UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

ARTHUR VIANA LOPES

A ESTRUTURA PSICOLÓGICA DO CONCEITO DE CONHECIMENTO

Porto Alegre, Junho de 2014

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

ARTHUR VIANA LOPES

Tese apresentada ao departamento de Filosofia da UFRGS como parte dos requisitos para a obtenção do título de Doutor em Filosofia.

Orientador: Prof. Dr. Paulo F. E. Faria

Porto Alegre, Junho de 2014

TERMO DE APROVAÇÃO

ARTHUR VIANA LOPES

A ESTRUTURA PSICOLÓGICA DO CONCEITO DE CONHECIMENTO

Tese de doutorado sob o título “A estrutura psicológica do conceito de conhecimento”, defendida por Arthur Viana Lopes e aprovada em Junho de 2014, em Porto Alegre, Estado do Rio Grande do Sul, pela banca examinadora constituída pelos professores:

Prof. Dr. Paulo F. E. Faria
Orientador/ UFRGS

Prof. Dr. Eros Moreira de Carvalho
Examinador/UFRGS

Prof. Dr. César Schirmer dos Santos
Examinador/ UFSM

Prof. Dr. André Joffily Abath
Examinador/UFMG

Prof. Dr. Roberto Horácio de Sá Pereira
Examinador/UFRJ

Aos meus pais.

Agradecimentos

Sou grato por ter encontrado na UFRGS um ambiente produtivo e que favoreceu minha formação no campo da epistemologia. No período do doutorado pude não só me dedicar à minha pesquisa individualmente, mas tive o privilégio de cursar várias disciplinas que foram do meu interesse pessoal em um nível que só poderia encontrar em poucos departamentos de filosofia do país. Sou grato aos professores do departamento com que pude trabalhar e agradeço especialmente a Paulo Faria por todo aprendizado, experiência enquanto fui atenciosamente acolhido como orientando. Agradeço à Priscila pelo inestimável incentivo e compreensão durante esses anos, e aos meus pais por seu apoio e auxílio imprescindíveis. Tenho gratidão também por todos que me ajudaram em minha adaptação a Porto Alegre, como os amigos acadêmicos que tive o prazer de encontrar. Por fim, a gradeço a Capes pela concessão da bolsa durante o período de realização deste doutorado.

Sumário

RESUMO

A proposta geral desta tese é discutir argumentos de uma linha cognitivista sobre o uso típico de intuições na literatura epistemológica. Em particular, o uso feito por epistemológicos interessados no projeto conhecido como análise do conhecimento. A questão central desta tese já está formulada em seu título: qual a estrutura psicológica do conceito ordinário de conhecimento? Em outras palavras, investigamos qual a organização psicológica da unidade mental que responde por nossos julgamentos intuitivos ordinários sobre casos de conhecimento. Argumentamos que a resposta a esta pergunta pode gerar lições importantes para o projeto epistemológico, em especial quanto a sua satisfatoriedade. Nossa investigação percorre literaturas da epistemologia, psicologia comparativa, psicologia do desenvolvimento, e a psicologia popular (*folk*). Em uma segunda parte da tese, tratamos de outros argumentos que surgem desta linha cognitivista na literatura epistemológica, tais como argumentos empíricos sobre a robustez de intuições – no sentido de serem amplamente compartilhadas – e considerações sobre as bases cognitivas de uma intuição. Por fim, descrevemos uma forma de dar sentido à metodologia epistemológica que leva em consideração argumentos cognitivistas em termos da noção de equilíbrio reflexivo.

Palavras-chave: Análise do conhecimento, conceitos, estrutura conceitual, filosofia experimental.

ABSTRACT

The proposal of this dissertation is to discuss issues from a cognitivist line about the typical use of intuitions in the epistemological literature. In particular, issues about the use of epistemologists interested in the traditional project of the analysis of knowledge. The central question of this dissertation is already formulated in its title: What is the psychological structure of ordinary concept of knowledge? In other words, we investigate what is the psychological organization of the mental unity that responds for our intuitive judgments about cases of knowledge. We argue that the answer for this question can provide important lessons for the epistemological project, especially about whether it can be satisfied. Our inquiry goes through the literature of epistemology, comparative psychology, psychology of development, and folk psychology. In a second part, we deal with other kinds of arguments from this cognitivist line in epistemology, such as empirical arguments about the robustness of intuitions – in the sense of being widely shared – and considerations about the cognitive basis of intuitions. Finally, we describe one way of making sense of the epistemological methodology which takes into account cognitivist arguments in terms of the notion of reflective equilibrium.

Key words: The analysis of knowledge, concepts, conceptual structure, experimental philosophy.

Introdução

Recentemente a questão do uso de intuições na metodologia filosófica ganhou seu próprio tópico na literatura. Uma das principais razões para isso foi o também recente renascimento do interesse pela análise conceitual gerado por autores como David Lewis, David Chalmers, George Bealer e Frank Jackson na década de 90 (Lewis 1994; Chalmers 1996; Bealer 1998; Jackson 1994, 1998). Esta metodologia evidentemente não é nova. De uma forma geral, a análise conceitual é tomada como uma metodologia que exemplificaria a autonomia e o caráter a priori da filosofia, o que torna sua importância óbvia. Sua história é praticamente tão longa quanto a história da filosofia, remetendo de forma clara pelo menos até os escritos de Platão, onde Sócrates e seus discípulos tentavam discernir a essência de coisas como piedade e conhecimento. O que autores como Bealer e Jackson fizeram foi tentar gerar defesas sofisticadas deste método, e especificar qual seria seu escopo e objetivos principais.

Estes trabalhos sobre análise conceitual suscitaram uma extensa discussão sobre questões metodológicas. O núcleo destas discussões, especificamente, diz respeito ao uso de intuições. Ao tentarem desenvolver uma teoria particular sobre coisas como conhecimento, o bem, verdade, referência, justiça, ação moral, livre arbítrio, etc., filósofos engajados na análise conceitual se apoiam fortemente em suas intuições pré-teóricas sobre o que são essas coisas e este é um aspecto fundamental deste método. Intuições tipicamente são descritas como possuindo conteúdo proposicional: filósofos afirmam possuir a intuição de que p, ou que nossa intuição sobre certo assunto é de que p, i.e., que nós compartilhamos um estado interpessoal. O que é mais importante, também é comumente assumido que este conteúdo possui um status evidencial especial. Teorias podem ser desenvolvidas a partir do que dizem inicialmente nossas intuições, uma teoria pode ser refutada se ela implica em consequências anti-intuitivas, e putativas descobertas filosóficas, assim como sucessos teóricos, em geral são atribuídas à concordância intuitiva sobre algum assunto. Em suma, intuições são vistas como a base com que filósofos testam suas teorias. Além disso, às vezes uma função ainda mais significativa é atribuída a intuições, como ser a fonte de todo conhecimento a priori (Bealer

2002). Como reação à recente revisão sobre a análise conceitual, no entanto, muitas questões importantes sobre a natureza de intuições – que afetam seriamente a plausibilidade da análise conceitual – foram levantadas. De uma forma geral: existe algo que realmente justifica o status e a autoridade atribuídos a intuições? Intuições são realmente confiáveis? Pessoas não podem simplesmente ter intuições conflitantes ou mudarem facilmente de intuições?

Particularmente, várias preocupações motivadas por uma orientação filosófica naturalista têm sido levantadas sobre intuições. Uma série de autores, como Robert Cummins, Stephen Stich, Jonathan Weinberg, Stephen Laurence, Eric Margolis, Joshua Alexander, Shaun Nichols, Joshua Knobe, Hilary Kornblith, dentre outros, afirmam que filósofos engajados em análise conceitual estão comprometidos com pressuposições empíricas muito significativas, viz., sobre os mecanismos cognitivos que geram intuições (Cummins 1998; Laurence & Margolis 1999, 2003; Stich & Weinberg 2001; Nichols et al. 2001, Kornblith 2007). De fato, algo que também explica a recente atenção dada ao tema de intuições é a repercussão dos trabalhos da chamada filosofia experimental, que levantam dúvidas sobre a normatividade de intuições – pelo menos da maneira que tipicamente são utilizadas na literatura analítica. Grosso modo, a ideia geral da filosofia experimental é que questões sobre concordância interpessoal e estabilidade intrapessoal de intuições, a ideia de que as intuições de filósofos podem representar as intuições do senso comum sobre assuntos particulares, etc., são pontos essencialmente empíricos, e devem ser investigados empiricamente. Os resultados da pesquisa empírica – feita por eles próprios –, no entanto, supostamente desfavorecem o status de intuições. Nichols et al. (2001), por exemplo, sugerem que fatores culturais podem afetar nossas intuições sobre questões particulares e, portanto, que elas não são universais como seria normalmente assumido.

Atualmente existe um acirrado debate sobre quais são as implicações da filosofia experimental. Os próprios filósofos empíricos, por assim dizer, discordam sobre qual posição geral devemos tomar quanto a intuições. Alguns se esforçam para mostrar o quanto nossas intuições podem ser não-confiáveis diante de questões filosóficas, parecendo sugerir que esta é uma metodologia a ser abandonada (Nichols et al. 2001, 2003). Outros parecem sugerir que embora os resultados empíricos contradigam alguns filósofos, eles são úteis justamente para corretamente favorecer alguma teoria ou desfazer algum problema filosófico (Nahmias et al. 2005). É também uma questão em aberto o quão conclusivos ou relevantes são estes resultados empíricos. Autores como Ernest Sosa (2008, 2009) e Henry Jackman (2009), por exemplo, demonstraram ceticismo quanto a isto. Independentemente de qual seja realmente a interpretação correta dos resultados da filosofia experimental, entretanto, ela possui o mérito de ter introduzido na agenda filosófica questões fundamentais sobre a natureza de intuições. Uma consequência importante, em particular, é o fato de que agora o debate sobre o status de intuições e a análise conceitual predominantemente possui um pano de fundo naturalista.

Os resultados desta discussão obviamente são de grande impacto para projetos tradicionais. A consulta a intuições tem sido uma característica marcante da filosofia analítica contemporânea e tem sido responsável por guiar as discussões teóricas de muitas áreas. Uma forma específica de consulta a intuições tem sido predominante a partir da década de 60, i.e., a análise de casos. Basicamente, situações hipotéticas ou atuais são descritas e filósofos se questionam se estes casos constituem instâncias de um conceito relevante. Existe uma vasta e conhecida literatura, passando por tópicos como teorias da referência, ações morais, implicações conversacionais, entre outros, onde casos imaginários são descritos na intenção de salientar uma intuição particular ou mostrar uma consequência anti-intuitiva de uma teoria. Se existem realmente problemas intrínsecos à consulta a intuições e filósofos estão simplesmente enganados sobre suas pressuposições, então a consistência dessas discussões fica seriamente ameaçada.

Talvez em nenhuma outra disciplina, no entanto, a análise de casos tenha ganhado tanta proeminência quanto na epistemologia. A principal razão para isso foi a imensa repercussão do célebre artigo de Edmund Gettier (1963) que reascendeu o interesse pela análise de conceitos epistêmicos ou, simplesmente, pela análise do conhecimento (AC). Gettier notavelmente determinou a agenda da epistemologia pelos trinta anos seguintes ao levantar contraexemplos à definição milenar do conhecimento como crença verdadeira justificada, e o modo como fez isso – através da descrição de dois contraexemplos intuitivos – resultou no grande uso de intuições a partir da descrição de casos na (AC). Graças às novas discussões sobre intuições, entretanto, agora é uma questão em aberto o quanto a (AC) é um projeto consistente. A proposta desta tese é justamente analisar as consequências dos atuais debates naturalistas sobre intuições no que diz respeito especificamente ao caso do conceito de conhecimento.

A questão sobre quais os mecanismos que geram intuições não é nova. Na verdade, muitos filósofos parecem assumir, ainda que implicitamente, uma mesma concepção geral herdada da psicologia empírica para justificar a consulta a intuições. Linguistas, por exemplo, se apoiam amplamente nas intuições linguísticas de falantes para desenvolver ou defender suas teorias. Todos nós possuímos intuições sobre a ambiguidade de sentenças, sobre se uma sentença é bem formada ou não, se uma nova palavra que ouvimos pertence ou não à nossa língua simplesmente por sua fonética, sobre se dois termos são sinônimos, se uma palavra é banal, etc. (Fiengo 2003). A ideia, defendida principalmente pelos adeptos da gramática gerativa, é que intuições linguísticas fornecem dados sobre a estrutura dos mecanismos cognitivos que constituem a competência gramatical dos falantes (Chomsky 1965). A ideia básica que parece ser assumida por filósofos engajados em análise conceitual é a de que, da mesma forma, intuições na metodologia filosófica servem para informar propriedades importantes sobre nossos conceitos. De fato, Jaakko Hintikka (1999) afirma que a origem moderna do termo "intuição" se deve à própria descrição de Noam Chomsky da metodologia da linguística.

Mais precisamente, uma concepção básica da literatura psicológica é a de que intuições dizem coisas importantes sobre o conteúdo de nossos conceitos, sendo estes tomados como entidades mentais relacionadas ao funcionamento de inúmeros mecanismos cognitivos. Por exemplo, as tendências de categorização intuitiva de sujeitos sobre pássaros informam aspectos do conteúdo do conceito PASSÁRO – a entidade mental que armazena informações sobre instâncias de pássaros. Intuição, portanto, a princípio seria uma espécie de acesso cognitivo ao conteúdo de conceitos. Esta, ao menos, parece ser a concepção básica assumida por alguns filósofos e herdada da literatura em psicologia de conceitos.

No caso específico da epistemologia, esta concepção é mais explicitamente adotada por filósofos de tendência naturalista como, por exemplo, Alvin Goldman (1976, 1986, 1992, 2007). Goldman, conhecidamente, é um entusiasta quanto à análise conceitual ao mesmo tempo em que fortes concepções naturalistas nortearam seu trabalho. Seu naturalismo é marcado não apenas pela sua posição quanto à relação entre filosofia e ciência natural, mas também pelo comprometimento em desenvolver uma epistemologia estritamente não-clássica ou não-cartesiana. Em "Discrimination and Perceptual Knowledge" (1976), por exemplo, Goldman coloca:

> The trouble with many philosophical treatments of knowledge is that they are inspired by Cartesian-like conceptions of justification or vindication. There is a consequent tendency to overintellectualize or overrationalize the notion of knowledge. In the spirit of naturalistic epistemology (...) I am trying to fashion an account of knowing that focuses on more primitive and pervasive aspects of cognitive life, in connection with which, I believe, the term 'know' gets its application (p. 102).

O que não só Goldman, mas uma série de outros autores está fazendo ao adotar o método da análise conceitual em epistemologia, é optar por analisar um conceito de conhecimento que está enraizado em um âmbito com aplicações familiares ao senso comum. Em particular, uma característica de teorias epistemológicas clássicas é elevar os critérios para o conhecimento de forma que uma enorme parte das atribuições de conhecimento que fazemos ordinariamente deixe de ser correta, e com isso diminuem drasticamente o escopo do conhecimento. Se ao invés disto um filósofo adota a análise de CONHECIMENTO – sua entidade mental que armazena informações sobre instâncias de conhecimento –, ele se compromete em desenvolver teorias cujos critérios para conhecimento e justificação ainda mantenham pelo menos uma parte desejada do que já temos elegido como conhecimento, i.e., ele analisa o conceito ordinário de conhecimento. Em "What is justified belief?" (1979), já com o propósito de desenvolver uma teoria de justificação, Goldman exemplifica claramente esta pretensão ao afirmar que "I do not try to prescribe standards for justification that differ from, or improve upon, our ordinary standards. I merely try to explicate the ordinary standards, which are (...) quite different from those of (...) 'Cartesian' accounts" (p. 106). Uma concepção naturalista da (AC), i.e., que pretende tornar este projeto aceitável de um ponto de vista científico, portanto, parece estar comprometida com seguinte hipótese básica: Nossas intuições epistêmicas fornecem informações sobre o conteúdo do nosso conceito CONHECIMENTO.

Epistemólogos envolvidos com a (AC), portanto, parecem estar comprometidos com a hipótese básica de que nossas intuições informam algo sobre o conceito CO-NHECIMENTO. Uma breve olhada na literatura, entretanto, mostra que o uso de intuições a partir de casos imaginários não tem sido efetivo para alcançar os objetivos da (AC), i.e., alcançar definições explicativas sobre conhecimento. Além disto, resultados e discussões de trabalhos da filosofia experimental têm desafiado pressuposições sobre a consulta de intuições e, consequentemente, gerado mais dúvidas sobre a utilidade destas consultas. Este quadro pessimista, contudo, certamente não implica na conclusão de que devemos desistir de obter informações importantes do que acreditamos ser o nosso conceito de conhecimento. Pelo contrário, um projeto de investigação que olhe diretamente para este conceito ordinário pode ajudar a jogar luz sobre o projeto epistemológico tradicional.

Em particular, alguns filósofos como William Ramsey (1992), Laurence & Margolis (1999), dentre outros, já atentaram para o fato de que a psicologia de conceitos pode ter um grande papel explicativo para o que ocorre na análise conceitual. Um exemplo é fornecido pela relação entre análise conceitual e a chamada teoria clássica de conceitos. Ramsey argumenta que um modo de fazer sentido ao empreendimento da análise conceitual é pensar que os filósofos estão implicitamente assumindo a teoria clássica de conceitos. Ramsey lembra que comumente filósofos tentam gerar definições através de uma conjunção curta de propriedades separadamente necessárias e conjuntamente suficientes. O argumento de Ramsey é que fazendo isto, e consultando intuições, filósofos supõem que nossos conceitos são mentalmente estruturados como definições – esta, grosso modo, é a teoria clássica ou ortodoxa de conceitos. Assim, por exemplo, o conteúdo do conceito SOLTEIRO seria 'HOMEM NÃO-CASADO'. Ambos os critérios desta definição – 'ser homem' e 'ser não-casado' – são separadamente necessários e conjuntamente suficientes para que algo seja solteiro. Ramsey então argumenta que para a análise conceitual, ao menos do modo que é tradicionalmente feita, ser bem sucedida, a teoria clássica deve ser correta. Existem, porém, muitas razões na literatura psicológica para considerar essa teoria falsa.

Algo que em princípio já torna a teoria clássica de conceitos implausível é o escasso número de definições consensuais. Se fosse o caso que nossos conceitos são armazenados mentalmente em forma de definições seria de se esperar que fosse mais fácil conseguir outros exemplos além de SOLTEIRO. O fato é que poucos conceitos parecem realmente suportar definições. A principal razão para assumir a falsidade da teoria clássica, no entanto, foi a constatação de uma série de fenômenos que não podiam ser explicados pela teoria clássica, os chamados efeitos de tipicidade. Por exemplo, uma série de experimentos influentes, especialmente de Eleanor Rosch (1973, 1975, 1978), mostra que pessoas intuitivamente categorizam alguma instância não como uma questão de "sim ou não". Isto é, pessoas comumente possuem julgamentos gradativos sobre o se algo é uma instância de uma categoria, e.g., "maçã é um bom exemplo de fruta", "figo não é um bom exemplo de fruta". Se nossos conceitos se estruturassem de forma definicional, seria de se esperar que um elemento simplesmente fosse julgado apenas como sendo ou não uma instância de categoria. Ao contrário, pessoas normalmente podem intuitivamente julgar o quanto uma instância de um conceito é "típica", "um bom exemplo" ou "representativa". A descoberta destes e outros dados levaram ao desenvolvimento da teoria prototípica de conceitos.

A teoria prototípica oferece um modelo diferente para a estrutura de conceitos. A ideia geral desta teoria é a de que o conteúdo principal de um conceito é um protótipo, i.e., um conjunto abstrato de propriedades típicas das instâncias do conceito, e que algo é categorizado como uma instância de c se ele é suficientemente similar ao protótipo de *C*. Categorizações intuitivas, portanto, seriam uma questão de acessar similaridade e não de aplicar uma definição. Uma consequência deste modelo é a de que dois elementos que possuem conjuntos de propriedades muito diferentes podem ser considerados como instâncias de uma mesma categoria. É necessário apenas que cada conjunto tenha uma soma de valores de tipicidade suficiente para se igualar ao valor do protótipo. Este modelo tem consequências bastante distintas das da teoria clássica. Ramsey (1992), por exemplo, importantemente argumenta que se é verdade que nossos julgamentos intuitivos de categorização estão ligados à tipicidade das instâncias de um conceito, então filósofos devem perder a esperança de (1) alcançar definições conjuntivamente simples e (2) que não possuem contraexemplos intuitivos. O problema é que um conceito pode possuir muitas propriedades a serem listadas de acordo com sua tipicidade e diversos conjuntos dessa lista podem gerar um julgamento de categorização. Dada essa estrutura cognitiva, qualquer formulação de uma definição curta simplesmente estará tratando arbitrariamente um subconjunto dessas propriedades como necessárias e suficientes. Inevitavelmente essa definição sofrerá com um contra-exemplo intuitivo.

Existe muita discussão em jogo atualmente sobre qual a teoria correta para conceitos. O que é mais importante aqui, no entanto, é o fato de que diferentes teorias sobre a estrutura de conceitos dizem coisas diferentes sobre a análise conceitual e a (AC). Por exemplo, se é o caso que CONHECIMENTO, especificamente, possui uma estrutura prototípica, isto poderia explicar por que não foi possível encontrar uma definição satisfatória de conhecimento, e porque é possível que a descrição de casos possa gerar intuições que favorecem teorias distintas – porque filósofos estariam descrevendo casos com diferentes conjuntos de propriedades, mas com valores de tipicidade suficientes para gerar a categorização intuitiva "este é um caso de conhecimento". Diferentes teorias sobre a estrutura de conceitos implicam diferentes perspectivas para projetos de análise conceitual. Uma investigação que tenta revelar a estrutura de CONHECIMENTO, portanto, poderia revelar lições importantes para o a análise do conhecimento.

