C. (2003). *Psicopatologia dos ataques de pânico*. São Paulo, Brazil: Escuta.

Peterfreund, E. (1983). *The process of psychoanalytic therapy*. Hillsdale, MI: The Analytic Press.

Pièron, H. (1963). *Vocabulaire de la psychologie*. Paris, France: Presses Universitaires de France.

Pinker, S. (2015). *The sense of style: The thinking person's guide to writing in the 21st century*. London, England: Penguin.

Pinto, W. C. F. (2016). *Do círculo à espiral: Por uma história e método da recepção filosófica da psicanálise segundo o freudismo filosófico francês (Ricoeur) e a filosofia brasileira da psicanálise (Monzani)* [Doctoral dissertation, Universidade Estadual de Campinas].

PiPS (2021). *Our team*. http://programinplacebostudies.org/about/people/

Plato (1997). *Plato: Complete works*. (J. M. Cooper & D. S. Hutchinson, Eds.). Indianapolis, IN: Hackett.

Politzer, G. (1928). *Critique des fondements de la psychologie: La psychologie et la psychanalyse*. Paris, France: Quadrige/PUF.

Popper, K. (1963). *Conjectures and refutations*. London, England: Routledge & Kegan Paul.

Porta dos Fundos (2012, November 8). *Baby's Name* [Video]. YouTube. https://www.youtube.com/watch?v=m1HnHVZHAgA

Pullman, P. (2003). *His dark materials*. New York, NY: Yearling.

Putnam, H. (2002). *The collapse of the fact-value dichotomy and other essays*. Cambridge, MA; London, England: Harvard University Press.

Quine, W. V. O. (2013). *Word and object*. Cambridge, MA and London, England: The MIT Press.

Rachman, S. J., & Wilson, G. T. (1980). The effects of psychological therapy (2nd enlarged ed.). New York, NY: Pergamon Press.

Rangell, L. (1988). The future of psychoanalysis: The scientific crossroads. *The Psychoanalytic Quarterly*, *57*, 313–340.

Rapaport, D. (1960). *The structure of psychoanalytic theory: A systematizing attempt* (Psychological Issues, Mono. 6). New York, NY: International University Press.

Rappaport, S. (1996). Inference to the best explanation: Is it really different from Mill's methods? *Philosophy of Science*, *63*(1), 65–80.

Reiser, M. F. (1986). Open peer commentary: Grünbaum's critique of clinical psychoanalytic evidence: A sheep in wolf's clothing? *The Behavioral and Brain Sciences*, *9*, 255-256.

Renik, O. (1998). The analyst's subjectivity and the analyst's objectivity. *International Journal of Psychoanalysis*, *79*, 487–97.

Ricoeur, P. (1970). *Freud and philosophy*. New Haven, CT: Yale University Press.

Ridge, M. (2019, Fall). *Moral non-naturalism*. The Stanford Encyclopedia of Philosophy. Retrieved October 19, 2021, from https://plato.stanford.edu/archives/fall2019/entries/moral-non-naturalism/

Ritvo, L. (1990). *Darwin's influence on Freud: A tale of two sciences*. New Haven, CT; London, England: Yale University Press.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, *3*, 377–415.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psycho-therapy. *American Journal of Orthopsychiatry, 6*(3), 412–415.

Rosenzweig, S. (1992). Freud and experimental psychology: The emergence of ldiodynamics. In S. Koch & D. E. Leary (Eds.). *A century of psychology as science* (pp. 135-207). American Psychological Association.

Rosenzweig, S., & Mason, G. (1934). An experimental study of memory in relation to the theory of repression. *British Tournai of Psychology (General Section)*, *24*, 247-265.

Rotaru, T.-Ş., & Dafinoiu, I. (2014). Attachment and suggestion-related phenomena. *International Journal of Clinical and Experimental Hypnosis*, *62*(2), 195–214.

Rothman, K. J. & Greenland, S. (1998). *Modern epidemiology (2nd edition)*. Philadelphia, PA: Lippincott Rawen.

Rothman, K. J. (1976). Causes. *American Journal of Epidemiology*, *104*, 587–592.

Rubinstein, B. B. (1975). On the clinical psychoanalytic theory and its role in the inference and confirmation of particular clinical hypotheses. *Psychoanalysis & Contemporary Science*, *4*, 3–57.

Rubinstein, B. B. (1980). On the psychoanalytic theory of unconscious motivation and the problem of its confirmation. *Noûs*, *14*(3), 427–442.