Nosso objetivo geral nesta tese é discutir questões relacionadas ao uso de intuições epistêmicas que está tipicamente presente na literatura epistemológica, i.e., atribuições intuitivas a partir de casos imaginários. Para isso, a tese é dívida em duas partes. Na primeira parte trataremos da questão que motiva esta tese, i.e., a questão sobre qual é a organização psicológica de CONHECIMENTO. Esta é uma questão psicológica que é interessante por si só, e em boa parte esta investigação se parecerá como uma investigação puramente psicológica. Todavia, usamos aspectos da literatura epistemológica como evidência e esperamos com essa questão tirar conclusões sobre as expectativas sobre a análise do conhecimento assim como sobre o uso atribuições intuitivas por si só. No primeiro capítulo avaliamos algumas hipóteses estruturais básicas aplicados ao caso de CONHECIMENTO e defendemos que a categoriza conceitual deste conceito é a mesma de outros conceitos mentais como CRENÇA e CONHECIMENTO. Para isso fazemos uma breve revisão da literatura da *psicologia comparativa* e da *psicologia de desenvolvimento*. No segundo capítulo avaliamos as duas principais teorias com respeito a conceitos mentais, *viz.*, a *teoria teoria* e a *teoria simulacionista* e qual delas melhor explica a evidência que podemos encontrar com respeito a CONHECIMENTO, respondendo à questão central da tese. Na segunda parte tratamos da linha de investigação que tem sido chamada recentemente de *virada cognitiva da epistemologia* (Brown & Gerken 2012). Essa linha de investigação inclui os problemas levantados pela filosofia experimental assim como argumentos psicológicos sobre a base cognitiva de intuições. No terceiro capítulo revisamos uma série de trabalhos experimentais sobre duas intuições particulares, a intuição de casos Gettier (Gettier 1963) e o "efeito do erro" (Nagel 2012) e tentamos concluir se de fato estas intuições são robustas para o uso na teorização epistemológica. No quarto capítulo, avaliamos algumas tentativas de explicar as bases cognitivas dessas intuições e descrevemos uma forma de dar sentido à metodologia da epistemologia dados os tipos de argumentos que surgem nesta virada cognitiva. Esta posição explica como alguém pode defender a manutenção do projeto tradicional da análise do conhecimento ou teorias específicas de conhecimento diante tanto das conclusões da primeira parte da tese como das considerações sobre a base cognitiva de intuições que são introduzidas no quarto capítulo.

PART I – The problem of the psychological structure of KNOWLEDGE

Chapter 1

The folk concept of knowledge: KNOWLEDGE as a mental state concept

One of the most fundamental and traditional projects of philosophy is the attempt to develop theories of folk concepts of interest like *knowledge*, *belief*, *justice*, *free will*, *moral wrongness*, etc. Although in a sense we have a good grasp of those concepts and are typically able to apply them competently on a daily basis, that is a tacit understanding and does not fully serve to philosophical purposes. In particular, we cannot answer substantive questions like "can we have knowledge of the external world?" and "do we really have free will?" without an explicit understanding of what makes a situation, action, event, etc., c a case of concept C. When asking substantive questions like these, philosophers are interested in determining what exactly things like knowledge, moral wrongness and free will are, and folk concepts play an important role in the pursuit of this goal. It is often assumed, for instance, that they account for the *intuitive ascriptions* that we make on an ordinary basis. So when one characterizes actions c_1, c_3, c_7, etc., as wrong, for example, one is relying in one's concept WRONGNESS. Folk concepts, therefore, can provide a starting point for philosophical theorization. If our ordinary ascription patterns are significantly consistent, we can then try to systematize a theory which characterizes C – typically in the form of a definition with necessary and sufficient conditions that captures as much as possible these intuitive patterns. In sum, the project of *conceptual analysis*, as it is known, puts folk concepts in the central spot of philosophical scrutiny.

The *analysis of knowledge*, the particular conceptual analysis of KNOWLEDGE, occupies a large part in the history of epistemology. Epistemologists have always been interested in defining what knowledge is, and after the famous article by Edmund Gettier (1963), which brought unexpected problems for the then accepted definition of knowledge as justified true belief, the analysis of knowledge gained a renewed interest by philosophers who now saw themselves as having no reasonable characterization at hand. The ensuing developments of epistemology were mainly guided by the attempts to generate a new definition of knowledge, resulting, for example, in the great divide between *externalist* and *internalist* theories. Much of those developments can be seen as relying in the folk concept of knowledge. Sometimes this is more explicit as philosophers clearly try to incorporate an ordinary standard to their theories. Sometimes this is only implicit, as philosophers make great use of intuitive ascriptions, and these seem to come from our ordinary conceptual competence regarding knowledge. The fact that we still do not have a consensual definition of knowledge – we had a constant discovery of intuitive counterexamples to proposed definitions – led to a decline of the analysis of knowledge as it is classically understood, but substantive characterizations of knowledge are still in dispute, and those disputes also often rely on the folk concept. Even epistemologists who are not exactly trying to generate a definition of knowledge still strongly rely on intuitive knowledge ascriptions.[1]

Given the importance assigned to the folk concept of knowledge, one would expect us to know a lot about it, about its properties as a psychological entity, but this is not the case. Epistemologists are constantly trying to infer something about its content, but we do not know anything about how this content is organized. Indeed, some have suggested that philosophers may be making substantive empirical assumptions regarding this organization when using folk concepts (Ramsey 1992; Laurence & Margolis 1999). In particular, one apparent common assumption, at least in the first post-Gettier moment, is that this concept is *analyzable*. That is, many have seemed to assume that we can make explicit the content of KNOWLEDGE in terms of a definition. Does the *structure* of KNOWLEDGE enable this? William Ramsey (1992), for example, shows skepticism about this possibility by arguing that the main general theory from the psychology of concepts implies that we can never generate a definition that is intuitively satisfactory. This would explain the successive failures to generate a definition of knowledge which is not subject to intuitive counterexamples. The view assumed by Ramsey's argument is in the midpoint of possible views about the structure of KNOWLEDGE: a structured concept that does not admit of a definition. The thesis that it is in fact organized in terms of a definition, in its turn, is in one of the extremes. In the other extreme,

[1] This is the case, for example, of epistemic contextualism (DeRose 1992, 2009) and subject-sensitive invariantism (Hawthorne 2004; Stanley 2005).

KNOWLEDGE is not analyzable precisely because it is a *primitive* concept and, therefore, does not have a structure. If this is the case, then we have a strong reason to rethink the paradigm of analyzing such a concept and maybe to adopt something like Timothy Williamson's "knowledge first" approach to epistemology (Williamson 2000). More generally, what is the form or forms of the representations that KNOWLEDGE has?

Besides questions about the structure of KNOWLEDGE, we can also find many doubts about its content. For instance, we do not know the extent to which it constitutes the cognitive basis which answers for our knowledge ascriptions. Take the example of Gettier intuitions. We can interpret these as evidence that there is a conceptual incompatibility in thinking of an agent that could easily be wrong as being in a knowledge state. But if this is right, how does KNOWLEDGE respond for it? How is this condition stored in a concept? Also, some recent arguments in the literature try to dismiss certain patterns of intuitions by revealing their cognitive basis and how problematic they would be (Hawthorne 2004; Williamson 2005; Nagel 2008). We can interpret these arguments as showing that if we look closely to some epistemic ascriptions, we may find that they are explained by mechanisms not directly related to our conceptual competence about knowledge[2]. So we can say that one of the reasons for those possible mistakes is that we know so little about the actual content of KNOWLEDGE.

We are interested here in this surprising gap between the philosophical reliance in such a concept and what we actually know about it as a psychological entity. We think that this psychological question is interesting itself, but it is clear now that issues about the structure and content of the folk concept of knowledge may be highly relevant for the defense of epistemological projects or particulars theories[3]. In this chapter, however, we are going to focus only on questions about

[2] There is an ongoing dispute about what constitutes *conceptual competence* and what distinguishes it from mere *performance*. Ned Block (1986), for instance, claims that any inference or judging involving a concept is constitutive of that concept. This holist view contrasts with views which claim that not every inference or judging is constitutive of a concept, but only some more representative class of them (Peacocke 1992). The arguments we mentioned clearly assume the latter view. We cannot address this dispute here, but we think that it is much more a subsequent matter to our specific interests than a predetermining one. We can try to establish the content of KNOWLEDGE, but if we find that it is really difficult to determine its conceptual extension, then we may have reasons to adopt a more holistic view about the sense in which KNOWLEDGE constitutes a concept.

[3] One may wonder here about how much we really rely on a folk concept in our processes of theorization. Herman Cappelen (2012), for example, recently argued against the accepted general view

the structure of this concept. In particular, our objective is to evaluate some initial hypotheses about the structure question. These initial hypotheses are related to assumptions about the kind of concept KNOWLEDGE is. The orthodox view in the philosophical literature is that knowledge is a composite state of things and that this is something which has an intuitive basis. So, given the orthodox philosophical view, one immediate assumption is that KNOWLEDGE represents an *abstract concept* constituted by a relation between distinct states of things. We will argue that the structural hypotheses related to this assumption are problematic and, in a second moment, argue that KNOWLEDGE is actually a mental state concept.

1.1. THE DEFAULT VIEW OF CONCEPTS AND TWO BASIC HYPOTHESES

Before we proceed it is prudent to clarify what we mean here by "concept". In its minimal characterization, concepts are the constituents of thoughts. This characterization, however, fits into very different philosophical views about the nature of concepts. Beyond the common view that takes concepts as *mental representations*, opposing views claim that concepts should be understood as *abilities* (Dummet 1993) or as *abstract entities* like Fregean senses (Peacocke 1992). These views are generally motivated by skepticism about the explanatory utility of mental representations, but this skepticism in its turn, is often motivated by the very fundamental approaches that are being adopted to the study of mind and language. One of the ways in which these approaches can oppose is in their *focus.* Different focus can put mental representations in an inadequate explanatory level, as in the case one is primarily interested in *propositions* (Peacocke 1992). On the other hand, if one is interested in psychological-level explanation about a number of cognitive

that philosophers rely strongly on intuitions. We believe it is very difficult to defend this claim in the specific case of epistemology, but we will not address this question here. We think, however, that beyond the reliance on a folk concept, other significant criteria come into the equation of epistemological theorization, like general criteria for a theory to be satisfactory, accordance with logical principles, and maybe even plain observation. On this view, the criteria for knowledge that derives from KNOWLEDGE is also subject to adjustment coming from these other criteria. As will be clear the last chapter, we are inclined to believe that the relation between KNOWLEDGE and a substantive epistemological theory is something like a *reflective equilibrium* (Goodman 1955; Rawls 1971; For a meta-epistemological view in terms of reflective equilibrium, see Goldman 1986), but, anyway, the direct investigation of this concept can be seen as a potentially important source for theoretical adjustment.

processes such as categorization, inference, learning, etc., there is no reasonable choice besides mental representations. Given our interest here, we assume the *default view* of psychology that concepts are mental representations; internal symbols which participate in a number of cognitive processes.[4]

Concepts are initially divided into primitive concepts, which are not constituted by any other concept, and complex concepts, which are formed by simpler or primitive ones. A fundamental objective of a theory of concepts, therefore, is to explain how complex concepts are psychologically organized and very different structures were postulated by the main views in the literature, e.g., *prototypes*, *exemplars*, and *theories*. The strength of these theories is determined by how much the structures they postulate can explain the cognitive processes aimed by psychologists, but the fact is that there is no consensus about one best general theory. Indeed, most reviews now suggest that it is a mistake to assume that the explanatory power of a theory must be generalized to every cognitive phenomenon that psychologists attempt to account for or, even further, that a single phenomenon must be accounted by only one theoretical framework (Laurence & Margolis 1999; Murphy 2002; Machery 2009; Weiskopf 2009). Furthermore, a general theory of concepts must also say something about the nature of primitive concepts and how they can participate in cognitive processes.

Anyway, it can still be the case that specific processes attributed to a particular concept are better explained by particular structures. In view of this, we can try to establish whether it makes more sense to speak about KNOWLEDGE either as a structured or as a non-structured concept and, assuming that the former is true, what structure or structures better explain the processes related to KNOWLEDGE. Under the default view of concepts, therefore, there are two primary hypotheses to be explored. We shall call them the *non-structured hypothesis* and *the structured hypothesis*:

[4] Alvin Goldman (2007), for instance, claims that we must adopt the psychological notion of concepts if we want to make sense of the philosophical practice of consulting intuitions about particular cases. It allows us to explain how intuitions can be a reliable source for epistemological theorization – possessing a concept generates a disposition to make correct applications of this concept – and how they can also be unreliable – we can apply concepts incorrectly in a number of conditions, as when we are misinformed about the case, or we fail to process some property of the case.

Non-structured hypothesis (H1): KNOWLEDGE is a primitive concept.

The structured hypothesis (H2): KNOWLEDGE is a structured concept.

1.2. SOME HYPOTHESIS ABOUT COMPLEX STRUCTURE

Although our subject matter is so little explored, it is safe to say that (H2) is the common view regarding the folk concept of knowledge. As we will see, the few philosophers who said something related to this topic assumed (H2) and, furthermore, this seems to be the implicit assumption following the widely shared view that knowledge is a composite state of things. In this section we will review how (H2) appears in philosophy and discuss some particulars versions of it. We will argue that these particular versions all have difficulties.

1.2.1. The composite assumption

One usual way to start talking about knowledge is questioning what distinguishes it from mere *true belief*. This question is not only an inheritance from Plato, but also reflects the common view among most accounts of knowledge that for someone to be in a state of knowledge he must satisfy at least two individually necessary conditions: he needs to *believe* in a certain proposition, and this proposition needs to be *true*. These two conditions, especially the belief condition, are not free of controversies (Radford 1966), but are generally accepted and are often the starting point for theories. What the other conditions for knowledge are, on the other hand, is a far more controversial issue. Important for us here, these conditions are intuitively defended. It seems obvious that one cannot know something that is false, so that "know" is taken as a factive verb, and virtually all cases of propositional knowledge we can think about seem to involve the agent believing in a certain proposition[5].

[5] One well-known example against the belief condition is the case of the student who will submit to a quiz but thinks he does not know any of the answers (Radford 1966). All his answers to the quiz, however, are correct, which suggests that he knows, for example, that "Queen Elizabeth died in 1603", although he does not believe this proposition. It is debatable, however, if this case really contradicts the belief condition, since it is unclear how to interpret "belief" here. One can argue, for

These two conditions, and whatever additional conditions for knowledge, imply that knowledge is a composite states of things composed by very different things as internal conditions, like a mental state (and maybe justification), and external conditions, like truth (and maybe reliability of the belief-forming processes, or justification if that is an external property). In general, the idea that knowledge is not a composite state of things is just seen as counter-intuitive (Brueckner 2002). Given their intuitive basis, therefore, it is natural to assume that KNOWLEDGE is also composed by different properties. If every instance of knowledge is composed by a set of necessary and sufficient conditions which are characterized by very distinct properties, and at least some of these necessary conditions are intuitively defended, then it seems like our concept of knowledge must grasp these distinct properties and, therefore, it is a complex representation constituted by other concepts. So (H2) seems to follow naturally from the basic picture of conceptual analysis. If KNOWLEDGE is a complex concept it has a particular structure. But what would this structure be like?

1.2.2. The classical theory

One first hypothesis is that it has something equivalent to a definition. Indeed, William Ramsey (1992) suggested that most philosophers engaging in conceptual analysis are assuming something like the *classical theory of concepts*, which says that concepts are structured by underlying representations of necessary and sufficient conditions, i.e., they have a definitional structure. Ramsey argues that philosophers commonly have two requirements regarding the kind of definition they seek in conceptual analysis. First, that "the definitions be relatively straightforward and simple" (p. 60), what is illustrated by the standard way of defining a concept: "*c* is an instance of *C* if and only if *c* satisfies…", where the other side of the biconditional is a *small* set of properties. Second, for a definition to count as robust it cannot admit any intuitive counterexample. These are strong requirements for our concepts and seem to characterize most attempts of defining knowledge, demanding that the classical theory be right about KNOWLEDGE. One cannot achieve

example that the fact that his answers are correct is evidence that he actually believe in those propositions, although he erroneously does not recognize it.

a short definition of C that is free of intuitive counterexamples if C itself does not have a definitional structure.

Of course, this hypothesis is highly problematic. As a general theory of concepts, the classical theory is widely rejected. The main reason for this is that pretty much everything we would expect if it were the case that our concepts have definitional structure does not occur at all. One obvious problem is that it is really hard to find satisfactory definitions of any lexical concept. The recurring example is that of BACHELOR, which is defined as UNMARRIED MAN. Everything that is a bachelor is unmarried, and is a man. These two properties are all that is needed to define a bachelor. But how many definitions like that do we have? How one define PERSON, DETAIL, or, HAND, for example, in terms of necessary and sufficient conditions? We can probably find more examples of definable concepts, but why is this so difficult if our concepts have definitional structure? Applying the classical theory to the case of KNOWLEDGE has the consequence of making an absolute mystery the reason why we have such a difficult in generating a satisfactory definition of knowledge. After all, if there is any consensual conclusion from the analysis of knowledge is that it is hard to define such a concept. Indeed, one important lesson here is that whatever is the correct psychological theory for knowledge, it must provide a good answer for the following question: (Q1) what explains the difficulty to find a satisfactory definition of our concept of knowledge?

Furthermore, the classical theory is undermined by a number of experiments regarding cognitive tasks related to concepts. Influential works of Eleanor Rosch and her colleagues (1973, 1975, 1978; Rosch & Mervis 1975) made clear that there is an aspect of our processes of *categorization* which strongly contradicts what we would expect if the classical view were true. A class of findings known as *typicality effects* led to development of an alternative theory which could naturally explain them. The *prototypical theory*, the theory that emerged from those findings and substituted the classical theory, was able to do that by postulating a totally different kind of structure for concepts. Ramsey's general attack on conceptual analysis continues as he assumes the prototypical theory and shows how it implies the failure of this project. One possible second version of (H2), therefore, is the hypothesis that KNOWLEDGE has a prototypical structure. As we will see, if Ramsey's assumption is right, we could explain how the structure of KNOWLEDGE has made the task of defining knowledge so hard.

1.2.3. The prototypical theory

The experiments by Rosch and colleagues suggest that categorization is not a "yes or no" question as drawn by the definitional view, but instead corroborates Ludwig Wittgenstein's (1953) claim that categorization is more a matter of *family resemblance* than meeting necessary and sufficient conditions. The data at issue showed that instead of only two kinds of judgments – "x is a *C*" or "x is not a *C*" – our categorization judgments often reflect a taxonomic system of properties in which we assign *graded values* in a consistent way to members of a category. Crucially, the consistency of this system reflects *statistical* properties of the members. This means that the properties that determine conceptual membership are not defining, but characteristic: *c* is considered an instance of *C* if *c* has properties that are characteristic enough of *C*. There are different versions of the prototypical theory, but the central idea is that the main content of a concept is a *prototype*, i.e., an abstract set of properties of typical instances of the concept. Something is categorized as an instance of *C* if it is sufficiently similar to the prototype of *C*.[6]

A significant initial finding was that subjects can easily generate an intuitive ranking on how instances of a category are "typical", "representatives", or "a good example" of the category in question. These rakings enjoy considerable interpersonal agreement. For example, in an experiment involving FRUIT, apple and peach were intuitively judged by most subjects as typical instances, whereas raisins and pumpkins were considered atypical (Rosch 1973). Rosch & Mervis (1975) also found that these typicality judgments reflect an important pattern. When subjects are asked to list the typical properties of different instances of a category, the highest rated typical instances share a larger number of properties with the members of the category than the less typical ones. In other words, there is a function relating the number of shared properties and the typicality judgments, which allows the

[6] There are different models to explain how prototypes work and to answer questions as: how many properties the prototype and the instance must share to trigger a positive categorization? What determines which properties belong to the prototype? What determines the values that are assigned to each value? We will ignore these complications here. For our purposes it is enough that these models, contrasting with the classical theory, share a statistical account.

prediction of performance in a number of other tasks: when asked to list the instances of a category, subjects tend to list first the typical ones (Rosch 1973; Rosch Simpson, & Miller 1976). When asked to categorize something as quick as possible, they are quicker when categorizing typical instances (Rosch 1973; McCloskey & Glucksberg 1979). Typical instances are also more easily learned as members of a category (Rosch 1973; Rosch Simpson, & Miller. 1976).

If the classical view were correct, every instance of a concept would be equally judged as a good example and there should be no disparity in the time of categorization. Because the criteria for something to be categorized as *C* would be the same for every instance of that concept – a set of necessary and sufficient properties – the typicality of instances should not affect our intuitive categorizations, but they do. Furthermore, there is no requirement for the properties responsible for the typicality of an instance to be necessary properties (Rosch & Mervis 1975). Take the example of BIRD. A list of properties that predicts typicality judgments for this concept includes "flies", "sings", lays eggs", "is small", "eats in sets", "makes nests in trees" (Rosch 1973). However, none of these properties is really necessary for something to be a bird.

These findings reveal a much more complex picture for our concepts than was supposed by the classical theory. They suggest that our intuitions about conceptual membership are not defined by definitional information, but by the statistical information stored in prototypes and consist in the *assessing of similarity*, i.e., one instance is intuitively judged to be a member of *C* if it is similar enough to the prototype of *C*. If we generalize these results, the prospects of conceptual analysis are strongly affected. Ramsey, for example, emphasizes that one concept can have many properties to be listed according to their typicality and that several sets of those properties can trigger a categorization judgment. Add to this the fact that philosophers are always generating imaginary cases too freely, and then it seems very unlikely that the two requirements regarding the definition of a concept can be satisfied. Suppose, for example, that the prototype of *C* contains the properties $\{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}\}$, which are listed here in decreasing order of typicality. For any given *c* to be intuitively categorized as *C*, it must only be sufficiently similar to the prototype of *C* or, according to some models, the sum of the typicality values of its properties must reach a certain value. Therefore, any definition of *C* in terms of necessary and sufficient conditions could not be given by a simple small set of properties, but it would contain a minimally extensive disjunction of sets, e.g., "*c* is an instance of *C* if and only if *c* satisfies {f_1, f_2, f_3} or {f_1, f_2, f_{10}} or {$f_6, f_7, f_8, f_9, f_{10}$} or...". To propose a simple definition is to arbitrarily treat a subset of this disjunction as necessary and sufficient and to submit it to intuitive counterexamples, especially when it is so easy to create the most varied sets of properties through imaginary cases.

Of course Ramsey is assuming here that the prototypical theory can be taken as a general theory of concepts and, was we said above, this is extremely doubtful. But for us it still matters if it applies to the case of KNOWLEDGE. We will call the hypothesis of KNOWLEDGE having a prototypical structure (H2-*a*). If (H2-*a*) is true, then we have an available answer to (Q1): the analysis of knowledge has failed to produce a satisfactory definition of knowledge because KNOWLEDGE has a prototypical structure. Every time a definition is proposed it fails to capture all the sets of the extensive disjunction that reflects our intuitive ascriptions of knowledge. Besides, we can always manipulate the typicality values of imaginary cases by adding or taking out atypical and typical properties and produce intuitive counterexamples. This would explain a lot and would place a proper emphasis in our practice of generating very imaginative cases. But do we have reasons to believe that KNOWLEDGE has a prototypical structure?

There is no direct answer to that question. Until now, no one ever tried to empirically detect typicality effects regarding KNOWLEDGE and it is not clear just how plausible this possibility is. The problem, basically, is that the prototypical theory is mostly motivated by experiments dealing with *concrete concepts*, and we do not known the extent to which it can also account for *abstract concepts*, i.e., entities that are neither purely physical nor spatially constructed. The prototypical theory obviously can deal with some level of abstractness. One of its central ideas is that the prototype is formed by cognitive processes that *abstract* properties from particular instances. Also, proponents of it have long shown that concepts like WEAPON and VEHICLE, for example, present typicality effects and seem to store statistical information (Rosch & Mervis 1975). Nevertheless, they constitute superordinate categories, having as members categories which are diverse and that share few properties between them, as knife, gun, sword, car, truck, bicycle, etc. Their meanings, therefore, are somewhat abstract, and that does not preclude them of containing statistical information. As WEAPON and VEHICLE, the instances of KNOWLEDGE are also very diverse, so one may claim that it makes sense to think of it as having prototypical structure.

This analogy does not go very far, however. What would be the basic categories of knowledge? Would it be things like perceptual and testimonial knowledge? This is questionable. Compare with "weapon". Categories like knife, sword, and gun are in a more basic level of experience than "weapon", such that it is more natural to ordinarily think and categorize about KNIFE and GUN than WEAPON. Can the same thing be said about PERCEPTUAL KNOWLEDE or TESTIMONIAL KNOWLEDGE? We do not believe so. We think these categories can in fact be ordinarily identified, but are much less generic than KNOWLEDGE itself, at least in our culture[7], and instead constitute *subordinate* categories. Furthermore, even if it was easy to agree that KNOWLEDGE constitutes a superordinate category, the point is that it is not clear how plausible is the possibility that it has a prototypal structure as far it is a concept with a high degree of abstractness. One difficulty is that most properties we can think about as constituting particulars instances of knowledge are themselves abstract, e.g., "to have a belief", "to have a true belief", "to have good reasons", "to have favorable evidence", "to have a feeling of certainty", "to have a reliably produced belief", etc. Because the literature in the psychology of concepts focuses more on concrete concepts, it is not obvious how these properties can be represented and, more importantly, how they are represented in a way that allows us to store statistical information about them from particular instances.

There is some work that shows evidence that abstract concepts can in fact have a prototypical structure. Linda Coleman and Paul Kay, for example, ran a study which detected typicality effects relative to the concept LIE (Coleman & Kay 1981). Different cases are better or worse instances of a lie depending on dimensions such as whether what is being told is true or false, whether the speaker

[7] Some languages like Turkish and Korean have grammaticalized evidentials that indicate the source of the asserted proposition, e.g., inference, testimony or own perception (Aikhenvald 2004). This may suggest that in these cultures PERCEPTUAL KNOWLEDGE and TESTIMONIAL KNOWLEDGE are basic level categories. However, there are experiments that failed to find significant differences in the performance of young English-speakers and Korean-speakers regarding the tracking of evidential source (Papafragou 2007), suggesting otherwise.

knows that what he is telling is true or false, and whether the speaker has or not the intention to deceive. James Hampton, however, tested eight more abstract concepts for a prototypical structure and although he found some positive results, e.g., CRIME, SCIENCE, he also failed to detect evidence for prototypical structure for other concepts, such as BELIEF and INSTINCT (Hampton 1981). From this we can conclude that it is not safe to assume that the prototypical theory applies equally well to every abstract concept. Indeed, the idea that it should not been seen as a general theory of concepts is not new (Laurence & Margolis 1999). So we cannot be sure about when it applies, except by going case by case. Given what we know empirically, KNOWLEDGE may be more like BELIEF than LIE. We will explore this possibility in the last section. For now, we want to focus on one armchair objection to (H2-*a*).

One can object to (H2-*a*) by noting that it implies that we have judgments about how much certain situations are a good example of knowledge. But that does not seem to be the case. Observe that, as we mentioned, instances of knowledge are very diverse. We attribute knowledge to children, animals, senile people, beliefs acquired by perception, inference, testimony, explicitly justified beliefs, etc., and every case is particular to a specific situation and context. If such diversity were organized by a summary representation defined by statistical information it would be natural to think of some of them as being a better example than others. But that is not exactly what happens. Of course some cases are more confusing and seem to be in the boundary, but once we categorize something as a case of knowledge (and this happens despite the particularities of each case) it just seems like a good case of knowledge as any other. Qualitatively, it does not matter if it is perceptual, testimonial, held by a child or adult, it is just a case of knowledge. Then either (H2-*a*) is at odds with this, or every instance of knowledge and its properties are statically equivalent, what it is just implausible.

Of course, whether (H2-*a*) is true is ultimately an empirical question and we cannot refute it just by this objection. This is especially the case when we consider that there is no reason to assume that KNOWLEDGE, just like any other concept, stores just one kind of information (Laurence & Margolis 1999; Murphy 2002; Machery 2009; Weiskopf 2009). Anyway, this objection and the abstract nature of this concept suggest that it is implausible that it stores statistical information.

1.2.4. The exemplar theory

Alvin Goldman (1992), one of the few epistemologists to say something about our subject matter here, proposed a theory not about the structure of KNOWLEDGE, but of JUSTIFICATION. Goldman claimed that what explains certain epistemic intuitions against *reliabilism* is that JUSTIFICATION stores *exemplars*. If we extend this hypothesis to the case of KNOWLEDGE we have an alternative theory that can better handle the objection above. Let us call the hypothesis that KNOWLEDGE has exemplars in his structure (H2-*b*).

The exemplar theory of concepts first emerged as an alternative to the prototypical theory. Not every psychologist was convinced about the existence of summary representations formed through the abstraction of properties from particular instances. Shortly after the early development of the prototype theory a different view came up (Brooks 1978; Medin & Schaffer 1978). This view also tries to explain typicality effects on categorization by similarity judgments, but instead of positing summary representations for categories, it claims that a concept stores particular *exemplars* of the category, i.e., a set of detailed particular representations. So an exemplar is a body of information about the properties of particular instances. Roughly, according to *the exemplar theory of concepts*, to have a concept *C* is to think of *C* as being the class of entities similar to its set of exemplars stored in long-term memory. To have a concept FRUIT is to think of a class of objects similar to a certain set of objects, e.g., an apple, a peach, a watermelon, a tomato, etc. Categorization is a similarity judgment that compares an input to one or a set of stored particular representations. Important for us, the idea that a summary representation is limited can be used to handle the objection against (H2-*a*). About JUSTIFICATION Goldman says:

> The hypothesis I wish to advance is that the epistemic evaluator has a mentally stored set, or list, of cognitive virtues and vices. When asked to evaluate an actual or hypothetical case of belief, the evaluator considers the processes by which the belief was produced, and matches these against his list of virtues and vices. If the processes are

> matched partly with vices, the belief is categorized as unjustified. If a belief-forming scenario is described that features a process not on the evaluator's list of either virtues or vices, the belief may be categorized as neither justified nor unjustified, but simply *non*justified. (1992, p. 157).

A list of virtuous processes may well contain quite different cases, beliefs formed by vision, hearing, memory, and a number of approved kinds of reasoning. Vicious processes would include things like "wishful thinking", "guessing", and "to ignore contrary evidence". If JUSTIFICATION really constitutes part of the content of KNOWLEDGE and Goldman's hypothesis is correct, then KNOWLEDGE has some exemplars in its structure. That would be a weak interpretation of (H2-*b*). A stronger interpretation of it says that KNOWLEDGE contains a set of exemplars of particular instances of knowledge. This hypothesis enables the storage of a set of quite diverse instances of knowledge, what puts (H2-*b*) in a better position with respect to the argument above against (H2-*a*). It can explain the diversity of our knowledge ascriptions and, furthermore, neither requires us to store statistical information about these instances nor to have intuitions about their typicality: every instance categorized as knowledge is a good instance of knowledge as far it is sufficiently similar to one of the exemplars stored in long-term memory.

Regarding (Q1), just like the prototypical hypothesis, (H2-*b*) implies that we cannot achieve a satisfactory short definition of KNOWLEDGE. This is because the set of stored exemplars is too diverse to be captured by a small conjunction of conditions. Indeed, the exemplar framework in its original motivation is somewhat adverse to the possibility of summary representations of a class. The diversity of exemplars is a good explanation of why we may find counterexamples to the necessity of certain conditions: the definition just excludes one, or some, of the exemplars. Concerning counterexamples to sufficiency of conditions, in a similar way to Goldman's hypothesis, we may have to adjust (H2-*b*) to include the storage of *negative* instances. For it is not just the case that some instances said to be knowledge are not similar to other cases of knowledge, but we also make negative assessment of these instances as, for example, in Keith Lehrer's Truetemp case (1990), Laurence Bonjour's clairvoyance case (1980), Carl Ginet's fake barns case

(Goldman 1976), etc. Imaginative cases, therefore, would share some properties with stored negative exemplars. But, again, how plausible is (H2-*b*)?

Just like (H2-*a*), it is not clear how much an exemplar theory can be applied to the kind of abstract concept KNOWLEDGE is. Both theories equate categorization to processes of assessing similarity, but, contrary to concrete concepts, it is unclear whether these structures could contain the abstract properties that are presented in instances of knowledge. Of course this not makes (H2-*b*) implausible, just not obvious. However, there is at least one objection we can aim to it.

In the way we conceived (H2-*b*), KNOWLEDGE not only stores a set of exemplars, but this set is composed of diverse particular instances of knowledge and maybe also specific instances of non-knowledge. This could answer for the diversity of our attributions and their apparent equivalence of epistemic status. One problem, however, is that (H2-*b*) by itself cannot explain how these exemplars are acquired as instances of the same category. Consider how exemplars are used in learning. When one is learning a new category one is presented to an instance of that category, acquires a detailed memory of that instance, and uses it to categorize new instances of that category. This, however, can only happen if the instances of the category share many properties and we just said that by hypothesis this is not the case of KNOWLEDGE. In the case of a category with diverse instances we need to somehow learn that these are members of the same category. In other words, we cannot acquire a set of diverse exemplars of knowledge without having additional information saying they are instances of the same concept. (H2-*b*) by itself, therefore, makes it a mystery why these exemplars are acquired in the first place. Note that one does not solve the problem by postulating that we acquire them by being ostensibly presented to diverse instances. If KNOWLEDGE was acquired this way, and we do not believe that, it would still be mysterious why these instances are categorized as knowledge by those presenting them to us in the first place. This becomes even more complicated if exemplars are supposed to explain the negative judgments we make about some instances.

From this objection we can generalize another criterion for a theory about the structure of KNOWLEDGE. Such a theory must answer (Q2): How KNOWLEDGE is acquired? Any satisfactory theory must present a clear story of concept acquisition, or at least have a reasonable prospect of explaining it. One can try to defend

(H2-*b*) by arguing that the fact that we need additional information to make sense of the hypothesized diversity of exemplars does not preclude KNOWLEDGE from actually storing such exemplars. We agree with that. Neither (H2-*a*) nor (H2-*b*) requires that there is only one kind of structure stored in KNOWLEDGE. Indeed, processes of categorization in terms of similarity can be seen as just one kind of process related to certain cognitive tasks (Laurence & Margolis 1999). In particular, one general account says that they are responsible for *quick* categorization, while more reflected judgments depend on other kinds of structures (Osherson & Smith 1981; Smith et al. 1984). So, while (H2-*a*) still seems implausible because of the apparent lack of typicality of instances, it is possible that exemplars answer for similarity judgments at the same time other kind of information explains their acquisition and other kinds of categorization. This, however, opens again the possibility of KNOWLEDGE having some sort of summary representation. The plausibility of (H2-*b*) depends on us finding a good candidate for storing that other kind of information and that answers (Q1) and (Q2).

1.3. FROM ABSTRACT CONCEPTS TO MENTAL STATE CONCEPTS

We have no direct empirical evidence for or against (H2-*a*) and (H2-*b*), but we found problems with these hypothesis. Prototypes and exemplars are well suited to explain many concrete concepts, but they seem limited when applied to KNOWLEDGE. At this point we may wonder whether we are starting from the right assumptions while investigating the matter of the structure of this concept.

Maybe we are just raising the wrong questions about KNOWLEDGE and the kind of concept it is. We have argued that the correct theory about our folk concept of knowledge must be applicable to the kind of abstract concept it is. But what specific kind of category KNOWLEDGE represents? Given the orthodox philosophical view that knowledge is a composite state of things, an immediate answer is that KNOWLEDGE represents an abstract property constituted by a relation between distinct states of things. Nothing different from this was assumed so far. But there is another answer available. Alternatively, one can defend the view that KNOWLEDGE is not only an abstract concept, but essentially a *mental state con-*

cept. This contrasts with the orthodox view by claiming that knowledge is not conceptually seen as composite state of things constituted by the mental state of belief and other things, but as a mental state on its own.

A possible motivation for this view is in the observation that the general description of mental states concepts fits well with what we are finding about KNOWLEDGE. For example, Anna Papafragou and colleagues said about mental verbs that:

> [T]hey do not refer to perceptually transparent properties of the reference world; they are quite insalient as interpretations of the gist of scenes; (…) the concepts that they encode are evidently quite complex or abstract; and they are hard to identify from context even by adults who understand their meanings. (Papafragou et al. 2007, p. 126)

More than the acknowledged abstractness and complexity of these concepts, the lack of properties that can perceptually identify its particular instances can be an indication of the nature of KNOWLEDGE and potentially explain the difficulties of the prototypical and the exemplars hypothesis. As Papafragou et al. add, "words that refer to mental states and events lack obvious and stable observational correlates: as a general rule, it is easier to observe that jumpers are jumping than that thinkers are thinking" (p. 128). Given the lack of prototypical structure for BELIEF, it is plausible that the apparent limitation of these theories in relation to abstract concepts may apply to mental state concepts, at least some of them. This alternative view, therefore, may put us on the right track with respect to our central question.

It may be hard to accept that knowledge can be seen as a mental state, however. There are reasons why the orthodox view is directly opposed to this idea. Indeed, Williamson (1995, 2000, 2009), who is the leading voice in philosophy to advocate that knowledge should be understood as a mental state, has found much resistance to his view. Two main arguments appear here (Nagel 2013). One problem is that we have a tendency to consider mental states as something localized. Roughly, they are something "inside the head" of agents – they are in their neural states or supervene on them. A relation between external events and what is inside the head of the agent *prima facie* does not qualify as entirely mental. Besides being admitted that knowledge is a factive state and that the truth value of a mental state (among those that have truth-values) depends on its relation to the world, most theories are based on the idea that knowledge is more than true belief and many non-mental conditions have been postulated. So, how could knowledge be seen as a mental state? "This idea is just metaphysically odd", one could say.

Another problem is the view that our understanding of action is fundamentally and sufficiently determined by the attribution of belief, which follows a principle of belief-desire reasoning (Davidson 1963; Dennett 1971). One can argue that along with desires, beliefs can explain behavior more efficiently and economically than any other candidate mental state. For example, suppose that Jean opened his freezer and took out some ice cream after saying "I am dying for some ice cream". We could explain his behavior both by attributing knowledge or belief in a proposition *p* about the existence of ice cream in the freezer. Now suppose the alternative situation where there is no ice cream in the freezer, Jean says "I am dying for some ice cream", opens the freezer and seems frustrated. Jean's behavior would be equally explained by the attribution of the belief in *p*. So why would it be more accurate to say that Jean knows *p* in the first situation and believes *p* in the second situation instead of saying only that he believes *p* in the first and second situation? We could still attribute knowledge in the first situation, but that do not seem to add anything to the explanation or prediction of Jean's behavior. In response, Williamson has developed cases where the attribution of knowledge is supposed to better explain the behavior of agents. Important for us, arguments like these can be understood as motivating the rejection of the idea that KNOWLEDGE is a mental state concept: KNOWLEDGE is not a mental state concept because besides being *prima facie* metaphysically odd to think of knowledge as a mental state, we do not need KNOWLEDGE to understand others' behavior.

We reject this conclusion. Of course both of these arguments affect the motivation for saying that KNOWLEDGE is a mental state concept, but neither implies its falsity. This idea can be defended and, as we will see, this classification has important consequences for the structural question we are pursuing. So, again, how could knowledge be seen as a mental state? In sharp contrast with the orthodox view of philosophy, in the psychological literature knowledge is constantly listed as just another mental state alongside beliefs, desires, intentions, phenomenal states, etc. (Premack & Woodruff, 1978; Apperly, 2011; Baron-Cohen et al., 1994; Call & Tomasello, 2008; De Villiers, 2007; Heyes, 1998; Saxe, 2005; Sodian et al. 2006; Wellman & Liu, 2004). This happens in developmental, comparative, and social psychology. More specifically, such classification features in research on our *mindreading* abilities, which raises questions like: At what age do children start to track knowledge of others? How are the kinds of processes that allow us to discriminate knowledge? To what extent nonhuman animals can track the knowledge of others? Importantly, much of this literature treats these questions as a matter of conceptual acquisition. For example, Ian Apperly says that:

> (…) [M]any researchers hold that the development of mindreading consists of acquiring abstract mental state concepts, and that these concepts constitute a 'theory' about how the mind works. This has led to 'theory of mind' becoming the predominant term for mindreading in the academic literature. (2011, p. 2)

Under this framework, we are authorized to say that not only adults, but young children and some nonhuman animals have a concept of knowledge because they successfully pass cognitive tasks regarding the discrimination of knowledge. Motivated by the metaphysical argument above, one may doubt psychologists are meaning the same thing here as philosophers by "mental state". For, over again, a factive state could not be mentally localized. This, however, is unwarranted. More than a terminological issue, one big difference between literatures is in a presupposition about the way we understand mental states. In particular, this argument presupposes that it is not possible that the natural way we understand mental states already incorporate relations between the agent and the world, including a factive relation. But nothing in the argument prevents this. In contrast, this is part of the reasoning behind the psychologist's view about knowledge being a mental state. That is, psychologists readily accept that our mindreading processes are partly performed by taking into account relations between the agent and the world, and this is properly supported by the empirical evidence. Most of the cognitive tasks regarding states of knowledge, ignorance, and false belief, for example, consist precisely in testing subjects' ability of tracking specific relations. Under the framework of psychology, therefore, KNOWLEDGE can perfectly be a mental state concept (Nagel 2013).

In what follows, we will endorse the view that KNOWLEDGE is a mental state concept. Note, however, that one can concede the psychologists' point and still deny that knowledge really is a mental state. We will not try to argue in favor of the more strong position defended by Williamson that the state of knowledge really is a mental state. As far as this is a metaphysical matter we doubt that evidence from psychology, which is what concerns us here, can solve it. In contrast, we think it is reasonable to trust in the psychological literature to help us settle certain matters regarding KNOWLEDGE, especially when it comes to questions about which the empirical evidence has much to say.

One specific and crucial point of disagreement that can be empirically addressed is that of the relation between KNOWLEDGE and BELIEF. The philosophical orthodoxy claims that the recognition of knowledge starts from the recognition of belief and the truth of this belief. That is, it takes KNOWLEDGE to be a more complex concept than BELIEF. But the psychological literature opposes to this idea. The virtual consensus among psychologists is that BELIEF is a more sophisticated and later acquired concept than KNOWLEDGE, and this consensus derives from the evidence around the research on our mindreading abilities. In the next section we will review part of the empirical evidence supporting this. Besides rethinking the composite assumption about KNOWLEDGE and BELIEF, this will also be useful to undermine these philosophical arguments against KNOWLEDGE being a mental state concept.

1.3.1. Rethinking the composite assumption

There are different sources that we can use against the composite assumption. One of them is the evidence we got from psychology of development. Research on children abilities allow us to observe when we start to discriminate particular mental states and, as is usually interpreted, when we acquire their respective concepts. Of course, there are nuances in this literature. Despite the common focus on concept acquisition, there are fundamental controversies around this matter and difficulties with the interpretation of the data which can make it hard to assert when one possesses a mental state concept. We will discuss one of these fundamental issues further. For now, we are interested in what the empirical evidence has to say about the composite assumption regarding KNOWLEDGE and BELIEF.

One possible difficulty, for instance, is that there are a number of studies using indirect observations as eye-gaze (Clements & Perner, 1994; Garnham & Perner, 2001; Southgate et al. 2007), and shared attention behavior (O'Neill 1996; Moll & Tomasello 2007), showing that very young children (before 3 year olds) have some sensibility to what others perceive, know or believe; sometimes even before they can make explicit judgments about these mental states. These studies can pose problems for one trying to determine when someone can be credited with concepts of knowledge and belief. However, since this sensibility is distant from the competence of older children and in some important cases is not even present in their explicit judgments, the usual interpretation is that they do not yet represent the possession of the relevant conceptual competence – although it is still not clear what is the relation between these early sensibilities and the latter competences (Apperly 2011). Anyway, there are some kinds of cognitive tasks that are paradigmatic and revealing enough for our purposes here.

1.3.1.1. False-belief and knowledge-ignorance tasks

One of them is the influential false-belief task. This task basically consists in the presentation of a story in a child-friendly way in which the mental state of one of its characters is judged by the children. In particular, it is used to test children's capacity to detect beliefs by presenting them to a relatively simple case of false belief. In its original version (Wimmer & Perner 1983), the experimenter presents Maxi (a puppet) and shows the children Maxi placing his chocolate in a certain location *x*. Maxi goes away and the children sees Maxi's mother replacing the chocolate to location *y* in the absence of Maxi. Maxi comes back, and the children are asked: "Where will Maxi look for his chocolate?". Researchers found that most 4-5 years-olds (about 60%) wrongly answer that Maxi will look for his chocolate in location *y*, presenting a difficulty in detecting the relevant mental state of Maxi and an egocentric pattern of judging from his own point of view. In contrast, most 6-7 years-olds (about 90%) correctly answered that Maxi would look in location *x*. There are a number of variations of this task and although more extensive analysis corrected the general lack of competence for 3-4 years-olds and the competence achievement for 4-5 years-olds, this significant change in competence was not cancelled by any methodological variant, proving it to be a robust developmental phenomenon (Wellman et al. 2001).

False-belief tasks show the inability of 3-4 years-olds to construe others' point of view and this has often been interpreted as an indication of the lack of the concept of belief. Conversely, "(…) there is a general presumption that when children do pass false belief tasks we can be sure that they do have this concept" (Apperly 2011, p. 14). Crucial for us, this inability contrasts with children's competence to detect *ignorance* states and with what we would expect in case the composite assumption were correct. For example, in a comparison study, pairs of children were given a box with a certain object in it, in the absence of one of the children the other witnesses the object being replaced for another kind of object. In addition to the questions intending the detection of the belief of the absent child, e.g., "if we ask John what is in the box, what will he say?", some groups of children were asked questions intending his knowledge state, e.g., "does John know what is in the box?". While only 6% of the 3-years-olds correctly answered the attribution of false-belief, 39% were able to correctly deny knowledge to the agent. While still less than half of 4-years-olds (44%) correctly answered the belief question, 81% of them denied knowledge to the other child. With 5-years-olds the correct answers raised to 76% and 86% respectively. This seems to contradict the composite assumption. For if our attributions of knowledge starts with the attribution of belief, why would we see this gap between the performances of groups?