Rubovits-Seitz, P. F. D. (1998). *Depth-psychological understanding: The methodologic grounding of clinical interpretations*. Hillsdale, NJ; London, England: The Analytic Press.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. Bruce & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rumelhart, D. E., P. Smolensky, J. L. McClelland & G. E. Hinton (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart & the PDP Research Group (Eds.), *Parallel distributed processing (Vol. 2)* (pp. 7–57). Cambridge, MA: MIT Press.

Ruse, M. (1986). Open peer commentary: Grünbaum on psychoanalysis: Where do we go from here? *The Behavioral and Brain Sciences*, *9*, 256-257.

Sachs, D. (1989). In fairness to Freud: A critical notice of “The foundations of psychoanalysis”. *The Philosophical Review*, *98*(3), 349–378.

Sagan, C. (1980). *Cosmos*. New York, NY: Random House.

Sagan, C. (1997). *The demon-haunted world: Science as a candle in the dark*. London, England: Headline.

Sandler, J., Sandler, A.-M., & Davies, R. (Eds.). (2000). *Clinical and observational psychoanalytic research: Roots of a controversy*. Madison, CT: International Universities Press.

SCA (2021). *The single case archive project.* https://singlecasearchive.com/about

Schachter, H. (2003). *The serial killer files*. New York, NY: Ballantine Books.

Schafer, R. (1976). *A new language for psychoanalysis*. New Haven, CT: Yale University Press.

Schafer, R. (1978). *Language and insight*. New Haven, CT: Yale University Press.

Schepanck, H. (1974). *Erb- und Umweltfaktoren bei Neurosen: Tiefenpsychologische Untersuchungen an 50 Zwillingspaaren*. Springer Verlag.

Schepank, H. (1984). Das mannheimer Kohortenprojekt – Die Prävalenz psychogener Erkrankungen in der Stadt. *Zeitschrift fur Psychosomatische Medizin und Psychoanalyse*, *30*, 43-61.

Schmidl, F. (1962). Psychoanalysis and history. *Psychoanalytic Quarterly*, *31*, 532-548.

Schmidt, E. (2017). *Charcot e a escola de Salpêtrière: A afirmação de uma histeria neurológica* [Doctoral Dissertation, Universidade Federal de Juiz de Fora]. UFJF Repository https://repositorio.ufjf.br/jspui/handle/ufjf/6075

Schumaker, J. F. (1991). The adaptive value of suggestibility and dissociation. In J. F. Schumaker (Ed.), *Human suggestibility: Advances in theory, research, and application* (pp. 108–131). New York, NY: Routledge.

Schurz, G. (2014). *Philosophy of science: A unified approach*. New York, NY and London, England: Routledge.

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O.T., & Oitzl, M.S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1740–1749.

Schwanenberg, E. (1989). Suggestion as social biasing of meaning tests. In V. A. Gheorghiu, P. Netter, H. J. Eysenck, & R. Rosenthal (Eds.), *Suggestion and suggestibility* (pp. 263–278). Berlin, Germany: Springer-Verlag.

Seitz, P. F. (1966). The consensus problem in psychoanalytic research. In L. A. Gottschalk & A. H. Auerbach, *Methods of research in psychotherapy* (pp. 209-225). Boston, MA: Springer.

Shapiro, A. K., & Morris, L. A. (1978). The placebo effect in medical and psychological therapies. In S. L. Garfield (Ed.), *Handbook of psychotherapy and behavior change* (26th. ed.). New York, NY: Wiley.

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, *65*(2), 98–109

Sherif, M. (1936). *The psychology of social norms*. New York, NY and London, England: Harper & Brothers.

Sherwood, M. (1969). *The logic of explanation in psychoanalysis*. New York, NY; London, England: Academic Press.

Shevrin, H. (1986). Open peer commentary: An argument for the evidential standing of psychoanalytic data. *The Behavioral and Brain Sciences*, *9*, 257-259.

Shevrin, H., Bond, J., Hertel, R., Marshall, R., Williams, W., & Brakel, L. A. W. (1992). Event-related potential indicators of the dynamic unconscious. *Consciousness and Cognition*, *1*, 340-366.

Shevrin, H., Brakel, L. A. W., Snodgrass, M., Kushawa, R., & Kalaida, N. (2013). Alpha power and unconscious inhibition: Toward a causal relationship between unconscious conflict and conscious symptoms. *Frontiers in Human Neuroscience*, *7*(544).

Sidis, B. (1898). *The psychology of suggestion: A research into the subconscious nature of man and society*. New York, NY: D. Appleton and Company.