Since the most obvious state of belief that falls short of knowledge is a false belief, unless there is good explanation for this gap, it seems to reveal that is easier to attribute knowledge or its lack than belief *per se*. But at the same time there seems to be no such a reason, this conclusion is supported for further evidence. For instance, meta-analysis of diverse kinds of tasks involving the attribution of mental states puts knowledge-ignorance tasks in an easier degree than false-belief tasks in a developmental scale as children present competence to them earlier (Wellman &

Liu 2004). Furthermore, evidence from comparative psychology is also very suggestive.

1.3.1.2. Evidence from comparative psychology

The literature of comparative psychology went through exciting twists in the last years. Until recently, most researchers responded negatively to the question about whether nonhuman primates could understand the psychological states of others (Tomasello & Call 1997; Heyes, 1998; Povinelli & Vonk 2003) famously asked by David Premack and Guy Woodruff (1978). This was motivated by results with chimpanzees which *prima facie* showed them to have very poor understanding of mental states, or perhaps none at all. For example, in a situation where food was hidden and two humans point to different locations, young chimpanzees follow the gestures indiscriminately despite the fact they witnessed only one of the humans seeing the food being hidden (Povinelli, et al. 1994), which may suggest that they have no understanding of knowledge and ignorance. Even more striking, researchers found that chimpanzees gesturally beg humans for food even when they are blindfolded (so cannot see) or have a bucket in the head (so have no perceptual access at all to the begging) (Povinnelli & Eddy 1996). Results like this led some to the general conclusion that the capacity of chimpanzees to predict the action of others or apparently understand their psychological states is based on past experiences and maybe specialized cognitive adaptations, that would be what allows adult chimpanzees to prefer the indication of the human who saw where the food has hidden, for example. In other words, they do not understand what is inside the head of others, but only learn behavioral rules.

However, this conclusion has changed. A new methodological paradigm proposed that chimpanzees could be more skillfull in competitive situations than in situations requiring communicative cooperation with humans (Hare & Tomasello 2004). Under this paradigm researches began to find evidence of chimpanzees' competence with respect to certain mental states. Relevant for us, they tested whether chimpanzees can track what other chimpanzees have seen and, therefore, what they know. In one experiment a subordinate and a dominant chimpanzee were placed in two different rooms separated by a space where they could see each other. This space contained two barriers which were used to place one piece of food. Two conditions were tested. In one of them the dominant could see the food being placed behind one of the barriers. In the other condition a guillotine door prevented the dominant ape of seeing that a piece of food was placed in the room behind one of the barriers. In both conditions the food was placed in the side of the barrier that could be seen by the subordinate chimpanzee. Researchers found that the subordinate subjects were more likely to approach the food in the condition where the dominant chimpanzee had not seen the food being hidden, suggesting that they indeed can, in some degree, track the knowledge of others (Hare et al. 2001). This conclusion is supported by a number of other experiments using different scenarios, what led some initially skeptical researchers like Josep Call and Michael Tomasello to say "(...) we believe that there is only one reasonable conclusion to be drawn from the totality of the studies (...): chimpanzees, like humans, understand that others see, hear and know things" (Call & Tomasello 2008, p. 189). Furthermore, the tracking of knowledge by humans and nonhumans seems to clearly incorporate the relevant kind of relation between the agent and the external world, *viz.*, a factive relation.

In contrast, nothing changed in this literature relative to the state of belief. Previous studies already showed that chimpanzees and orangutans do not succeed in non-verbal versions of the false-belief task. (Children performance in this task follows their performance in the verbal versions of it – children from 4 years old succeed in it). In addition, no positive evidence was found in competitive versions of the false-belief task. For instance, the experiment from Hare et al. (2001) also included a condition in which the subordinate chimpanzee witnesses the dominant seeing the food being placed behind one barrier, but not seeing it being replaced to the other barrier. The subordinate thus could anticipate that the dominant would look for the food in the wrong place and compete more effectively for food, but even having a head start that was not what was observed – the subordinate kept distance from the food. Others studies which tested different competitive scenarios and detected a sensitivity of subjects in conditions involving uninformed competitors failed to detect sensitivity to misinformed competitors (Kaminski et al. 2008). In all these studies, chimpanzees prefer to approach food in conditions where their competitors are ignorant about the location of food, and avoid approaching the food both in conditions in which the competitors know where the food is and conditions in which the competitors are misinformed about the location of the food. So they seem to be able to track to some degree the knowledge of others, but when it comes to false-belief tasks, it is like they wrongly and egocentrically judge the competitors' mental state from their own point of view, just as human children do. Putting together the evidence from developmental and comparative psychology, therefore, it is tempting to conclude that understanding knowledge is easier than understanding belief.

1.3.1.3. Why KNOWLEDGE would be simpler than BELIEF?

Granting the suggestion of empirical evidence, what would explain the simplicity of reasoning about knowledge? This point is addressed by Jennifer Nagel (2013), who endorses both Williamson's view that knowledge really is a mental state and the assumption from psychology that knowledge is naturally seen as mental state. She emphasizes Williamson's claim that knowledge is the more factive state and thus is a state which essentially involves matching how things are. The essence of belief states, in the other hand, is not characterized by this correspondence, which, contradicting the judgments of parsimony from the composite view, actually makes it a more complex notion. That is, it is part of the essence of beliefs that they can be either true or false, so the understanding of the belief state incorporates the understanding that it may or may not match reality. Considering that our mindreading processes naturally assess the kind of relation that is in play between the agent and the environment, the "additional degree of freedom in belief attribution poses an additional computational burden, which matters because a very significant challenge in explaining mature human mindreading is explaining how it is computationally possible" (Nagel 2013, p. 299).

The explanatory challenge to which Nagel refers is sometimes made explicit by psychologists like Apperly who wonder how mindreading is implemented given that it is not clear how we go from non-obvious external cues to the detection of the relevant mental state. "[W]e do not have direct access to what other people know, want, intend or believe, but must infer these mental states on the basis of what they do and say" (Apperly 2011, p. 1). Comparing belief and knowledge from this perspective, a mental state that essentially reflects a matching with reality is less computationally demanding than a state whose relation with the world is more open-ended. This makes sense when we consider that the verb 'know' appear early and is more heavily used in children's vocabulary than 'think' (Shatz et al. 1983; Bartsch & Wellman 1995) and that this is a cross-cultural pattern (Tardif & Wellman 2000).

On second thoughts it is not surprising that a social agent first learns about knowledge than belief. If a rational agent is trying to understand reality, it makes sense that he first grasps a notion about knowledge, that he learns that sometimes he is right and sometimes he is wrong in his expectations about the environment, in order to grasp the fact that he has this general kind of attitude that essentially can be right or wrong about the environment. Accordingly, it makes sense that in order to a *social* rational agent – trying to also understand how he relates to others and how those relate to the environment – to grasp that others have a general kind of attitude which can be right or wrong about the world, he needs first to grasp, through particular situations, that others form the same correct expectations about the environment than him. The natural processes that allow this agent to acquire knowledge and to track others evidential states would guide his learning about others' knowledge and, later, about others having belief states.[8]

One could resist to these considerations by insisting in an apparent compositional role for belief in knowledge states. For instance, one may appeal to the intuitive argument we mentioned above that one cannot think about knowledge states in which an agent does not have a belief. But we can answer to this argument. Following Williamson (2000), we can concede that knowledge states imply belief states. We can realize that knowledge states have a mental attitude equivalent to beliefs, especially when we look to the mental attitude that is left from cases that fall short of knowledge. But that does not affect our arguments for the simplicity of KNOWLEDGE over BELIEF. We still naturally assess knowledge states as a mental state on their own, and the kind of relation one have to track for detecting belief is still more complex than the factive relation of knowledge. Indeed, we can

[8] This picture is supported by the studies using indirect measures showing that very young children, with 14-18 months-old, for example, presents sensibility to others' knowledge (Moll & Tomasello 2007). The behaviors highlighted by these studies suggest that our grasp of others' knowledge derives in part from innate mechanisms, like the ones related to shared attention, which allow us to track their evidential state.

follow the reversion of status we are defending here and explain the appeal of this argument in other terms. Once you describe a knowledge situation it is easy to infer a belief attitude in it, but this happens only because the description provides the necessary information for belief ascription. If you grant that one knows a certain proposition, you are granting that he has this mental attitude towards the proposition which is equivalent to a belief. But you recognize this only by reflection. In ordinary situations, where you have to infer one's mental state by observational or situational cues, you attribute knowledge or belief by their own. You do not need to start from belief attribution to attribute knowledge.

Furthermore, as we are not endorsing Williamson's stronger claim that knowledge really is a mental state, we have no problem admitting that belief is actually a simpler *state* than knowledge and that this can be a fact cognizable by reflection. We can agree that knowledge, in the end, is composed by the mental attitude of belief and other conditions. But we deny that the natural recognition of knowledge starts with the recognition of belief. Our point is a psychological one: knowledge is naturally seen as a mental state on its own right. As a folk concept or a notion, KNOWLEDGE is actually simpler than BELIEF.

Summarizing the attack to the composite assumption, the main lesson we draw from the psychological literature is that because KNOWLEDGE is an instance of a mental state concept on its own it is not partly composed by the mental state concept of BELIEF. The next step is to investigate what this can tell us about the structure of this concept.

1.3.2. Worries about empirical evidence

There are some worries that can be raised against this lesson and the meaning of the empirical evidence that led to it. First, motivated by the skepticism about animals having mental states, one may doubt that the studies from comparative psychology serve as evidence that reasoning about knowledge is easier than reasoning about belief. Indeed, it is always possible to formulate behavioral rules to explain results like the ones we saw about chimpanzees' ability to track knowledge. There is still much debate on how to interpret these data. Advocates of a more mentalistic interpretation, for example, will argue that behavioral rules cannot explain the flexibility of chimpanzees' behavior in front of new tasks which can be solved by reasoning in terms of mental states (Fletcher & Carruthers 2012). We will not address this kind of arguments here. One interesting answer, however, is to note that there is no reason to insist in explanations in terms of behavioral rules without also doubting about mentalistic explanations of children's abilities, after all, much of the tests with apes are molded from experiments designed to reveal children's reasoning about mental states or theory of mind (Call & Tomasello 2008). Are we willing to reject these explanations in the case of children? For our purposes of understanding which mental domain is more complex, we assume the conditional argument that insofar the experiments on children are interpreted as showing the development of their theory of mind, we can agree that some apes reason about mental states.

Another worry concerns the subject matter of the developmental and comparative literature. In particular, one may wonder if these disciplines are talking about the same thing as philosophers and other psychologists when they refer to KNOWLEDGE. Even if we concede that children and nonhuman animal can reason about mental states, is it the case that we should credit them with KNOWLEDGE? 3-years-olds, for instance, can track others' knowledge to some degree, but they are obviously far from fully understanding how knowledge works. For example, experiments show that even children with 5-6 years-old which have passed false-belief tasks can have trouble to understand how knowledge can be constrained by informational access (Apperly & Robinson 2003). If presented to an object with an ambiguous identity, e.g., an eraser that looks like a die, and to a puppet who only looks to the object, many will wrongly answer that the puppet knows that there is an eraser in the box, a competence acquired only latter. And this kind of problem affects belief attributions too. Also, in the competitive versions of knowledge-ignorance tasks, chimpanzees do no discriminate between conditions where a food was hidden noisily or quietly (Bräuer et al. 2008), suggesting they cannot track knowledge that originates from hearing. Commenting about that, Apperly says that "a core feature of the concept of 'knowledge' is that it provides some unification over the results of a variety of perceptual and inferential processes. If chimpanzees' understanding of 'knowledge' is modality-specific then it falls short of providing this" (2001, p. 53). So it may seem that we should not credit children and chimpanzees with KNOWLEDGE and, therefore, we are not allowed to conclude that it is simpler than BELIEF from the evidence of this literature.

One possible answer to this argument is that we can speak of proto-concepts. No psychologist assumes that children's mindreading abilities are identical to those of adults, so there is no reason for us to assume that their concepts are identical. If those early competences highlight the possession of a concept, it is a proto-concept which will still be developed, but which already contains some of its crucial features. Anyway, once our first understanding of knowledge appears early than our understanding of belief our considerations of parsimony are valid. That is, even if KNOWLEDGE becomes more complex with time, the evidence from developmental and comparative suggests that our understanding of knowledge does not depend on the understanding of belief. Why our latter concept of KNOWLEDGE would depend on BELIEF? Also, there are reasons to assume that BELIEF develops by itself. For example, the kind of problem above related to kinds of informational access affects belief attributions too (Apperly & Robinson 2003). Furthermore, with time we come to learn many different ways in which one can come to belief something. This includes diverse processes of inferences, delusion, wishful thinking, etc. The full understanding of belief plausibly occurs just in maturity. So should we dismiss both children discriminatory competences about knowledge and about beliefs? We know of no good reason to do this.

1.3.3. What we mean by mental state concepts?

So we have reasons to rethink the composite assumption and the subordinate status that it gives to the folk understanding of knowledge. We can assume now that KNOWLEDGE, as a folk concept, is not composed by BELIEF and is a mental state concept. But what does this tell us about our central question, about its structure? Despite the emphasis of this literature on conceptual acquisition, there is a surprising gap between attributing a mental state concept and explaining what it means to have a mental state concept. Apperly expresses his perplexity about this when he claims that such an emphasis "gives a deceptive impression of simplicity. It feels like we know what we mean when we credit a child who passes false belief tasks with the concept of belief, but do we really?" (2011, p. 2). In particular, what does it mean to say that KNOWLEDGE is a mental state concept? What does this say about the structure of KNOWLEDGE?

The answer to that question, however, is not simple. The problem is that it passes through an ongoing dispute about the very nature of mindreading. The two main theories about mindreading, the *theory-theory* and the *simulation theory*, say very different things about fundamental questions such as: How people attribute mental state to others and to themselves? How people acquire mental state concepts? How people represent mental states? Depending on which theory is correct, we have very different answers to the question about the structure of KNOWLEDGE. In the next chapter, we will try to solve the structure question. To do this, however, we have to take a position with regard the fundamental dispute between TT and ST.

Chapter 2

The structure of KNOWLEDGE

Having established that KNOWLEDGE is a mental state concept, our investigation becomes more specifically an investigation about the cognitive capacity of attributing and predicting mental states and behavior to others and oneself. That is, it turns to the nature of our everyday psychological competence, what is commonly called our *folk psychology* or *theory of mind*. These are misinformative terms since they are often used to refer to a particular position about this psychological competence, one that claims that we have a kind of *theory* of behavior. This position, for obvious reasons, was called *theory-theory* (henceforth, TT) (Morton 1980). At the same time, these terms are also informative as they signal the initial dominance of the *theoretical framework* on the topic. This dominance, however, has been disputed insofar the *simulation theory* (ST) has become a prominent alternative to the general view of our psychological competence being constituted by a kind of theory. In order to determine the structure of KNOWLEDGE, therefore, we have to understand the implications of these views regarding mental concepts and evaluate their theoretical merits. This is what we will do in this chapter. In what follows, we will, in a first moment, present the main versions of TT and ST and make explicit the structural hypotheses of these theories regarding mental concepts. Then, we will present and discuss the initial evidence for each of them with respect to the particular case of KNOWLEDGE, and, finally, propose an answer to the structural question. From now on, we will use the term *mindreading* or *mentalizing* to refer to our ordinary psychological competence.

2.1. THE THEORY-THEORY

The TT approach to mentalizing follows a paradigm in cognitive science in which a number of cognitive abilities are explained by the postulation of internally represented knowledge structures, i.e., a set of representations and principles[9]. Those principles consist in rules or laws relating elements of a certain domain and determine the agent's understanding of it. Because the principles operate on representations of things from the domain, those structures are equivalent to an interpretation or *theory* of the domain. Importantly, the idea is not that the theory is represented in an explicit conceptual level. It is not the case that the agent can make his theory explicit or even that it is easy to cognitive scientists to reveal the exact principles being used. Instead, the main assumption is that the theory is in a *subdoxastic level*, that it is only *tacitly represented*, although it can largely interact with doxastic levels. One useful example is the postulation of a folk or *naïve physics* to explain the understanding of laypeople about certain physical domains.

A number of experiments show that many people share a particular understanding of the physical behavior of objects that diverge from Newtonian mechanisms. In one experiment, people were asked to predict the path of a ball with a certain speed at the moment it exits a curved tube through which it traveled (McCloskey, Caramazza, & Green 1980). This tube is held horizontally, so the task allows subjects to ignore gravity. Many subjects, diverging from Newtonian physics, predicted a curved path to the ball instead of a straight path. In another experiment, subjects were given the task to predict the path of a falling object in different scenarios (McCloskey, Washburn, & Felch 1983). When asked to predict the path of a ball dropped by a person walking in a brisk pace, many people judged that it would fall straight down. In contrast, most people judged that a cannonball fired off from a cliff with initial velocity *v* would fall in a parabolic path. However, at the same time subjects most people judged that the ball fired off from a cliff would fall in a parabolic path, subjects were much more likely to judge that a cannonball would fall straight down when it is dropped in the cliff by a mechanism that was carrying it with initial velocity *v*, which is, obviously, an identical physical situation to the previous one. These results were interpreted as showing a "straight down belief", an expectative that objects dropped while being carried fall straight down and that objects moving independently fall in a parabolic trajectory.

Michael McCloskey and his colleagues concluded that what explains these judgments is that people possess a naïve physical theory much like the medieval

[9] "Knowledge structures" here are just equivalent to "information structures".

“theory of impetus”. Roughly, the naïve theory can be summarized as saying, first, that an object set in motion gains an impetus or internal force which responds for the continuity of its motion and, second, that this impetus gradually dissipates. The understanding of the subjects in these cases, therefore, seems to be that just being carried out does not add impetus to an object. Indeed, another common answer in the cliff scenario was that the cannonball would follow a parabolic path for a while and then fall straight down, as if the subjects were judging that what keep the cannonball moving would end during the fall (McCloskey 1983a). Again, the idea is not that one holds this naïve theory in an explicit conceptual level. Some subjects could indeed rationalize their answers and talked about “a force that has been exerted and put into the ball” (McCloskey 1983b), but the naïve physics is postulated as an *intuitive* theory. Naïve judgments are guided by a set of represented principles, but although these judgments are, in a sense, inferential, people are not aware of these principles. Those inferences take place in an intuitive level and people do not have conceptual access to every step of the process. It is a job for cognitive scientists to reveal the principles that constitute folk theories.

Similarly, many philosophers and psychologists claim that our ordinary mindreading abilities are explained by a folk theory of mind or folk psychology. That theory would explain general abilities as predicting other’s behavior, to pick up other’s mood, to choose adequate behaviors or outcomes in social situations, etc. The characterization of Paul Churchland about such a folk psychology is representative of the theoretical approach coming from cognitive science. He says that “each of us understands others, as well as we do, because we share a tacit command of an integrated body of lore concerning the lawlike relations holding among external circumstances, internal states, and overt behavior” (1990, p. 207). According to him, these law-like relations should be understood as a “large number of universally quantified conditional statements, conditions with the conjunction of the relevant explanatory factors as the antecedent and the relevant explanandum as the consequent” (1991, p. 52-53).

Obviously, it is possible to dispute the force of the relations that compose our mindreading abilities. For instance, one can question if they really are law-like or if, instead, they are more like normative principles (Churchland 1991). But this is to dispute the theoretical approach itself. As a general psychological hypothesis, the TT posits that the structures that account for our mindreading abilities are constituted by relations that have the same sorts of commitments of an empirical theory, such as ontological commitments and the possibility of their principles being wrong. A folk theory may tacitly presuppose the existence of certain entities, e.g., an internal force keeping objects in motion, or mental states as beliefs, desires, etc., and it may well be wrong about both the existence of such entities and the exact relations in which those entities are involved in reality. What are the exact formats of these relations (universal generalizations, *ceteris paribus* laws, etc.) and their contents are legitimate and relevant questions to TT, but to assume the theoretical approach to mindreading is to assume that this set of abilities is constituted by the deployment of structures which are equivalent to an empirical theory about the mental realm.

Anyway, there are further important questions that must be answered by advocates of TT and there are different views within the theoretical approach regarding them. In particular, the two most influential views from the psychological literature disagree about how such a theory would be acquired and how to properly interpret the empirical data.

2.1.2. The child-scientist view

Alison Gopnik is one of the main defenders of the developmental view known as the *child-scientist view.* According to Gopnik, the internally-represented structures that account for our mindreading abilities are similar to scientific theories with respect to several features, and not only in having empirical commitments (Gopnik & Wellman 1992). Roughly, the tacit theories that account for our understanding of various domains are also *acquired* and used much like any scientific theory, and this is something we do since our cribs. "[T]he processes of cognitive development in children are similar to, indeed perhaps even identical with, the processes of cognitive development in scientists" (Gopnik & Meltzoff 1997, p. 3). For instance, a scientific theory is *abstract* in the sense that it appeals to constructs that go beyond the evidence that supports it. That is, it posits entities and relations whose properties are abstracted from the more apparent properties of evidence, e.g., species, physical forces, viruses, structures, etc. It presents *coherence* as the entities postulated by it are closely interrelated with one another. Related to these two features, it appeals to *causality* to explain the more apparent properties of evidence and the interrelation between the entities of the theory. Thus, all these characteristics would be present in children's folk theories and, in particular, "these characteristics of theories ought also to apply to children's understanding of mind, if such understandings are theories of mind. That is, such theories should involve appeal to abstract unobservable entities, with coherent relations among them" (Gopnik & Wellman 1992, p. 148).

According to this developmental picture, we manage to form a theory of mind through the same learning mechanisms that support all cognitive development, i.e., through *domain-general* mechanisms. In the specific account of Gopnik and Laura Schulz (2004), we abstract causal information from patterns of statistical data through *graphical causal models* or *Bayes nets* (Pearl 2000; Spirtes, Glymour, & Scheines 2001). Another important point of the child-scientist view is that folk theories would also share the *dynamic properties* of scientific theories. So, after acquisition, a theory is subject to *theoretical change* in the face of the accumulation of counterevidence. Indeed, this is how the view interprets a number of developments regarding mindreading.

Josef Perner (1991), for example, claims that what explain the gap of competence we see between 3-4 years-olds and of 4-5 years-olds regarding false-belief tasks is a matter of theoretical change from a theory which does not differentiate between *pretense* and *belief* states to a theory that does. It is a shift between a theory which understands that one can act as if something was something else, e.g., one can act as if a banana were a telephone, to a theory which grasps that one can be in a state that in essence can *misrepresent* things in the world. This understanding would be acquired altogether with children's acquisition of a general concept of representation. "Young children fail to understand belief because they have difficulty understanding that something represents; that is, they cannot *represent* that something is a *representation*" (1991, p. 186). The child-scientist view, therefore, posits that acquiring mindreading is in part a matter of acquiring a *metarepresentational theory*, a theory about the existence of mental states and what they are.

2.1.3. The modularity view

In contrast with the child-scientist view, the modularity view claims that we do not acquire a metarepresentational theory at all, but that mentalizing is a matter of maturation of a *specialized cognitive module* or *core system*[10]. According to this position, instead of being acquired, revised and stored just like a scientific theory, the information that postulates mental states is stored in one or more *innate* modules. Jerry Fodor (1983) influentially describes a cognitive module as an input system with most or all of the following criteria: (1) they are domain specific, i.e., they operate exclusively on certain types of input; (2) their operation is unconscious and there is limited central access to their representations; (3) their operation is fast; (4) they are *informally encapsulated* in the sense that they are not affected by the information contained in other mental systems; (5) their operation, given relevant input, is mandatory; (6) their outputs are shallow in the sense that the information they carry are simple and serve low-level cognition; (7) they exhibit *pathological universal*, i.e., specific breakdowns patterns; (8) they exhibit *ontogenetic universals*, i.e., specific developmental pace. It is currently doubted whether modularity requires most of these features and different definitions propose different combinations of them (Elman et al. 1996). Anyway, the existence of specialized and independent devices is generally accepted in cognitive science.

In particular, these devices are postulated to explain a number of *innate* capacities. Although a cognitive module typically involves maturational processes, the information and representations that constitute its operations are not empirically acquired. We can make an analogy with a language module. Following Noam Chomsky's (1957, 1965) original hypothesis that language acquisition is facilitated by an innate device, many cognitive scientists postulate modules that would be responsible for the learning of natural language syntax, morphology and phonology. Steven Pinker (1984), for example, provides a detailed account of a syntactic module. The problem with the origin of syntax is that although children are born

[10] Considering the common characterizations of modules and core systems, these are very similar notions and seem to refer to the same phenomena. However, just as there are different specific definitions of modules and core systems, it is possible to distinguish them (Spelke 1988; Carey 1995). Anyway, for our purposes, we will use the terms interchangeably.

without language, at about 2 and a half years and 3 years-old they already manifest some competence to form new sentences by their own. How children manage to learn such an abstract and complex thing as syntactical categories and rules? Pinker's modularist response is that they do not learn syntax from the ground up. Instead, we have an innate module which already contains syntactic categories and constrain the possible constructions of sentences. This saves children from the wide demand that they be little linguists and allows them to learn the grammar of a specific language. Similarly, a modular view of mentalizing posits that *mental categories* and *principles* can be stored in innate module which allows one's acquisition of mindreading abilities.

Alan Leslie and Simon Baron-Cohen are two important defenders of the modularist view. The inspiration for their view comes from evidence regarding autistic children. Along with the developmental pattern from false-belief tasks, another major find in the developmental literature is that autistic children with mental age exceeding 4 years fail in this test, whereas children who also have a disability and with identical mental age, as children with Down syndrome, can succeed it (Baron-Cohen, Leslie, & Frith 1985). In the experiment of Baron-Cohen, Leslie and Uta Frith, 84% of preschool children and 85% of children with Down-syndrome succeed it, while inversely, 80% of autistic children failed it. What is the reason for such a high contrast? Previously, Leslie was interested in the problem of how children are capable of the puzzling ability of *pretense* and *identification of pretense*, an ability they can present already at 18-months-olds (Dunn & Dale 1984). This early capacity requires complex *representational* abilities. They are not only able to represent aspects of the actual situation ('this is a BANANA') and of the pretense situation ('this is a TELEPHONE'), but relate them in specific ways ('is this banana that is a telephone'). Furthermore, they are often able to identify when other children are engaging in pretend play and what the contents of their pretense are. They do not see certain situations as purely behavioral, but interpret them in mentalistic terms, in terms of pretense. Noting the striking similarity in the semantic structures of mental states expressions and pretense (from a logical point of view, both contain embedded propositions and do not present normal reference and truth relations), Leslie (1987) proposed that the same mechanism respond for the representational demands of pretense and understanding of mental states. He postulated an innate *representational mechanism* that normally matures between 18 and 24-months, the *theory of mind mechanism* (ToMM). Besides providing the necessary representational abilities for engaging in pretense, the ToMM directs children's attention to relevant behavior and use representations of mental states, *M-representations*, to represent them. Putting together these results and also the fact that autistic children lack pretend play (one of the behavioral grounds for autism diagnosis), Leslie argued that what explains autistic children's poor performance in false-belief tasks is precisely an impairment in the ToMM.

Another component of Leslie's view is a non-modular mechanism of *inhibitory control* he calls *selection processor*. If ToMM maturates in the second year of life, why children with 3-4 years-olds fail in false-belief tasks? Do they not already possess the concept of belief? While the child-scientist view interprets the lack of competence in this task as a lack of theory, Leslie's answer is that it is a matter of *performance*. The representational demands of the task involve children's representation of the agent being in a belief state, the representation of the content of the agent's belief, and his representation of the actual situation. The representation of the agent being in a mental state is accounted by ToMM, but correctly selecting the content of the agent's belief relies on more general problem-solving processes, as inhibiting the representation of the actual situation and, instead, memory consulting of the situation to which he was exposed. The inhibitory capacity is a relatively late developed executive function and, being related to several cognitive tasks, it is not a specialized mechanism. In that respect, Leslie's view is akin to many who claim that the failure in false-belief tasks are importantly related to *processing* matters (Lewis & Osborne 1990; Mitchel & Lacohée 1991; Zaitchik 1991; Lewis et al. 1994; Carlson, Moses, & Hix 1998). Difficulty with false-belief tasks, therefore, can be explained without us having to postulate a theoretical lack or change about mental states. According to the modularity view, 3-years-olds would already think about belief and other mental states despite the evident difficulties regarding certain mentalizing tasks.

Finally, the idea of the modularity view is not that everything we need to know about mental states is already stored in an innate representational module. Leslie's proposal is that the ToMM works as a mechanism of *selective attention*, allowing the brain to attend to mental states and their properties by providing the agent with mental states representations. According to the modularity view, mindreading development is dependent on the representations of an innate module, but Leslie promptly admits that adults' mindreading abilities may be significantly more sophisticated than children's abilities. Brian Scholl and Leslie say:

> [W]hen we talk about ToM in the context of the modularity theory, we intend to capture only the origin of the basic ToM abilities, and not the full range of mature activities which may employ such abilities. It is certainly the case that these basic ToM abilities may eventually be recruited by higher cognitive processes for more complex tasks, and the resulting high-order ToM activities may well interact (in a non-modular way) with other cognitive processes, and may not be uniform across individuals or cultures. (1999, p. 140)

2.2. SIMULATION THEORY

The general notion of simulation, in terms of *mental mimicry* or *empathy*, is already present in the philosophical literature since the nineteenth-century in the work of philosophers such as David Hume, Adam Smith and Friedrich Nietzsche, and in the twentieth-century with Theodor Lipps and, more recently, Willard Quine. However, it was only with Robert Gordon (1986) and Jane Heal (1986) that simulation (or "replication") was proposed specifically as a hypothesis regarding folk psychology and as an explanatory alternative to TT. As such, ST explains the same phenomenon accounted by TT, or part of it, but positing a mechanism that does not require theoretical information for that. ST claims that mindreading requires no internally represented structures about both what mental states are or about psychological rules. Its advocates find especially doubtful that what explains mindreading are folks acquiring psychological generalizations and metarepresentations, but they also doubt the need for innate metarepresentations. The central thesis of ST is that mindreading is achieved by a particular use of the *agent's own cognitive apparatus* (Gordon 1986). Roughly, we form predictions and explanations of someone's mental states by "putting ourselves in another's shoes", running our own cognitive mechanisms and seeing the resulting psychological states. In other words, we do not need to develop a folk psychology about the working of others' behavior because we have a perfect behavior-producing mechanism ourselves and we can use it as a model to understand others. Furthermore, because we share similar cognitive mechanisms and principles, to use this model allows successful prediction of others' psychological states.

As a kind of psychological mechanism, simulation provides a radically different picture of the processes underlying mindreading in comparison to the theoretical approach. But how does simulation exactly works? The prototypical example of simulation is that in which one makes a *third-person attribution* of a certain psychological state. Take the example of Mr. Tees and Mr. Crane used in an experiment from Daniel Kahneman and Amos Tversky (1982) on counterfactual reasoning.

> Mr. Crane and Mr. Tees were scheduled to leave the airport on different flights, at the same time. They traveled from town in the same limousine, were caught in a traffic jam, and arrived at the airport 30min after the scheduled departure time of their flights. Mr. Crane is told that his flight left on time. Mr. Tees is told that his flight was delayed and just left five min ago. Who is more upset? (Kahneman & Tversky 1982)

Unsurprisingly, the overwhelming majority of subjects (96%) answered that Mr. Tees is more upset than his limo colleague. More relevant, simulationists see this piece of mindreading as a representative case of simulation. Assuming their interpretation, how we are able to simulate Mr. Crane's and Mr. Tees's states and compare them? One obvious obstacle is that we are not really in their situations. We are not in a limo on our way to the airport trying to catch a flight, and neither are we in any of their specific situations of delay. More generally, in order to properly predict the resulting state of someone, simulation requires a way to use other's relevant initial states as input. Accordingly, one essential aspect of the simulationist proposal is the fundamental role it confers to *imaginative processes* in the mindreading of cases like this. Alvin Goldman (2006), who has provided the more detailed defense of ST so far, proposes the notion of *enactment imagination* or *E-imagination* to account for what is the first step of this kind of simulation. To e-imagine a certain state is to recreate the *feeling* or what is like to experience that state. For example, to e-imagine being late to catch a flight or to miss a flight involves an imaginative process that creates a pretend state of being late or missing a flight that phenomenally resembles the actual states. The way we overcome the initial interpersonal distance, therefore, is by generating *pretend states* that are relevantly similar to those of the target.

In a second step, after creating pretend states of the relevant initial states of Mr. Tees and Mr. Crane, we can just run them into our own cognitive system for, lastly, check what their resulting states are like. Similarly, to predict someone's decision or epistemic state about a certain matter, we create pretend states that enact his initial states and we run them into our own decision-making mechanism. These initial states can include propositional attitudes themselves, as desiring, believing, knowing, doubting, etc. Obviously, however, we do not process these pretend states as we normally process the inputs we find in tasks not related to mindreading. Typically, in real life situations, cognition about a certain situation is related to practical matters. In face of a certain situation *s*, we run the situational inputs in our cognitive system and besides form particular mental states regarding *s*, we take a decision on how to act regarding *s* and we actually act. In imaginative simulations, therefore, we have to make the system "*off-line*", disconnected from our action-controllers. Computationally, thereby, simulation just requires the cooptation of existing mechanism, instead of the computation of an entire body of information – Gordon (1986) and Goldman (1995) indeed use this idea as an argument from parsimony in favor of ST. Another important aspect of this step of simulation is the necessity of "*quarantine*" or *inhibition* of the agent's own states when running his cognitive system. (Goldman 2006). The agent's own mental states must not interfere in the process or else it may no longer resemble the target's processes. Importantly, what becomes a source of evidence of simulation, failure to do so leads to an *egocentric bias* by the agent – we will return to this point latter.

The last feature in simulating the states of Mr. Tees and Mr. Crane is attributing the resulting state of simulation to *them*. This is the main source of disagreement between proponents of ST, however. On Goldman's proposal, third-person mindreading is fundamentally dependent on *first-person mindreading.* In particular, he claims that people's understanding of mental states is preceded and grounded by the awareness of their own states. He calls his position about first-person attribution the *special method view*, and the idea is that one can detect one's own mental states *introspectively* or by *self-monitoring* (Goldman 2006). So, in third-person mindreading, before representing the resulting state as being the target's state, the agent *accesses* the product of his simulation. In our particular case, we predict that Mr. Tees is very upset by introspectively accessing the resulting state of the simulation of his situation. Being a piece of e-imagination, the resulting state has a distinct phenomenology which can be accessed by the simulator – although, Goldman's (2006) proposal emphasizes the use of *neural properties* as the proper input for mental states detection.

Gordon (1995; 1996), however, disputes this fundamental role of first-person mindreading. Gordon opposes to most "traditional" simulationist accounts of mindreading which draw simulation as an analogical inference from oneself to others. He claims that instead of an implicit inference from pretend states, simulation requires an "*egocentric shift*" and, accordingly, he uses the slogan that simulation is "not a transfer but a transformation" (1995, p. 54). What one would pretend in imaginative simulation is not that one is in someone else's shoes, but that *one is* the owner of the shoes himself. In our particular case, he says, "I have the option of imaging in the first person Mr Tees barely missing his flight, rather than imaging myself, a particular individual distinct from Mr Tees, in such a situation and then extrapolating to Mr Tees" (p. 55). This allows the agent to rule out one step of the simulation. Instead of having to make an analogical inference from himself to the other, he becomes the target itself. More radically, simulation would require no grasp of mental states. To defend this, Gordon appeals to a procedure he calls "*ascent routine*" in which one transforms a mental task into a lower semantic question. He believes this is a very common procedure. For example, if someone is asked "do you believe that penguins can fly?" he can just ask himself "do penguins fly?" and his answer would need no understanding of beliefs. Similarly, after transformation, if asked whether "S beliefs that *p*", one can just express "one's" propositional attitude regarding *p*.

2.3.1. High-level and low-level mindreading

Although it is the paradigmatic example of mindreading, the kind of imaginative simulation we presented so far correspond to what Goldman (2006) thinks constitutes only *one kind* of mindreading. When ST first emerged as an alternative to TT it was mainly concerned with the debate around folk psychology, whose paradigmatic states are propositional attitudes as beliefs and desires, and the kind of cases that imaginative simulation accounts for. This view gained even more attention when psychologist Paul Harris (1992) initiated the discussion about the origins and the role of simulation from the developmental perspective. Anyway, a major boost in favor of ST was caused by the late neuroscientific findings regarding *mirror neurons* (Gallese & Goldman 1998; Currie & Ravenscroft 2002; Decety & Greze 2006; Goldman & Sripada 2005; Gallese 2007). Mirror neurons are a group of neurons discovered in monkeys' premotor cortex and which fire both when a monkey performs an action as when it sees another agent performing the same action. Their importance for ST is in the suggestion that they may be not only the neurological basis of imitation, but of "a more general mind-reading ability" (Gallese & Goldman 1998, p. 493), i.e., simulative mindreading. Unlike some of his cognitive scientists colleagues (Gallese, Keysers & Rizzolatti 2004), Goldman (2006) does not defend that mirror neurons are the basis of all social cognition. Instead, he proposes that they are related only to a kind of simulation distinguishable from the imaginative simulation we saw so far. His proposal is that simulation should be divided between *high-level mindreading* and *low-level mindreading*.

The distinction between low-level and high-level mindreading is more characteristic than definitional, where the latter is described by Goldman as being about states of a more "complex nature such as propositional attitudes" (p. 147), being more subject to voluntary control and more accessible to consciousness in comparison to low-level mindreading, which "is comparatively simple, primitive, automatic, and largely below the level of consciousness" (2006, p. 113). While imaginative simulation falls into Goldman's characterization of high-level mindreading, mindreading based on mirroring or "*unmediated resonance*" constitutes his paradigmatic example of low-level mindreading. Roughly, the idea is that mirroring allows an agent to see certain kinds of behavior as meaningful – instead of purely behavioral – because he *experiences* them in some level, e.g., seeing someone hopping on one leg after kicking a stone. Exploring the neuroscientific evidence on mirroring systems in humans, Goldman then tries to link mirroring to mindreading. Crucially, he limits his case to *emotions*, *feelings* and *intentions*, relying on evidence showing, for example, overlap brain activation when one fells disgusted by smelling or tasting unpleasant odors and tastes and when they observe faces expressing disgust (Phillips et al. 1997), when one feels pain and see another feeling pain (Singer et al. 2004), and when one performs and observes an agent performing the action (Grèzes & Decety 2001). This distinction is important for pointing to the possibility of simulation being achieved by different mechanisms.

2.3.2. Simulation and false belief

Similarly to the modularist view, ST explains 3-4 years-old children's lack of competence regarding false-belief tasks not as a problem of theoretical or representational nature, but as a matter of performance. In other words, the failure indicates that the task is beyond children's *abilities*. The older children would just acquire the ability to simulate the state of someone whose perspective upon the world is different from her perspective. For instance, as we already said, to succeed in simulating the false belief of the target in those tasks, the children must *inhibit* her own beliefs about the situation, something for which she is still struggling. Thus, as far as simulation accounts for false beliefs, success on false-belief tasks would follow, at least, the development of inhibitory control.

ST, therefore, also finds support on the evidence mentioned above suggesting that competence in false-belief tasks is related to *information processing*. For example, some experiments show that 3-years-olds can pass modified versions of the false-belief task which change the *salience* of the events involved in the task. In one study, Deborah Zaitchik (1991) tested two new conditions: the "seen" and the "unseen" condition. In the seen condition, the doll of a bird shows the child the location of a toy in a particular box and tells her that it will tell to the frog a lie, it will tell the frog that the toy is in another box. In the unseen condition, the same story is used, but this time the location of the toy is not showed to the child, she is only told where the toy is. When asked where the frog will think the toy is, most children in the seen condition fail the test, while in the unseen condition children tend to succeed. Zaitchik explanation is that in the seen condition the actual location of the salient toy is just too salient to the child. In another experiment, Chris Moore and colleagues (Moore et al. 1996) increased the inhibitory demand in a task structurally similar to the false-belief task, but regarding desire instead. They created a game where the child plays with a competitor and where both have to draw cards from a desk in order to complete a puzzle. At one point, both have to draw a red card, and then, after in fact drawing the red card, a blue card. So, their initial desire for the red card changes when they get it, and the desire becomes a desire for the blue card (resembling the structure of false-belief tasks). The child is the first to draw the red card, so her desire changes. At this point, she is asked about the desire of Fat Cat (his puppet competitor), who still did not get the red card. The result is that 3-5 years-olds perform at this test as poorly as in the false-belief task, something with is easily explained in inhibitory terms. All those experiments present evidence of a role for inhibitory control in mindreading. However, as far as those experiments do not show that the specific processes at issue are simulations, they favor both ST and the modularity view.

2.4. THEORIES AND MENTAL STATE CONCEPTS

Now, what these views tell us about the structure of mental states concepts? At first sight, it seems like we can get a straight answer from TT. For, although TT originates as a theory about our understanding of human psychology, it also constitutes one of the main theories of concepts besides the prototypical and exemplar theories. Roughly, as a general theory of concepts, TT holds that concepts are *theory-like* entities. Thus, its immediate answer is that mental concepts have theory-like structures. Even though it is unwise to take the general claim of TT in its strongest sense – as if every concept has a theoretical structure – given the solid evidence in favor of the existence of prototypes and exemplars, we can still keep the straight answer from TT in the specific case of mental concepts. The reason is that TT's framework is especially suitable for abstract domains. Indeed, part of what led it to also constitute a general theory of concepts is the same insight which originally motivated it as a theory of psychological reasoning, *viz.,* that theories can abstract from evidence and encode information about hidden properties. Non-observable properties and entities can be addressed by the kind of causal or explanatory relations which appears in theories. A theoretical approach to concepts, therefore, is *prima facie* appropriate for the category of abstract concepts, which includes mental concepts. Thereby, as far as mental concepts are theory-like entities, they have a theory-like structure.

TT's initial answer, however, is not as straight as it may seem. Most theory theorists are not clear about what they mean with the central tenet of TT, but to say that mental concepts have a theory-like structure is a vague statement, and there are different interpretations available. It is doubtful, for instance, that Gopnik, Perner and modularists as Leslie understand "theory" and "concept" in the same way. Modules can be considered theory-like because they store representations and principles which make implicit ontological presuppositions and are empirically committed. So, it seems like mental concepts are already present in modules. But they also are essentially meant for learning, so a number of the information we have as adults and which we use to think about mental states is acquired due to modular mechanisms. Should that information be considered theoretical? If so, this theoretical information is part of mental concepts' constitution? Gopnik and Perner, on the other hand, see the acquisition of mindreading as a matter of acquiring a metarepresentational theory of mind and emphasize that understanding mental states consists in the deployment of theoretical information. Mental state concepts, therefore, are not acquired until the acquisition of such a theoretical knowledge. But, what exactly is the relationship between concepts and theories?

In its weakest interpretation, in the sense that distinguishes TT from the prototypical and exemplar views of concepts, to say that a concept has a theory-like structure is merely to say that it stores information which is distinct from sensory information. It is to say, in particular, that it stores abstract information which is concerned with the *understanding* of the world. So, instead of having only sensory information and categorization processes based on similarity, we have information structures providing explanatory content of certain aspects of the world.

This can hardly be considered controversial. We are obviously capable of explanation-based reasoning, and if concepts are the constituents of thought, concepts must store theoretical information. This is also hardly informative, however. Such an idea poses no constrains on both what should be considered a mental theory, equating every piece of explanatory reasoning to theoretical reasoning, and on the exact relationship between concepts and theories. Under this idea, for instance, the classical view of concepts is integrated to TT, for definitions can be considered theories. Furthermore, if what determines a mental theory is so inclusive as to include any explanatory reasoning, then the resultant information acquired via modules counts as theoretical information. But, again, what is the relationship between concepts and theories?

We can find two answers in the literature on concepts regarding that question, and one of them is from Gopnik herself. Some psychologists talk about concepts *being* theories (Rips 1995; Rehder 2003, 2003a), while others talk about concepts being *elements* or *terms* of theories (Carey 1985; Gopnik and Meltzoff 1997). One source of confusion is the fact that most psychologists are not really clear about their specific positions or even seem to slip from one position to the other (Murphy & Medin 1985). Anyway, according to the concepts-as-theories interpretation, the relation between concepts and theories is one of identity. Assuming this to our case of interest, to have a mental concept is to have a theory or a "mini-theory" about a mental state. According to the concept-inside-theories interpretation, however, concepts relate to theories at a different level. Inversing the relation, concepts are constituents of theories. A concept, in turn, is constituted by the *roles* it plays inside a theory. Susan Carey is especially and exceptionally clear about this when she says that "[o]ne solution to the problem of identifying the same concept over successive conceptual systems and of individuating concepts is to analyze them relative to the theories in which they are embedded. Concepts must be identified by the roles they play in theories" (1985, p. 198). Thus, a mental concept would be constituted by its explanatory and inferential roles, by its associative relation with other concepts inside the theory in which it is embedded, and so forth.