Silva, C. N. (2010). Ex post facto study. In N. J. Salkind (ed.), *Encyclopedia of research designs*. Los Angeles, CA: SAGE Publications.

Silverman, L. H. (1976). Psychoanalytic theory: The reports of my death are greatly exaggerated. *American Psychologist*, *31*, 621-637.

Simanke, R. T. & Caropreso, F. (2011). A metáfora psicológica de Sigmund Freud: Neurologia, psicologia e metapsicologia na fundamentação da psicanálise. *Scientiae Studia*, *9*(1), 51-78.

Simanke, R. T. (2002). *Metapsicologia lacaniana: Os anos de formação.* São Paulo, Brazil: Discurso Editorial; Curitiba, Brazil: UFPR.

Simanke, R. T. (2009). Realismo e antirrealismo na interpretação da metapsicologia freudiana. *Natureza Humana*, *11*(2), 97-152.

Simanke, R. T. (2013). Causalité et emancipation dans la cure psychanalytique. In M. Coelen, C. Nioche & B. Santos (Orgs.). *Jouissance et souffrance (Vol. 1)* (pp. 119-138). Paris, France: Campagne Première.

Smith, C., Cook, R., & Rohleder, P. (2017). 'When it comes to HIV, that's when you find out the genuinity of that love': The experience of disclosing a HIV+ status to an intimate partner. *Journal of Health Psychology*, *24*(8), 1011–1022.

Smith, M. L., Glass, G. V, and Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.

Sober, E. (1987). Parsimony, likelihood, and the principle of the common cause. *Philosophy of Science*, *54*(3) 465–469.

Sober, E. (2001). Venetian sea levels, British bread prices, and the principle of the common cause. *British Journal for the Philosophy of Science*, *52*(2), 331–346.

Socarides, C. (1970). Homosexuality and medicine. *Journal of the American Medical Association*, *212*, 1199-1202.

Spence, D. P. (1982). *Narrative truth and historical truth: Meaning and interpretation in psychoanalysis*. London, England and New York, NY: W. W. Norton & Company.

Spence, D. P. (1986). Open peer commentary: Are free associations necessarily contaminated? *The Behavioral and Brain Sciences*, *9*, 259.

Stokvis, B., & Pflanz, M. (1961). *Suggestion*. Stuttgart, Germany: Hippokrates.

Storr, A. (1986). Open peer commentary: Human understanding and scientific validation. *The Behavioral and Brain Sciences*, *9*, 259-260.

Strachey, J. (1999). Classic article: The nature of the therapeutic action of psycho-analysis. *Journal of Psychotherapy Practice and Research*, *8*(1), 66–82.

Strupp, H. H. (1986). Open peer commentary: Transference: One of Freud's basic discoveries. *The Behavioral and Brain Sciences*, *9*, 260-261.

Strupp, H. H., Hadley, S. W., and Gomes-Schwartz, B. (1977). *Psychotherapy for better or worse: The problem of negative effects*. New York, NY: Jason Aronson.

Stukat, K. G. (1958). *Suggestibility: A factorial and experimental analysis*. Stockholm, Sweden: Almqvist & Wiksell.

Sulloway, F. J. (1979). *Freud, biologist of the mind: Beyond the psychoanalytic legend.* Cambridge, MA; London, England: Harvard University Press.

Suppe, F. (1986). Open peer commentary: Grünbaum, homosexuality, and contemporary psychoanalysis. *The Behavioral and Brain Sciences*, *9*, 261-262.

Szpitalak, M., & Polczyk, R. (2015). Reinforced self-affirmation and interrogative suggestibility. *Psychiatry, Psychology and Law*, *23*(4), 512–520.

Thomas, F.-N., & Turner, M. (2011). *Clear and simple as the truth: Writing classic prose* (2nd ed.). Princeton, NJ: Princeton University Press.

Timpanaro, S. (1975). The Freudian slip. *New Left Review*, *91*, 43-56

Toulmin, S. (1961). *Foresight and understanding: An enquiry into the aims of science*. Bloomington, IN: Indiana University Press.

Tucker, A. (2004). *Our knowledge of the past: A philosophy of historiography*. Cambridge, England: Cambridge University Press.

Van Nieuwenhove, K., Truijens, F., Meganck, R., Cornelis, S., & Desmet, M. (2019). Working through childhood trauma-related interpersonal patterns in psychodynamic treatment: An evidence-based case study. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*(1), pp. 64-74.

Van Vleet, J. E. (2021). *Informal logical fallacies: A brief guide* (revised edition). London, England: Hamilton Books.

Verztman, J. S. (2009). A estratégia de estudo de casos múltiplos na pesquisa clínica em psicanálise. *Colóquio Internacional sobre o Método Clínico. Anais do Colóquio Internacional sobre o Método Clínico, São Paulo (Brazil), 1*, 5-10.

Verztman, J. S. (2021, March 29). *Currículo Lattes*. CNPq. http://lattes.cnpq.br/5049979430862547

Von Eckardt, B. (1986). Open peer commentary: Grunbaum's challenge to Freud's logic of argumentation: A reconstruction and an addendum. *The Behavioral and Brain Sciences*, *9*, 262-263.

Von Korff, P. (1987). *Referential activity and the therapeutic process* [Unpublished doctoral dissertation]. Adelphi University.

Wachtel, P. L. (1986). Open peer commentary: Early Freud, late Freud, conflict and intentionality. *The Behavioral and Brain Sciences*, *9*, 263-264.

Waelder, R. (1962). Psychoanalysis, scientific method, and philosophy. *Journal of the American Psychoanalytic Association*, *10*, 617–637.

Wagman, M. (2000). *Historical dictionary of quotations in cognitive science: A treasury of quotations in psychology, philosophy, and artificial intelligence*. Westport, CT; London, England: Greenwood Press.

Wallace, E. R. (1985). Historiography and causation in psychoanalysis: An essay on psychoanalytic and historical epistemology. Hillsdale, NJ: The Analytic Press and Lawrence Erlbaum Associates.

Wallace, E. R. (1989). Pitfalls of a one-sided image of science: Adolf Grünbaum's "Foundations of psychoanalysis". *Journal of the American Psychoanalytic Association*, *37*(2), 493–529.

Wallerstein, R. S. (1986). *Forty-two lives in treatment: A study of psychoanalysis and psychotherapy*. New York, NY: Guilford Press.

Wallerstein, R. S. (1995). *The talking cures*. New Haven, CT: Yale University Press.

Wallerstein, R. S. (2006). The relevance of Freud's psychoanalysis in the 21st century: Its science and its research. *Psychoanalytic Psychology*, *23*, 302-326.

Wallerstein, R. S. (2009). What kind of research in psychoanalytic science? *International Journal of Psychoanalysis*, *90*, 109–133.

Walsh, W. (1969). Positivist and idealist approaches to history. In R. Nash (Ed.), *Ideas in history (Vol. 2)*. New York, NY: E. P. Dutton.

Ward, P. D. (1990). The cretaceous/tertiary extinctions in the marine realm. In B. Sharpton & P. Ward (eds.), *Global catastrophes in Earth history* (pp. 425-432). Boulder, CO: Geological Society of America.

Watermeyer, B., Hunt, X., Swartz, L., & Rohleder, P. (2019). Navigating the relational psychic economy of disability: The case of M. *Psychoanalytic Dialogues*, *29*(5), 515-531.

Weeks, G. R. & L'Abate, L. (1982). *Paradoxical psychotherapy: Theory and practice with individuals, couples and families*. New York, NY: Brunner/Mazel Inc.

Willemsen, J., Della Rosa, E., & Kegerreis, S. (2017). Clinical case studies in psychoanalytic and psychodynamic treatment. *Frontiers in Psychology*, *8*(108).

Willemsen, J., Inslegers, R., Meganck, R., Geerardyn, F., Desmet, M., & Vanheule, S. (2015). A metasynthesis of published case studies through Lacan's L-schema: Transference in perversion. *International Journal of Psychoanalysis*, *96*, 773–795.

Wolman, B. (1971). Sense and nonsense in history. In B. Wolman (Ed.), *The psychoanalytic interpretation of history*. New York, NY: Basic Books.

Wolpe, J. & Rachman, S. J. (1963) Psychoanalytic evidence: A critique based on Freud's case of Little Hans. In S. J. Rachman (Ed.), *Critical essays on psychoanalysis*. New York, NY: Macmillan.

Woolfolk, R. L. (1986). Open peer commentary: Hermeneutics and psychoanalysis. *The Behavioral and Brain Sciences*, *9*, 265-266.

Wright, R. W. (1988). Causation, Responsibility, risk, probability, naked statistics, and proof: Prunning the bramble bush by clarifying the concepts. *Iowa Law Review*, *73*, 1001–1077.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Los Angeles, CA: SAGE Publications.

Žižek, S. (2006). *How to read Lacan*. New York, NY and London, England: W. W. Norton & Company.