Psychologists holding the child-scientist view, such as Gopnik, talk about children acquiring intuitive theories of certain domains, including an intuitive theory of the mental realm (Gopnik & Meltzoff 1997). Their account of the evidence around false-belief tasks is in terms of a conceptual change that goes from a theory of the mental on which false beliefs are not predicted to a (metarepresentational) theory of the mental on which they are. One natural reading for the child-scientist view, therefore, is in the terms of the concepts-inside-theories. Perner, for example, says that "each particular mental concept gets its meaning not in isolation but only as an element within an explanatory network of concepts, that is, a theory (1991, p. 109). Accordingly, mental state concepts are terms of a larger theory (of the mental realm), and they consist in the inferential, explanatory, associative, etc., roles they play inside this theory. Despite this natural reading, however, one can easily dismiss the distinction between concepts taken as elements of theories and as theories themselves. Leslie (2000), for example, says that Perner "(…) is, however, committed to the child acquiring an explicit understanding of belief-as-representation, to the notion of conceptual change, (…) and, therefore, to the idea of *concept-as-theory*" (2000, p. 7, our emphasis). The reason for this is not necessarily that the distinction is just ignored by some psychologists, but instead that it is still not clear what counts as a theory on the TT approach. In particular, if we assume, as many theory theorists do, that a central aspect of a theory is being about a certain domain, in the absence of other criteria for theories, it is not obvious that BELIEF, for example, does not constitute a theory itself. If we take the domain of the mental as primary, BELIEF is an element in a theory about a larger domain, but if we are interested in the domain of beliefs, can we not consider BELIEF as a mini-theory in itself? Furthermore, Leslie interprets Perner's position as an instance of a descriptivist account in which the meaning of a concept is the entity that is "picked out" by its descriptive content. On this interpretation, indeed, BELIEF is equivalent to a theory of what beliefs are.

Although it does not provides a complete account of what a theory is, the child-scientist view provides a more restricted notion of mental theory than the one provided by the weaker interpretation of TT as a theory of concepts. In contrast, on the perspective of a theoretical approach to concepts, it is now less clear what a modularist view as Leslie's theory has in common with the child-scientist view. For Leslie, mental state concepts are *prior* to the accumulation of information about this concept. Previous representations of mental states are what allow children to identify mental instances of mental states and learn from them. Contrasting the modularity view with the child scientist view, Tim German and Leslie say:

> [E]arly developing abstract concepts are much more likely to depend upon mechanism rather than upon knowledge [theory]. A cognitive mechanism may play the role of enabling, and even directing, attention to a particular property or set of properties which then become a topic for knowledge acquisition. On this view, concept possession is prior to knowledge. Consequently, a concept may be innate without innate knowledge. (German & Leslie 2001, p. 80)

Assuming that the late information that one acquires about a mental state can be considered theoretical, one possible interpretation, therefore, is that mental concepts, for the modularist view, are independent of theory. A mental concept is a theory-like entity only in the sense that the representations and principles stored in a module may resemble a tacit empirical theory, but this obviously contrasts with the views from Gopnik and Perner. So, if BELIEF, for example, has a theory-like structure, it is in an entirely different sense that BELIEF has a theory-like structure on the child-scientist view. Since one of the central tenets of the representational theory of concepts is that a concept is the mental unity which responds, among other things, for our categorization processes, however, this distinction between a mental concept and the information related to it may be blurred. Assume, for instance, that the information acquired as a result of a modular system responds for a large number of our attributions or categorizations of mental states. Is it not correct to say that this information is part of the constitution of mental concepts? If those acquired structures are theory-like, are not the innate mental concepts embedded in them? We are inclined to think so, but we will not try to solve these issues here.

To circumvent so many complications, we can do what most psychologists do in order to point to the evidence in favor of TT as a theory of concepts, *viz.*, to focus on the *kind* of information which is stored in a theory. Since theories account, for example, for our nomological, causal, modal, and functional knowledge of things, it is evidence in favor of the hypothesis that a concept has a theory-like structure that it stores nomological, causal, modal, or functional information (Machery 2009). For our purposes here, we will assume this default interpretation of the theory-theory of concepts for the specific case of mental concepts. We can finally turn to our case of interest. Following TT, therefore, another structured hypothesis arises with respect to KNOWLEDGE. In particular, (H2-*c*) says that KNOWLEDGE has a theory-like structure (in the sense that it stores causal, modal, functional, or nomological information).

2.5. SIMULATION AND MENTAL STATE CONCEPTS

So TT provides an alternative structured hypothesis in relation to the ones seen before. But what about ST? Assuming ST as the correct account for mindreading, what does it tell us about the structure of mental concepts? Similarly to TT, ST provides an immediate initial answer which turns out to be non-obvious. One crucial aspect of the simulationist proposal is that, since we use or own cognitive apparatus as a model to predict others' states, simulation dismisses the necessity of internally represented structures about what mental states are or about psychological rules for mindreading. One can predict and attribute mental sates to others without storing information about mental states. This is a view radically different from the theoretical approach to both mindreading and mental concepts. One initial answer, therefore, is that in fact we may have no mental concepts at all. Due to simulative processes, we manage to mentalize without making use of mental concepts. This would be a striking result, but we think it is just wrong. For instance, to conclude this, we need to assume that the theoretical account of concepts, in general, is the correct one, and that is something which is in dispute. ST is in sharp contrast with TT in being an "information-poor" approach to mindreading, while TT is "information-rich" (Goldman 1995), but it does not follow from that, that there are no mental concepts or that no concepts are used in simulative processes.

We obviously can think about mental states. We make attributions, talk about them, and we even think about them in counterfactual situations. To do this we use some sort of representation and, assuming just the default view of concepts, we can say that we do have concepts which are about mental states. Indeed, this works against Gordon's radical position in which there is no room for grasping mental states in mindreading. Assuming that ST is the only mechanism for mindreading, as we are doing, how would one be able to think about mental states without any mental symbol? Every piece of reasoning about the mental would be an instance of accent routine? That is implausible. Goldman's proposal of simulation, in contrast, does not need to face this problem. Remember that according to his theory, first-person attribution plays an essential role in mindreading. Roughly, one needs to introspectively detect one's own mental states in order to attribute them to the other. He says:

> My central thesis is that mental concepts (partly) employ *introspection-derived*, or *introspection-associated*, mental representations. The hypothesis is that there is a proprietary code, the *introspective code* (I-code), used to represent types of mental categories and to classify mental-state tokens in terms of those categories. (2006, p. 260)

In other words, the I-code consists in properties which are accessible to introspection. Instantiation of mental states also instantiate those properties, and an agent classifies what mental state he is in virtue of them. Goldman uses "partly" because he does not exclude the possibility of other kinds of representations also being part of the representations that are used to classify a mental state. However, he is careful to say that the "vast majority of self-attributions of current mental states, I suspect, use introspection only" (2006, p. 263). Given that third-person attribution fundamentally depends on first-person attribution, one possibility, then, is that certain mental concepts consist only in introspective classification. Assuming this is true for DESIRE, for example, instantiations of DESIRE consist only in certain mental states tokens (desire states) being recognized by the agent as pertaining to the same state kind. Given that introspective classification is a "perception-like process" (Goldman 2006, p. 246), DESIRE is essentially a *recognitional concept*.

A recognitional concept is a concept whose possession conditions include a recognitional requirement according to which one needs to have the ability to recognize instances of the things that fall under that concept (Fodor 1998). Importantly, one needs to possess no descriptive content or informational structures in order to possess a recognitional concept. A concept like RED, for example, would be a recognitional concept because it includes the requirement that one recognizes instances of red in order to possess RED. But one does not need to store any descriptive or informational structures about what red things are in order to possess RED. Furthermore, even if a concept is constituted only by a recognitional ability, one can still talk about it, think about it, or use it to compose new thoughts or concepts. The meaning of this concept is just the entities or the class of entities to which it refers. Thus, when one thinks and talks about DESIRE, one refers to the states one recognizes as desire states.

Applying this idea to KNOWLEDGE, a new and surprising hypothesis arises: (H1-*a*) KNOWLEDGE is a recognitional concept. That is, KNOWLEDGE may consist only in the ability to recognize and classify certain instances of mental states as pertaining to the same kind of mental state. Assuming Goldman's proposal of I-code, KNOWLEDGE may consist only in a primitive representation, which we shall call *k-states*[11]*,* of a class of mental states. So, even abandoning ST's first answer, we still came up with a radically different hypothesis from the simulationist approach. Contrasting with all hypotheses we saw so far, we found a non-structured hypothesis: what underlies our ordinary understanding of knowledge is a primitive concept. No structure, however, does not mean no concept. This is a radical change of perspective for anyone who assumed that the term 'understanding' fundamentally involves some sort of substantive internal content or a body of stored information. If the (H1-*a*) is right, we simply do not need such a substantive content to ordinarily think about knowledge. KNOWLEDGE consists only in the ability to discriminate *k-states*, one that is supported by introspection and simulative processes.

(H1-*a*) it is not the only hypothesis that can be drawn from ST. First, still following the I-code proposal, it is possible that KNOWLEDGE is a recognitional concept and that it is constituted by more than one primitive representation. In that case, KNOWLEDGE would be constituted by a *schema* which connects those representations in a single structure. Thus, we have another structured hypothesis: (H2-*d*) KNOWLEDGE has a recognitional structure. Note, however, that although (H2-*d*) differs from (H1-*a*), it still radically differs from all the structured

[11] We do not take k-states in the same vein of Williamson (2000), i.e., as identical to knowledge states. k-states are the class of mental states classified by our cognitive system as a proper mental state.

hypotheses we saw before, for, in a sense, KNOWLEDGE would still be informatively primitive. For this reason, in the following sections we will not distinguish between (H1-*a*) and (H2-*d*). Second, there is no need to assume a radical position regarding ST. Goldman himself adopts a *hybrid theory* in which we can sometimes use theorization to infer a mental state. Being a simulationist, however, he claims that our "fundamental or default practices (…) are of a projective, or simulative, character. Absent special circumstances, we presume our own mental contents to be suitable in kind to match those of our targets" (2006, p. 176). Anyway, if we assume a hybrid view regarding knowledge categorizations, a natural step is to conclude that although KNOWLEDGE is a recognitional concept, it also has theoretical structures in its constitution. This is a valid possibility. In that case, we would achieve a pluralist conclusion, that is, (H2-*c*) plus a primitive or structured representation of k-states which is used in introspective process. What we must do now is to evaluate the evidence in favor of TT and ST with respect the particular case of KNOWLEDGE.

2.6. KNOWLEDGE CATEGORIZATIONS: SIMULATION OR THEORETICAL GENERALIZATION?

Let us start assessing the radical hypothesis from ST. How plausible is (H1-*a*)? How plausible is the idea that our ordinary understanding of knowledge consist only in recognizing a certain kind of mental states? To answer that, we need first to clarify the notion of "understanding" in play here. As we tried to make clear while motivating our investigation, what we have assumed so far with respect to concepts were the default assumptions from psychology of concepts. In particular, concepts are taken to be those mental unities responsible for a number of cognitive processes like categorization, inference, explanation, learning, etc. The structure question concerns these processes. Depending on our patterns of inference, categorization, learning, etc., we get a glimpse of how the information stored in a concept is organized. Given the typical features of the epistemological activity, we have focused on categorization, but the content of a concept need not be restricted to the information involved in categorizations processes. As we also mentioned, there is no reason to think that only one structure must underlie the cognitive processes related to a concept. The answer to the structure question can be a pluralist one. Anyway, regarding the relevant question here, if the information stored in a concept *C* is organized in only one structure, then the conceptual understanding of *c* consists in the content of that structure. Therefore, if (H1-*a*) is right and KNOWLEDGE consists only in a primitive recognitional concept, our conceptual understanding of knowledge consists only in a recognitional capacity.

One may initially think that (H1-*a*) is implausible because it empties too much our concept of knowledge. That is, it implies that we have no substantive understanding of knowledge states, but only a primitive way to categorize them. But do we really have reason to think otherwise? One may argue, for instance, that this hypothesis does not account for the apparent fact that we understand some crucial features of knowledge states. For example, we seem to understand that knowledge is a factive state, i.e., we seem to understand that one cannot know something that is false. If categorizations of knowledge depended only in identifying certain internal states (k-states) then we could not understand that knowledge implies truth, which is an external property. We do not think this objection is convincing, however. It is possible to motivate (H1-*a*) in such a way that it accounts for our "understanding" of crucial aspects of knowledge states. To do this, it is important to distinguish between the categorization of k-states and the *working* of k-states.

2.6.1. Tacit understanding, egocentrism and normativity

The idea of k-states follows the thesis defended in the last chapter that there is a class of mental states which are recognized as being a mental state kind on their own. If a simulationist theory like Goldman's is right, then we have a way to identify those states introspectively. Those who are skeptical about the privilege of introspective processes may doubt that k-states may be introspectively accessed. Much less disputable, we claim, it is the idea that our *cognitive system* recognizes or represents certain internal states as knowledge. To use the vocabulary of artificial intelligence, humans are "information processors", machines which have a proper way to regulate the processing and producing of information from its interaction with the world. The way we do this is the quintessential question of the artificial intelligence field, at least for those interested in the actual working of human's cognitive system. What is relevant here is that this system attributes different status to its represented information, and that this is crucial for its working. Which action should be taken before a certain situation, for example, is a task that can only be solved by assessing the information available to the system. Depending on what is at stake, this information may not have the right epistemic status to the system and a decision cannot be reasonable taken. In contrast, the conclusions of several reasoning processes have a special status which authorizes the agent to act and think according to them. Our claim here is that k-states are simply one kind of status that may be attributed to the information available to the cognitive system. It is quite obvious that many pieces of information are simply internally treated as pieces of knowledge. Countless outcomes of perceptual processes seem to have the internal status of knowledge, some actions are carried forward only if the reasons that motivate them are taken as knowledge, and the outcomes of reasoning processes have such a status only if the status of their inputs, individually or collectively, are good enough according to some determinate standards. Our cognitive system must have a way to recognize this internal status.

The claim that there are different epistemic statuses and the use of expressions like "special status" and "good enough" may sound as if a piece of information must be very strongly supported to count as k-states. We are not trying to suggest that. Indeed, we believe, there are two properties of k-states that contradict this idea. First, it is plausible that many stored information are simply internally considered knowledge *by default*. That is the case, for example, for most acquired information. Most outcomes of perceptual processes, most of our conclusions, most of the information that we receive by testimony, etc., are internally perceived as knowledge. Second, k-states are *defeasible*. Although much of our represented information is just considered knowledge by default, this fact, by its turn, does not mean that such a status is not *defeasible*. Indeed, this is a platitude. One can, for instance, reflect upon his reasons and change his mind about a certain mind. Even simpler, one can discover evidence that seems to contradict what one learned before or learn new facts that simply disprove them. It is true that some information survives scrutiny and it still counts as k-state, which, in a sense, suggests that it is stronger than other kinds of information, but information certainly does not need to undergo scrutiny to receive the status of k-states.

Now, even that "that k-states be factive" is not a condition used in the ordinary categorization of knowledge, i.e., even if this condition is not part of the information stored in KNOWLEDGE, it is still plausible that it is one of the *constitutive properties* of "k-states". That is, the production of our epistemic states works in such a way that it necessarily considers k-states only those states which it represents as true. A piece of represented information cannot be considered a k-state without also be considered true because this is exactly one of its constitutive properties. Of course the representations of k-states can actually be false. The cognitive system can obviously misrepresent a piece of information as being true. The point, however, is that a k-state could not be *internally* represented as false. To internally consider information as k-state is simply to internally represent it as true. Also, the claim is not that the information must be explicitly represented as true to be considered a k-state. One does not need to both represent *p* and to represent that *p* is true. A proposition *p* is implicitly represented as true if the system's attitude toward *p* is such that it is processed as true. This is a functionalist point. Propositional attitudes can be, at least partly, distinguished by their functional properties to the system, and those properties in certain cases involve implicit representations of the proposition's truth value. To doubt that *p* is to implicitly represent *p* as false. On the other hand, to regret that *p*, if Williamson (2000) is right, is a factive state, so it implicitly represents *p* as true. Similarly, k-states are also factive states, i.e., they work in such a way that the represented information is processed as true.

Of course, nothing we said so far makes a case against the argument that (H1-*a*) empties our understanding of knowledge. These claims, however, help to clear the ground for doing this. One straightforward answer to this objection, therefore, is to bite the bullet. There is nothing wrong in suggesting that KNOWLEDGE is an information-poor concept, one may say. It is not obvious that people "understand", for instance, that knowledge is factive state. Is not this a point that always need to be made explicit in the epistemology classroom? It is certainly something on which people can easily agree, but that requires a little observation or reflection. In fact, there is no incompatibility between some reflective understanding about knowledge states and KNOWLEDGE being only a recognitional concept. According to (H1-*a*), however, such understanding is not constitutive of KNOWLEDGE and is not used in the ordinary categorization of knowledge states. We can make sense of this argument by saying that of our ordinary "understanding" of knowledge states, including factivity, is actually tacit. This tacit understanding, indeed, serves as evidence for the reflective understanding of factivity. Pointing to reflective understanding does not preclude the plausibility of (H1-*a*).

What about positive evidence for this hypothesis? The stronger evidence we can find for (H1-*a*) is in the experimental literature regarding *egocentric effects* in the reasoning about mental states. This literature concerns a number of particular tendencies we have to both overestimate our own mental states and to attribute other's with aspects of our own viewpoint. This is common theme in the developmental literature, in which children are often described as egocentric. Jean Piaget entitled a proper developmental stage of thought "egocentric thought", a stage where children seem to be unable to take the perspective of others. In a classic experiment, Piaget demonstrated how children can be egocentric with respect to, for example, spatial perspective (Piaget & Inhelder 1956). In the "three mountains task", a child is presented with a three dimensional model of three mountains, which have different sizes, colors, and features, like a cross, a house, snow. After being familiarized with the mountains, a doll is presented and placed in a position in which it "sees" the mountain for a different angle and the child is then asked about what the doll can see and to indicate it by pointing to one of a range of pictures of the mountains from different angles. What it was found is that 4 years-old always chose a picture of their own point of view, 6 years-old sometimes presented awareness of the doll's point of view, and only 7 and 8 years-old consistently chose the picture of the doll's point of view. His conclusion is that children are no longer egocentric at age 7. Piaget's conclusion and experiment is criticized by being inaccurate (Glucksberg, Krauss, & Higgins 1975; Shatz 1983), but there is little doubt that children present egocentric error and that their capacity to take another's perspective increases over time (Brandt 1978; Kurdek 1977). These egocentric effects are not restricted to spatial perspective, but extend to feelings and other sorts of viewpoints. Important for us, these effects include a specific class of errors or biases regarding the attribution of knowledge, constituting what is called *epistemic egocentrism*, and are not restricted to children.

Epistemic egocentrism is the tendency we have to both overestimate our own knowledge and attribute others in a more naïve situation with our own knowledge. The reason why epistemic egocentrism favors (H1-*a*) is that ST naturally predicts that egocentric effects may occur. Remember that according to a simulationist account like Goldman's, to successful predict someone's mental state, one needs to put their own mental states in "quarantine", to inhibit them from entering in the simulative process, otherwise they will interfere in the process and no longer resemble the target's states. If mental state attribution is a matter of having the right concepts or theory, on the other hand, then egocentrism should not be such an issue in the presence of such concepts and theories. Evidence of epistemic egocentrism, specifically, serves as evidence for simulation in knowledge attribution. The most famous example of epistemic egocentrism is the very failure of 3-4 years-old in the false belief task. For when children fail this task they reason as if the target knows what they know. But as is already clear, psychologists dispute whether this failure is a conceptual or processing matter, so it not immediately favors a simulationist interpretation. However, if epistemic egocentrism is also a common tendency in adults, who would have already acquired all the relevant concepts or theory for knowledge attribution, then we have a stronger reason to think that particular egocentric errors regarding mindreading of knowledge states consist in instances of simulation.

Another famous example of epistemic egocentrism is what is called the *hindsight bias* or "*knew-it-all-along effect*". This bias is characterized as the tendency to see past events as having been more predictable than actually was the case, making us to believe that we knew that it would occur, even when evidence indicates that we did not know. In an influent study, Baruch Fischhoff (1975) asked people to judge the likelihood of some historical events on the basis of written descriptions of them. Participants were presented with a short story with four possible outcomes, and asked to assess the likelihood of each individual outcome, where one group of participants was informed of the actual outcome of the story. Fischhoff found that participants in the informed group were much more inclined to attribute higher likelihood to the outcomes they were told to be true. This result is representative of the difficulty we have to suppress the knowledge available for us when assessing our previous situation, a difficulty that was found in several other studies (Hawkins & Hastie 1990; Baron & Hershey 1988). Indeed, the hindsight bias proved to be very resilient, continuing even after the subjects were fully informed about their bias (Pohl & Hell 1996), which suggests that it may have an automatic processing source[12]. It is implausible that the hindsight bias has a single underlying mechanism. It is probable, for example, that sometimes it is simply motivated for the agent's desire to appear more knowledgeable. Anyway, several cases can be accounted by models that explain the bias in terms of a *memory distortion* (Blank & Nestler 2007), and this fits naturally with a theory that assigns a fundamental role for introspection in mental state reasoning. When an agent judges his past mental state regarding a certain issue *x*, the actual internal availability of information regarding *x* may causes him a memory distortion so that he overestimates his past mental state, or, alternatively, whatever causes the memory distortion makes the new information to be introspectively seen as something that was known far longer than it actually was. One may claim, therefore, that the effect happens because the default strategy in mental state reasoning is introspective. Furthermore, the basic simulationist idea that we assess our own mental states to mindread others can also be used to explain egocentrism biases with respect to others' epistemic states.

In another experiment, Fischhoff asked people to judge the likelihood of certain outcomes of historical events, but also to estimate the prediction of others participants that did not know the actual outcome. As in the previous experiment, one group of subjects was told the actual outcome of the event. The result was that the informed group overestimated naïve subjects' prediction with respect the "right" outcome. That is, the informed group answered as if the more naïve subjects share some of their knowledge. Interesting, adults also may present difficulty to suppress their own knowledge in a task very similar to the classic false-belief tasks (Birch & Bloom 2007). In a more complex version of the 'Sally-Ann' task, subjects were asked to judge the probability of the protagonist of the story to first look in each of four boxes in search of her violin, which was moved in her absence. One group of subjects was told the exact location of the violin in the description of the story. Even knowing where Vicki originally left the violin, the more informed group

[12] An interventional strategy called "*considering the opposite strategy*", however, proved to be effective to eliminate the hindsight bias (Lord, Lepper & Preston 1984).

attributed a higher probability to the possibility of Vicki first looking in the new location of the object than the more naïve group. Results like this led psychologists Susan Birch and Paul Bloom to talk about us being "cursed" by our own knowledge in face of certain tasks. This bias is easily explained by a simulationist hypothesis like (H1-*a*). The problem in these cases is that the agent's k-states interfere in the simulative process. On the other hand, which pieces of theory could explain these errors? More generally, even if one develops an explanation in the terms of TT, it is simply implausible that every instance of the widely reported cases of epistemic egocentrism, in children and adults (Birch & Bloom 2003), is explained by theoretical mistakes.

The literature on epistemic egocentrism, therefore, provides the best empirical evidence in favor of (H1-*a*). We think, however, that one can make a stronger case for such a hypothesis by speculatively approaching the working of k-states and epistemic egocentrism. In particular, we can make (H1-*a*) especially suited to deal with an aspect of knowledge which has so far been overlooked in our investigation, *viz.*, the *normativity of knowledge*. In the philosophical literature the concept of knowledge is often described as a normative concept. Reasoning about knowledge, one may claim, has to do with reasoning about the *correct* attitude regarding a proposition. Accordingly, some epistemic intuitions are obviously normative, they evaluate that one should not have assumed that attitude, or that that attitude is appropriate, etc. Thereby, it may seem preposterous to most philosophers that a psychological analysis of KNOWLEDGE ignores this aspect of knowledge. This criticism is not warranted, however. It is not obvious that this aspect of the philosophical analysis is constitutive of the ordinary concept KNOWLEDGE or, more generally, that any condition defended in the more robust philosophical concept of knowledge is constitutive of the ordinary concept. In fact, we already rejected the composite assumption that comes from the philosophical literature in the previous chapter. If we relate normativity or normative judgments to the most reflective methods of thinking, as the ones present in the systematic philosophical methodology, for example, it is certainly doubtful that this kind of normativity is constitutive of KNOWLEDGE. Even that philosophers use them extensively to make theoretical epistemological evaluations, this kind of judgment may be simply distinct from the categorization judgments derived from the ordinary concept of knowledge and which are our focus here. Anyway, a question remains: what is the relation between normativity and KNOWLEDGE?

The answer from (H1-*a*) is that no normative judgments related to knowledge is constitutive of KNOWLEDGE. Nevertheless, its defender can argue, normativity judgments still can have important relations with k-states. First, in a sense, it may be determinant to certain processes that produce k-states. In many times, k-states are produced automatically, with no room for conscious processing, e.g., perceptual processes. In many other times, however, the production of a k-state involves some degree of reasoning. In those cases, the agent rationally assesses the matter in question through processes like factual reasoning, applying logic, insights and other intuitive processes, causal reasoning, etc., and the resulting outcome is a k-state. If we assume a naturalistic position regarding rationality, reasoning processes can be described in terms of *procedural norms* (Pollock 1995). That is, there are certain tacit rules or principles that authorize or prohibit the steps of those processes. Defined broadly enough, the procedural norms include not only the procedures that determine the more automatic decisions, but also the normative judgments that guide more conscious reasoning processes. Thus, if the normative steps involved in reasoning, automatic or conscious, are determined by the procedural norms, then they also determine the production of k-states. Furthermore, because k-states are defeasible, they can be the target of normative judgments themselves. One can rationally revise ones' attitudes and change his mind.

Second, in accordance with (H1-*a*), we can speculate that we use normative judgments to simulatively categorize others' epistemic states. When judging a certain subject matter, procedural norms give us personal normative judgments about how to proceed, what to think, what to conclude, etc., about the matter. It is plausible, therefore, that when one simulate other's mental state regarding a particular subject matter, those normative judgments can constitute part of the simulative process, especially if the matter or the situation trigger more conscious reasoning. In other words, to assess *S*'s epistemic state, regarding *p,* the agent needs to run his own cognitive mechanisms regarding *p*, and these mechanisms may include normative judgments. Now, we can go further and speculate that this opens the possibility of particular cases of egocentrism. For instance, if one fails to inhibit his own mental states when simulating *S*'s mental state regarding *p*, this affects one's normative judgments about how to proceed or about what one is authorized to conclude, resulting in an inaccurate simulation of *S*'s mental state. It is possible, therefore, that an agent fails to detect that *S* is in a k-state because failing to inhibit his own mental states when running his cognitive mechanisms causes him to judge that *S* is not authorized to being a k-state. This would happen, for example, in cases in which the agent is in a privileged evidential position with respect to *p* in comparison to S.

This idea, indeed, is already present in the philosophical literature. Jennifer Nagel has recently used the notion of epistemic egocentrism to explain certain patterns of epistemic intuition (Nagel 2010, 2012). Although she remains neutral about the dispute between ST and TT and the more specific details of the conceptual processes involved in simulation, we can still argue that this explanation is more naturally accounted by ST and a hypothesis like (H1-*a*) – we will look more closely to Nagel's proposal in the last chapter.

Importantly, we can now use this idea to answer (Q1), i.e., to explain the difficulty of finding a satisfactory definition of knowledge in the philosophical enterprise. As many imaginary cases of the literature describe situations where the epistemic agent is in a more naïve condition than us, as evaluators, e.g., cases where we are told about possibilities of error unknown by the agent, it is plausible that we egocentrically categorize their mental states from our point of view. Even worst, because it is always possible to artificially create cases where the agent is in a more naïve situation then the evaluator, it is always possible to create intuitive counterexamples to proposed definitions, which adds even more explanatory power to (H1-*a*). This way, using the distinction between the categorization of k-states and their working, the idea of epistemic egocentrism, and the place for normativity in the production of k-states, we made the best case we could for the radical simulationist hypothesis about KNOWLEDGE. Does this close the case for (H1-*a*)? No. The problem is that there are also positive evidence and arguments that can make a good case for the structured hypothesis we derived from TT too.

2.6.2. Systematic errors

Although egocentric effects are more naturally predicted by ST, TT also predicts a proper kind of error. Remember that a good explanation for people's expectation of the physical behavior of objects is the postulation of a naïve physical theory. The reason why that is a good explanation is because people's wrong expectations present a distinct pattern. The individual errors do not diverge radically from each other, but are instead coherently related. This is what would happen if people share the same underlying understanding of the relevant domain. For instance, a mistaken theory about the domain, or a mistaken piece of theory, would cause *systematic errors* in tasks which require a more suitable understanding of it. Systematic errors, therefore, are evidence of a theoretical structure underlying the reasoning about a domain. ST predicts egocentric errors in reasoning about the mental domain, but notice that those errors are not exactly systematic according to the simulationist approach. To make an egocentric error in simulation one needs to have certain mental states and fail to inhibit them in the process. If one does not have mental states that are significantly distinct from the target's mental states, them one will not be egocentric in simulation. Simulative egocentrism depends on the particular circumstances of simulation. If someone has the wrong theory or wrong piece of theory, in contrast, he will present these errors more consistently.

The idea of systematic errors is a constant part of TT's argumentation. Its interpretation of the data about false-belief task, for example, is in terms of 3-4 years-olds lacking a non-metarepresentational theory of mind, so their answers in mindreading tasks will reflect systematic mistakes that are consistent with the non-metarepresentational theory they possess. Relevant for us here, there are specific systematic errors regarding knowledge states recorded in the developmental literature. These errors are progressively overcome with development, as if children learn new things about the mental, suggesting that children's understanding of knowledge really reflects theoretical content. 4 years-olds, for example, seem to not distinguish between "ignorance" or "not knowing" and "being wrong". This is demonstrated by an experiment from Ted Ruffman (1996) in which children are placed in front of two dishes full of sweets and next to a puppet observer. The round dish contained red and green beads, and the square dish contained only yellow beads. One bead from the round dish is moved to a bag. Both the children and the puppet know this (they observed the bead be moved under cover), but only the children know that the bead moved was green. The children are then asked what color the puppet thinks the moved bead is. “He does not know” or “he thinks it is red or green” would be correct answers. Surprisingly, however, most children answered “red”. This is not a random answer. While a few children answered “green”, resembling the failure in false-belief tasks, no children answered “yellow”. Ruffman argues that this pattern is best explained by children possessing an inaccurate generalization or rule about knowledge, *viz.*, “ignorance = you get it wrong”. Only latter children learn that one can simply be in an ignorant or “guessing” state. Furthermore, this result cannot be easily explained in simulationist terms. First, the children know the actual color of the bead themselves. So if their mental states were getting in the way, the result would be that the majority of answers would say “green”. Second, we cannot say that children are themselves incompetent in inference, so they fail to properly simulate the adult’s inference process. As it was cleared demonstrated in control tasks, children were able to make the inference themselves.

This experiment from Ruffman suggests not only a conceptual generalization regarding ignorance, but also that children do not conceptualize inference as a source of knowledge. The “inference neglect”, as it is called, is clearly illustrated by another experiment in which children and an observer were presented to a transparent container with balls of the same color (Sodian & Wimmer 1987). The observer is told that one of the balls will be placed in a bag and leaves the room for a second. The ball is placed in the bag before the observer returning, and the children is asked a knowledge question: “Does the observer knows the color of the ball in the bag?”. Despite the fact that they have seen the observer seeing the transparent container and that they were perfectly able to respond correctly when put in the observer’s situation (and therefore successfully inferring the color of the ball), most of the children under 6 years old denied knowledge to the observer, as if they could not understand that one can make inferences and that they lead to knowledge. The fact that children increasingly become apt to attribute inferential knowledge and to give an inferential explanation for others’ knowledge can be seen as evidence of theoretical adjustment.

More generally, the inference neglect is an instance of systematic errors with respect children's conceptualization of the *sources of knowledge*. For instance, 3-5 years-olds do not seem to understand adequately that knowledge depends on the mode of the agent's informational access. For example, one experiment put children in the position to assess puppets' knowledge states from different perceptual access they had to certain objects. In particular, they were presented to a tunnel whose ends were covered by felt flaps and to pairs of objects which either looked the same but felt different, e.g., two identical sponges, one wet and the other dry, or felt the same but look different, e.g., two toy footballs, one green and the other red (O'Neil, Astington & Flavell 1992). The task involved placing one object of the pair at a time inside the tunnel and varying the perceptual access one has to that object, e.g., the puppet could only look or just touch the object. When asked to judge which puppet know what object is inside the tunnel, 3-5 years-olds had a poor performance in predicting that puppets that only saw the object would not identify one of the objects that is distinguished by its tactile properties and seemed to understand that the mere fact that seeing the sponge, for example, was enough to the puppet gain knowledge of the sponge being wet or dry. In contrast, 4-5 years-olds properly attributed knowledge to the puppets which saw the objects visually identifiable.

Results like these made many psychologists like Ruffman to conclude that children's understanding of knowledge is constituted by some generalizations or rules which are improved with time. One rule that could explain most of 3-5 years-olds knowledge attribution are rules like "seeing = knowing" and "not knowing = getting it wrong" (Ruffman 1996). With time, children would correct their generalizations and acquire more rules, making them more competent to understand the informational sources that lead to knowledge. This interpretation would explain a number of other attribution patterns, like why children go through a period in which they tend to overestimate the information that can be gained from verbal messages, e.g., they think that an observer know where a piece of chocolate is by just being told "the chocolate is in the red drawer", even if there is two red drawers in the referred cupboard and the message is ambiguous (Sodian 1988).

These systematic errors present cogent evidence that we store theoretical information about knowledge and use it to make categorizations about instances of knowledge, and, therefore, favors (H2-*c*). To store a rule or generalization about the knowledge sources is to store (at least) causal information about knowledge. If TT's interpretation of the data is right, then KNOWLEDGE has, in a sense, a theory-like structure. However, one may not be convinced of (H2-*c*) by this developmental evidence only. For instance, if we use generalizations that we improved with time to think about and categorize knowledge states, what is the answer for (Q1)? That is, why is so hard to find a definition of knowledge? If KNOWLEDGE stores causal or nomological information about its instances, why we were not yet able to translate this content to the form of a definition?

As in the case of the simulationist hypothesis, there is some room for speculation and to make a better case for the hypothesis from TT. One can defend (H2-*c*) by arguing that there is no reason why we should think that conceptual generalizations are consistent enough to support a definition. One problem with TT's vague use of the notion of "theory" is that it is common to think of a theory, at least a real or robust theory, as providing *criterial conditions* about the domain of which it is about (Leslie 2000), but there is no reason to think that what determines that a rule or generalization is stored in a concept also determines necessary and sufficient conditions. For instance, some psychologists emphasize that the existence of naïve theories is explained by their utility as a heuristic mechanism (Saxe 2005). The naïve physics possessed by most laypeople, despite being simply false from the point of view of physics, is very useful to predict a number of instances of physical behavior and often generates accurate particular judgments. Similarly, a concept may contain certain generalizations just because of the usefulness criterion, just because they useful as a heuristic mechanism, providing accurate predictions or judgments to a number of situations. Furthermore, the problem is not that these generalizations can be false, but that a concept may contain generalizations that are not exactly consistent in relation to each other.

For a tentative example, let us take the "chick-sexer" case, often used in the epistemological literature. The case starts from the fact that there are individuals, working particularly in the poultry industry, that possess the ability of reliably sort male from female chicks. Their ability, indeed, needs to be highly reliable as their decisions have a direct impact on the industry profits. Most of them, however, according to the description of the case, cannot explain how they distinguish the chicks' sex. Indeed, the description goes, many of them give a false an explanation like "I do it on the basis of sight" when they actually do it on the basis of smell. The putative intuition about this case is that although the subjects lack internal justification, they know that a certain chick is male or female when they correctly sort them. We think that this intuition can easily be explained by the postulation of a piece of theory like, roughly, "consistent successful outcomes are caused by knowledge". That is, the intuitive categorization is determined by a generalization which says that the successful outcomes of the subjects are explained by them possessing knowledge. Given that many cases of successful outcomes are actually explained by one's having relevant knowledge, it is plausible that we use such a principle to categorize one's epistemic state. This would explain the intuitive categorization of cases like the chick-sexer case which is used in favor of the idea that internal justification is not necessary for knowledge. Nevertheless, it is also plausible that KNOWLEDGE stores other generalizations which generate intuitions that contradict this idea.

Another alleged intuition in the epistemological literature comes from cases like the "Mr. Truetemp" (Lehrer 1980) and the "clairvoyant" cases (Bonjour 1980). Both cases describe a subject who subtly acquires a highly reliable belief producing mechanism whose acquisition they are not aware of (a brain implanted mechanism that produce correct beliefs about the ambient temperature and a clairvoyance power). The mechanisms start to generate spontaneous true beliefs and we are asked about the status of these beliefs. Our intuition of these cases, however, at least epistemologists' intuition, is that the outcome of these mechanisms does not constitute knowledge states. One plausible explanation is that our categorization here is guided by another heuristic principle which relates knowledge with subjective processes. Because several instances of knowledge involve subjective access to evidential basis one has for his attitude, a theoretical piece that links this subjective access to knowledge would provide many accurate predictions. Cases like the ones exemplified by Mr. Truetemp and the clairvoyance or something in their descriptions triggers the use of this principle in categorization and, contradicting the previous generalization, causes one to deny knowledge to the subjects. Furthermore, remember that when we discussed the prototypical and exemplar hypothesis in the first chapter we argued that the class of knowledge states seems to be much diversified. This constitutes another reason to think that there are different theoretical principles stored in KNOWLEDGE. Thus, answering (Q1), assuming that KNOWLEDGE contains theoretical structures, they are not as rich as a real theory could be. In particular, theory-like structures seem to not contain criterial information – otherwise it would be easier to provide an intuitively satisfactory definition of knowledge. This being the case, distinct rules or generalizations can be each sufficient for particular intuitive attributions and, at the same time, not obviously favor a single analysis collectively. It is plausible that interpreting each principle behind intuitive attributions as providing a necessary criterion results in an (intuitively) inconsistent analysis, for each principle is sufficient for an attribution that contradicts the necessity of the other.

2.7. A POSITIVE PROPOSAL

We are finally in position to provide an answer to our central question here, at least a tentative one. We saw positive evidence and arguments for both (H1-*a*) and (H2-*c*). It is now obvious, however, that the empirical evidence for the latter contradicts the former. Evidence of systematic errors is evidence that we have some informational structure stored in KNOWLEDGE. And that is cogent evidence. However, is it sufficient for ruling out the idea of simulative processes for KNOWLEDGE? We do not think so. The evidence of epistemic egocentrism is at least as strong as the evidence for systematic errors. In addition, the explanatory power of the two hypotheses is really strong. They both allows us to explain why we have difficulty in defining knowledge from our intuitive ascriptions and provide good accounts for the role of normative judgments, tacit understanding, causal reasoning, etc. How should we accommodate all this evidence? Should we favor the simulative or theoretical account for KNOWLEDGE?

Our answer is now quite predictable: we should adopt a hybrid view. Like Goldman (2006) defends, it is reasonable to believe that both simulative processes and theoretical reasoning are involved in our reasoning about mental states. In particular, KNOWLEDGE involves both simulative processes and theoretical structures. Curiously, Goldman follows most philosophers and do not include

KNOWLEDGE in his list of mental state concepts. He says that the "verb 'know' is not a pure mental-state verb. The standard epistemological story says that knowing entails the truth of p; a variant of the standard story adds that the belief must be acquired by a reliable method" (p. 81). We think, however, that we made a strong case for the thesis that KNOWLEDGE – in contrast to philosophical analysis of knowledge – is a mental state concept and that nothing in the epistemological literature falsifies it. We claim now that we should adopt a hybrid theory of mindreading with respect the particular case of KNOWLEDGE. Thus, our answer is definitely a structural one: KNOWLEDGE contains, at least, theoretical structure. Yet, simulative processes are also responsible for several of our ordinary categorizations, and this, in its turn, involves introspective and recognitional processes. In other words, when mindreading others' epistemic states, sometimes we use causal or nomological information, and sometimes we just simulate their states and introspectively identify what state they are in.

Obviously, this is not the full story. A full story would explain how introspective, simulative, theoretical processes, and possibly modules, interact in conceptual development and form the mature concept of adults. This is a very complex question and we will not try to solve it here. Anyway, we cannot help but speculate that introspective process play a crucial role in conceptual development. We think it is plausible to believe that children acquire information about knowledge states in many occasions through the use of a recognitional concept. Being able to recognize k-states, they start acquiring information about this kind of states from particular instances and form generalizations about them. Anyway, although we do not have a detailed account of the conceptual development and acquisition of KNOWLEDGE, we think that the empirical evidence we evaluated and the explanatory power of the simulationist and the theoretical hypotheses authorize us to conclude in favor of a hybrid hypothesis.

Finally, there are important implications for epistemology in this conclusion. If our intuitive ascriptions of knowledge are determined both by simulative and theoretical processes, then, not only the traditional project of the analysis of knowledge is in deep problem – for now he have positive reasons to think that we cannot find a definition intuitively satisfactory – but, in principle, these intuitions are subject to *two* kinds of problems, *viz.*, egocentric and systematic errors. It is a job now to epistemologists to rethink our expectations regarding the traditional project and to discuss the exact implications of these findings about our epistemic intuitions on the way we theorize about knowledge. In the last chapter we will gave a taste of the implications of these results for epistemological theorization and provide a rough account on how to make metaepistemological sense of them.

PART II – The cognitive turn in epistemology

Chapter 3

The cognitive turn and the issue of intuitive agreement

3.1. THE COGNITIVE TURN

Among the various recent investigative paths of epistemology, one is characterized by the great attention it gives to methodological and empirical issues. This strand originates from some naturalist worries about the armchair theorizing of epistemologists and constitutes a *cognitive turn* on their methods and subject matters (Brown & Gerken 2012). That is our focus here.

The main concern of this strand refers to the widespread appeal to *intuitions* about epistemic imaginary cases. Every philosopher is familiar with it. One carefully describes a scenario containing properties relevant to his discussion, formulates a question about whether the situation or a component of the situation has a certain property, e.g., "is Jean's belief justified?", and then makes explicit his spontaneous judgment about the relevant matter. Significant results are achieved only when the intuition is shared among most philosophers. This is neither an exclusive practice of epistemology nor a recent one. It is very common in any discipline engaged in the general enterprise of *conceptual analysis*, and in epistemology it is as old as Plato's jury case in *Theaetetus*. However, since the well-known article of Edmund Gettier (1963) and the great revival of interest in the *analysis of knowledge* that followed from it, the use of intuitions from imaginary cases became more prominent in epistemology than any other philosophical area. This consequence and its subsequent theoretical contentions explain much of the emergence of the naturalist concerns leading to this recent cognitive turn. In particular, some philosophers recently argued that the traditional project of conceptual analysis – of which the analysis of knowledge is only one version – makes some substantive empirical assumptions, and that the empirical evidence does not really support it (Ramsey 1992; Laurence & Margolis 1999).

If we were to characterize the cognitive turn as limited only to problems with the analysis of knowledge, however, that not only would be misleading but also would take away much of its relevance. The use of intuitions by epistemologists is not limited to conceptual analysis so defined, remaining important roles for them. For instance, intuitive judgments are being used more recently for supporting or undermining general theories of knowledge which do not depend essentially on definitions of epistemic concepts. This is the case, for example, of *epistemic contextualism* (DeRose 1992, 2005, 2009) and *subject-sensitive invariantism* (Hawthorne 2004; Stanley 2005) which have concentrated much of the recent debate in the literature. Moreover, intuitions also play an essential role in the very status of paradox that some central arguments of the literature have, e.g., the *skeptical paradox* (Cohen 1988) and the *lottery paradox* (Kyburg 1961). The problematic aspect of a conclusion or a premise which constitute an epistemic paradox is a conflict between a logical and an intuitive level. It is because our intuitive judgments disagree with the logical conclusion of an argument, or the subsequent review of a premise, that we have a paradox. The roles that intuitions play in these cases are not affected by problems related with the attempt of achieving definitions, but these too are not free of naturalist cognitivist worries.

3.1.1. Questions about intuitive agreement

What the intuitions serving as evidence for these theories and the intuitions generating the mentioned paradoxes have in common is that they are judgments ascribing or denying knowledge to someone. That is, they are intuitive judgments which outcomes are *knowledge ascriptions*, and they play a central role in the theorization about the nature of knowledge.[13] One problem, however, is that epistemologists seem to be making significant empirical assumptions when using them. Stephen Stich and Jonathan Weinberg, for example, called attention to this by stating that when a philosopher points out to a certain intuition he assumes he is correctly predicting the intuition of others, i.e., that there is agreement around it and that it reflects the intuition of the folks. Their point is that interpersonal concordance is ultimately an empirical question (Stich & Weinberg 2001) and, along with

[13] We are not assuming that intuitive judgments are the only constraints in the theorization about knowledge. There are other constraints such as the pretension of many epistemologists that their theories are able to accommodate the logical constraint of epistemic closure, for example, and possible theoretical influences of all kinds. These constraints, however, do not affect the central role that knowledge ascriptions typically play in the development of the types of theories of knowledge which we are interested here.

Shaun Nichols, they presented empirical results which supposedly show how cultural background can determine individual intuitions about important epistemic cases, such as Keith Lehrer's Truetemp case (Lehrer 1990), and Gettier cases (Weinberg et al. 2001). Most naturalists are compelled to agree that the question of intuitive agreement is an empirical one, and with the threat of wide intuitive disagreement some general questions arises.

By trying to show how individual intuitions can disagree, Stich and others have a serious agenda in mind, *viz.*, to advocate that intuitions do not have evidential validity, so philosophers should radically rethink their methods. To conclude that from Weinberg et al. (2001) results would be too rushed, however. It is very debatable how conclusive or relevant these results are. Supposing they are accurate, the implications on the philosophical agenda pass through higher order issues. In particular, does the evidential status of epistemic intuition really depend on this concordance? Some doubt that when accepting that folks' intuitions might be inconsistent but claiming that what really matters for philosophical theorizing are the robust intuitions of philosophers, i.e., experts' intuitions (Kauppinem 2007; Jackman 2009). The strength of this position is questionable, however, due at least two reasons. It seems plausible to say that philosophers are really more competent in examining imaginary cases in the sense that they are generally more willing to engage in such an abstract task, or that they better respond to the unusual features typically present in the description of these cases. On the other hand, however, it seems equally plausible that philosophers may be influenced by their theoretical inclinations and do not form a homogenous class of intuitions. A good defense of the "philosophers' intuitions first" position has to explain how expert's judgment can be free of individual theoretical inclinations and at the same time distinct from ordinary competences. But, as we will argue below, even if we accept this position, it is also subject to important naturalist worries, and we can say that, in general, philosophers under the cognitive turn are not favorable to the idea of philosophical expertise on intuitions.

Besides, the necessity of agreement makes a lot of sense for those who see their projects as explicitly relying on ordinary intuitive ascriptions. For example, some philosophers see their theories as relying in the ordinary concept of knowledge or justification (Goldman 1986; Goldman & Pust 1998). Contextualists and subject sensitive invariantism, in turn, build their theories from certain intuitive patterns which are supposed to be ordinary. From this perspective, if the philosophers' intuitions do not respond for our ordinary concepts, why would they matter? For anyone with this common view the question of intuitive agreement becomes very relevant. So, more fundamentally, are the claims of Stich and his colleagues correct? Do we have substantive intuitive agreement? Should we worry about, for instance, the possibility that the folks, unlike most epistemologists, assign knowledge to subjects in Gettier cases, or have different intuitions about the bank cases supporting epistemic contextualism (DeRose 1992), etc.? As we will see, many philosophers have taken these issues seriously and have recently conducted their own experiments to answer them. However, in the following sessions we will focus on only two types of intuitions: the Gettier intuition, and the intuition from skeptical pressure cases.

3.1.2. Questions about the cognitive bases of intuitions

Furthermore, there are more specific questions about intuitive epistemic judgments relevant to anyone who uses them, even those who do not make explicit their meta-epistemological view, or who claim that their projects do not strongly rely on ordinary epistemic concepts. Consider, for example, how knowledge ascriptions are used by advocates of subject-sensitive invariantism. Those philosophers typically describe pair of cases that supposedly share all the traditional factors that are said to constitute knowledge, e.g., truth-values, justification, reliability of belief-formation processes, etc., but which differ with regard to the *stakes* involved to the subject. The alleged intuitive pattern is that the higher the stakes, the less we are willing to ascribe knowledge to the agent. This pattern is used to support the idea that the *practical interests* of the subject are also one of the constituents of knowledge. But, supposing they are right about the relevant intuitions on this type of cases, what explain them? What is happening at a mental level which produces this outcome? What the role of practical interests in the processes of knowledge ascriptions? Philosophers as John Hawthorne, Timothy Williamson and Jennifer Nagel (Hawthorne 2004, 2004a; Williamson 2005; Nagel 2008; 2010a) have noted that depending on the answer for these questions, we can have a traditional interpretation for this intuitive pattern and the subject-sensitive invariantism will have to find support elsewhere. For example, if an intuitive pattern proves to be a kind of systematic error performance, we have reasons to disregard its alleged evidential value for subject-sensitive invariantism.

Putting in more general terms, the cognitive turn is also interested in what explain the particular intuitive patterns found in the literature. Investigating what happens at a cognitive level help us to make a much better use of our knowledge ascriptions. We can find better reasons to accept an intuition, and we can know better what exactly it is evidence of. Some of the questions we can raise on this issue are: what is the type, or types, of processes related to knowledge ascriptions? That is, what are the kinds of psychological processes involved in the categorization of knowledge situations? There are different kinds of processes for specific kinds of intuitions? What, for example, explains Gettier intuitions, cases involving stakes, or skeptical pressure cases?[14] Note that even if we accept the idea of the priority of expert's intuition these are relevant questions to everyone using knowledge ascriptions as a source of evidence for their theories.

In sum, there are different ways of how cognitive research can affect our epistemological projects. There are some general empirical questions whose answers were simply presupposed until very recently, such as the question about the intuitive agreement over important epistemic cases. And there are more specific questions that directly affect the interpretation of particular knowledge ascriptions. How much these issues are decisive depends on the meta-epistemological view one has[15], but the more specific questions are relevant to anyone relying on intuitive judgments that derive from our ordinary patterns of knowledge ascriptions, even for those who are unaware of it. Of course, the two types of questions are not completely independent of one another. For instance, the issue of agreement passes through the question of what are the kinds of processes responsible for our knowledge ascriptions and, on the other hand, some of the specific questions only make sense if the general ones are answered. For example, there is no sense to ask what explains the Gettier intuition if in reality there is no intuitive

[14] Note that our focus here is on knowledge, but that does not mean that judgments about justification are not involved in knowledge ascriptions. How the concept of justification is related to the concept of knowledge is something to be investigated.

[15] We are not assuming here that these meta-epistemological views are not concurrent. The point is that is that they are out of our scope here.

agreement about it. So we will start with these general questions. In this chapter we will review the best evidence we have and draw some conclusions about the intuitive agreement issue and what are the intuitive phenomena to be explained by more fundamental cognitive questions. For the purposes of this chapter, we will discuss only the Gettier intuition and the intuition from skeptical pressure cases. In the next chapter we will look at the recent proposal of Jennifer Nagel to explain the cognitive bases of these intuitions.

3.2. EXPERIMENTS AND INTUITIVE PHENOMENA

We mentioned above that many philosophers currently take the empirical testing about our epistemic intuitions seriously. That is not only due the fact they are appealing for naturalists, but also due to a series of negative results with respect what the traditional literature says. Some studies, for example, have attempted to show that we not share the same conception of knowledge and that it varies across groups defined by factors such as ethnicity (Weinberg et al. 2001) and gender (Buckwalter & Stich 2010). This claim is supported by results showing evidence for a relation between these factors and different responses to well-known cases of the literature, e.g., Gettier cases and the Truetemp case. If different groups think differently about the same significant epistemic cases, then this may mean that they have distinct conceptions about knowledge. It is not easy, however, to show that there is genuine disagreement between these groups.

Since the appearing of those surprising results there has been much discussion about their significance, and now that the issue became properly empirical one fundamental matter is whether they are experimentally valid. i.e., whether their methodologies are valid. There has been, for example, much skepticism about the results of Weinberg et al. (2001). Ernest Sosa (2008), for instance, argued that there is no guarantee that the subjects of these experiments understood the cases properly and, as we will see, this claim can be supported by the lack of control parameters in their experiment. In fact, it is safe to say, the extent that experiments about epistemic intuitions are appearing, criticism and discussion over their results have been responsible for improving their methodology and now, as a consequence, we have experiments whose results contradict earlier experiments. Much remains to be learned and there is room for much more experimentation, but we are already in position to draw up a temporary balance. We can ask now: what the best evidence we have says about our epistemic intuitions? Does the evidence strongly contradicts or supports some of the epistemologists' predictions? Which ones?

Before we start, some remarks are important. First, what are we looking for? What would we find in a suitable experiment that would suggest that certain intuitive judgment is robust? One point is that since these experiments follow paradigms from social sciences, their statics methods predict some variability in the data. There are several reasons for this variability, from performance errors to legitimate differences in the mechanisms responsible for epistemic judgments. But as these experiments serve to illuminate these mechanisms only obliquely we cannot say what the divergent responses mean. Indeed, the result only tells us something about the patterns of our epistemic judgments. The patterns are informative, but as we will see, what exactly they mean for epistemological purposes depends on further investigation, one that justifies the cognitive turn.

What we are looking for now are strong tendencies. A suitable experiment shows evidence that a certain intuition is robust if its statistical inference shows a strong tendency in favor of that intuition. Second, we will focus here on only two specific intuitions which are central to epistemological discussions.

3.2.1. The Gettier intuition

Responsible for much of the popularity of imaginary cases in epistemology today, Gettier cases would not have achieved that if there was no broad consensus about them. Epistemologists strongly agree that the cases present in Gettier's paper are instances of justified true beliefs, but that they do not constitute knowledge. It is not surprising then, due to its importance, that it has been chosen to be empirically tested. It is rather surprising, however, that empirical results contradict such a strong conviction of epistemologists. If they are wrong about that intuition, in the sense that it does not reflect the ordinary judgment, so what others intuitions are they wrong about? The evidence at hand nowadays, nevertheless, does not authorize much concern.

3.2.1.1. The need for good experiments

Among the cases tested by Weinberg et al. (2001), they tested the following Gettier case:

> Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

At the end of the description subjects had to respond to the forced choice question that follows by either answering "really knows" or "only believes". They found that while most Western answered according the philosophical perspective (over 70%), most participants who referred themselves as East Asians answered that Bob "really knows" that Jill drives a American car (over 50%), and even more who referred themselves as from "Indian sub-continent" answered that Bob knows (over 60%). That is a surprising result and if accurate gives us not only suggestive evidence that Gettier intuition is not universal, but that it may be singular to Westerns. This experiment can be criticized for a number of reasons, however. One problem that justifies the skepticism expressed by Sosa (2008) is that this experiment lacks control over whether participants understand the story properly. Are they misinterpreting some aspect of the story? Particularly, are they misinterpreting something relevant for them realizing how luck enters in the story? This is a simple matter, but we should remember that we are trying to find empirical evidence that a certain intuition is robust, and if we test the wrong variable or if we miss a hidden variable we have an invalid result. Because there is the possibility of misinterpretation of the case – especially in the case of non-Westerns – this experiment lacks control over this variable, and this makes it unreliable.

Close to that, another weakness of the experiment is that only one Gettier case was used. Gettier cases are not limited to the cases used in the famous paper, rather, they form an entire kind of cases. Maybe cultural background really is affecting the responses of participants, but it may be only due an aspect of this particular case which makes them construe the case differently than Westerns. For instance, non-Westerns may have different presuppositions about Americans and their cars, e.g., "Americans typically prefer American cars". To see if an intuition is robust – whether it supports or contradicts the philosophical perspective – we need to test different versions of the case to reduce the chance of hidden variables in the specific descriptions (Nagel 2012a; Nagel et al., 2013).

A third worry is about engagement. How meaningful the responses for the participants themselves are? The results reflect the answers to the forced choice question, but they do not show how sure participants are about their answers. If the conviction average for one of the answers is low, is it still significant? In fact, the statistical test chosen by Weinberg and colleges is adequate for small samples, and the difference showed between groups about the Gettier case is statistically significant. However, if we look to the numbers of the non-Westerns groups themselves we may not be convinced about them. 13 East Asians answered that Bob "knows" opposed to 10 answering that Bob "only believes", only 3 more, and 14 subcontinental Indians answered that Bob "knows" opposed to 9 answering that Bob "only believes", only 5 more. Given the chosen statistical test, these are significant differences, but having in mind the worries we pointed out above, the differences are not cogent at all. Given the lack of control of this experiment over important variables, these numbers are indeed consistent with participants responding randomly.

A more convincing experiment then this one, therefore, would have to present different versions of the Gettier case, control over the understanding of the description, and some measure of the participants' confidence. Besides that, large samples are always desirable. In what follows, we will concentrate on what consider the experiments with the most important results to date.

3.2.1.2. A more worrying negative result

A more impressive series of experiments was recently provided by Christina Starmans and Ori Friedman (2012). In view of experiments like that of Weinberg et al. (2001), Starmans & Friedman emphasize the importance of studies paying close attention to control conditions for us having a better assessment of whether participants are responding to the features that turn a situation in a Gettier scenario and what these features are. They tested five different stories and some variations of these stories, most of them consisting in Gettier scenarios, and found evidence that people consistently attribute knowledge to subjects in Gettier scenarios, except in the cases where they involve *apparent evidence*. Importantly, all the stories are followed by comprehension and confidence questions, which result in the elimination of the answers of those who failed the former, and in an informative measure of the participants' conviction. Their results, therefore, account for the three worries we found in the experiment of Weinberg et al. (2001).

Starmans & Friedman begin by noting that there are two properties commonly attributed to Gettier scenarios, in addition to be justified beliefs. First, there are two kinds of luck involved in the scenario, an instance of bad luck, which interferes in the truth of the belief, and an instance of good luck, which cancels the effect of bad luck over the truth of the belief (Zagzebski 1994; Turri 2011). The other property is that there is a disconnection between what makes the belief justified and what makes it true (Goldman 1967). In one of their experiments, then, they used different versions of the following story, which were randomly assigned to participants:

> Katie is in her locked apartment writing a letter. She puts the letter and her blue Bic pen down on her coffee table. Then she goes into the bathroom to take a shower. As Katie's shower begins, two burglars silently break into the apartment. One burglar takes Katie's blue Bic pen from the table. But the other burglar absentmindedly leaves his own identical blue Bic pen on the coffee table. Then the burglars leave. Katie is still in the shower, and did not hear anything.

There are two versions of this story consisting in Gettier conditions. The version above is called the 2-thief Gettier condition. In the other version, called 1-thief Gettier condition, just one burglar breaks into the apartment and stole the pen, but replaces it with an identical pen (so it has an instance of bad luck and an instance of good luck). In both conditions, participants are asked whether Katie "knows" or "only thinks" that "there is blue pen on the coffee table". The difference between them is that in the 1-thief Gettier condition there is some causal connection between what makes the belief justified and makes it true (there is a pen on the coffee table *because* the burglar stole Katie's pen) and a causal connection between the bad luck and the good luck (they have the same source), so these connections could be variables determining participant's answers and should be tested[16]. There were also a false belief (where one of the burglars left his bandana in place of the pen) and a control condition (where participants are asked about whether Katie "knows" or "only thinks" that "the letter is in the coffee table"). In all conditions participants were asked to answer comprehension questions (e.g., is there a pen on the table? [Yes/No]; How did the pen get on the table? [Katie put it there/The burglar put it there]), and a conviction question, where they should score on a scale of 1-10 (1 meaning not at all confident, 10 completely confident). The score was multiplied by +1 when participants attributed knowledge and -1 in the case participants did not attributed knowledge.

Not surprisingly, most participants attributed knowledge to Katie in the control condition (79%), and the minority of them attributed knowledge to Katie in the false belief condition (14%), confidence rating exceeds chance in both cases, (M= 5.41; M= -6.69; respectively). However, despite the differences of causal connections, most answered that Katie knows (72%; M= 3.77) in the 1-thief condition and in the 2-thief condition (79%; M= 3.92), and the difference between the two Gettier conditions was not significant. So, if this is a robust result, there is evidence that people do attribute knowledge to gettiered cases of justified true belief. Furthermore, given the exemplary design of the experiment, we have no reasons to think it is not robust.

Observing that in many Gettier cases of the epistemological literature, unlike the story of Katie, the subjects form beliefs based in a piece of evidence that only *appears* to be informative, e.g., the interviewer which says the false testimony that the other candidate is the one that will be hired (Gettier 1963), Starmans & Friedman also did an experiment where participants were randomly assigned to one kind of version of two stories. The two versions differed in respect of whether

[16] This is particularly important because in the first experiment of the paper most participants attributed knowledge to the subject in the Gettier condition, but those causal connections were present.

the epistemic subject form his belief based on "apparent evidence", as they called it, or authentic evidence. For example, in one of the stories, Corey has been collecting coins in his piggy bank for years. One day he perceives that a coin he is about to put in his piggy bank appears very old, he reads in the coin that it dates from 1936, deposits it and goes to take a nap. He is not aware, however, that there is already a 1936 coin buried deep in the piggy bank. Corey's roommate comes home and needs some change, so he shakes the piggy bank and ends up getting the coin Corey just deposited. In the authentic evidence version of this story, Corey is right about the date of the coin, but in the apparent evidence version, he misread the date, which actually says 1938. If participants are sensitive to the authenticity of evidence, they should respond differently to the two kinds of versions of the stories.

Starmans & Friedman found that most participants ascribed knowledge to the subject in the authentic evidence conditions, with confidence rating exceeding chance (67%; M= 4.90) and that the minority of them attributed knowledge in the apparent evidence conditions, with confidence rating exceeding chance (30%; M= -6.88). This is evidence that people do deny knowledge to agents which are in one kind of Gettier scenario, *viz.,* cases where they form beliefs by apparent evidence. But what should we say about the other results? Why people seem to attribute knowledge to subjects that are in Gettier scenarios involving authentic evidence? There is luck anyway influencing the truth of their propositions. One of the questions that Starmans & Friedman raise throughout their discussion is whether folks' conception of knowledge is equivalent to the traditional justified true belief analysis. This possibility is obviously rejected, however, since cases of apparent evidence are also cases of justified true beliefs. A much more interesting question is what these experiments say about how a no-luck criterion is part of the folks' conception. Do they have no such criterion and the pattern found in the apparent evidence conditions means something else? Or is it a matter of performance? Is it the case that philosophers are just best detectors of luck than the laypeople? Starmans & Friedman left these questions open.

3.2.1.3. The hypothesis of sensibility to luck

The results from Starmans & Friedman contradict the philosophical expectation; we would expect laypeople to deny knowledge in authentic evidence Gettier conditions. But how bad the news are? On the bright side, the results from apparent evidence suggest that folks' conception of knowledge is not equivalent to the traditional analysis of justified true belief and do not exactly dismiss a non-luck criterion, so we are not completely wrong about the intuition of laypeople about Gettier scenarios. Furthermore, we do not know what the pattern from authentic evidence conditions means. Starmans & Friedman themselves raise the hypothesis that the difference lies in the sensibility to luck. This is an important possibility because, if true, then philosophers and the folks are not intuitively disagreeing about knowledge, but about luck or how much luck is involved.

Indeed, this hypothesis makes sense when we look at views which say that luck is a function of *chanciness* and *significance* (Pritchard 2005). Roughly, therefore, to an event count as lucky, it needs to be seen with low chance to happen and as significant. Assuming this is true, participants of authentic evidence may not be seeing the event of an ordinary blue Bic pen being stolen as significant, and that may be the reason they ascribe knowledge to the epistemic agent, despite the chanciness of the event. Similarly, the chanciness of the truth belief may be more salient in the apparent evidence conditions, and this is enough to get people to deny knowledge to the agent. Therefore, if we find evidence that in Gettier cases where the amount of luck involved in the possession of true belief is more obvious laypeople do not attribute knowledge to the agents, whether it is a case of apparent or authentic evidence, then we can sustain the thesis that philosophers' and laypeople's intuitive judgments about Gettier cases do not differ significantly.

3.2.1.4. Turri's tripartite structure

An interesting idea was recently presented by John Turri (2013). Based in his own experience, Turri proposes to dramatize the distinct parts of Gettier cases – the justified belief of the agent, the instance of bad luck, and the instance of good luck –

by presenting the case in a tripartite structure, i.e., in three separated parts, to see if that effectively affects laypeople's judgments.

His first experiment tests the case of Robert, who recently acquired a rare 1804 US silver dollar, which he keeps over the fireplace of his library. Participants were randomly assigned to either a control or a Gettier version of this story, but they all read the same first part of it – where they are informed that just before Robert going to receive visitors, he closes the door of the library – and the same third part – where they are informed that there is another 1804 US silver dollar in his library lost in the mortar mix. The second part of the story is what differentiates the two versions. In the control condition, participants are informed that the vibrations of the door shutting caused the coin to fall in the rug next to the fireplace. In the Gettier condition, they are informed that a thief quickly entered in the library, stole the coin, and escaped, just before he receives his visitors. Note that this Gettier condition is a case of authentic evidence – the reason Robert beliefs there is a coin over the fireplace is the fact the he put it there. In the end of each part, participants were asked the same comprehension question (when Robert greets his guests, is there an 1804 US silver dollar in his library? [Yes/No]), and at the end, they were asked if Robert "really knows" or "only thinks he knows" that there is an 1804 US silver dollar in his library when he greets his guests and. Similarly to from Starmans & Friedman (2012), they were also asked to score on a confidence scale.

Turri found that most participants in the control condition answered that Robert "really knows" (84%). More interestingly, in the authentic evidence Gettier condition, most answered that Robert "only thinks he knows" (89%). In both conditions, confidence rating exceeded chance. If this is a robust result, then we have evidence that laypeople do not attribute knowledge to authentic evidence Gettier cases when they are presented in a tripartite structure. Turri claims that it is not the tripartite structure *per se* that is leading people to deny knowledge to Robert; otherwise they would also deny knowledge in the control condition, but either way it is a weakness of this experiment's design that it did not also used a condition where the Gettier version would have been presented in a single structure. It would be much more significant if in that single structure condition most participants had answered “really knows”. Another of his experiments pretends to be more convincing, however.

Turri also replicated the result from Starmans & Friedman about the story of Katie and the burglars (2-thief condition), where he also found a majority of participants attributing knowledge to Katie (57%). He then used this as the control condition for a new experiment where he presented an adapted version of the same story in a tripartite structure. In his version, however, one of the burglars stole the pen, and is the other one that left a pen that effectively replaced it. The reason for this, is that despite the intention of Starmans & Friedman to present a story where there is no causal connection between the bad luck and the good luck, it is still easy to see a single source to the bad and good luck in their version (it is because a burglar stole the pen that he left another pen). In addition, Turri also used a condition where the husband of Katie, instead of one the burglars, is the cause why there is a blue pen on the coffee table in the end of the story. This separates the sources of luck more effectively. He found that in the burglars condition, although the difference were not significant, there was a reduction in the rating of knowledge compared to the control condition (44%), and that most participants in the husband condition answered at rating exceeding chance that Katie “only thinks she knows” (76%).

As in the previous experiment, we cannot say that if an authentic evidence Gettier condition to which most people would ascribe knowledge to the agent were presented in a tripartite structure that would make most people to not ascribe knowledge. Turri did not compare the exact same story, but an adaptation of it, to which he not even found a significant difference. A best design in Turri’s experiments would allow us a stronger conclusion. However, these results are still important. He legitimately found authentic evidence Gettier cases to which most people did not attribute knowledge. This contradicts one possible conclusion from Starmans & Friedman that laypeople consider authentic evidence Gettier cases to be knowledge. Now, at least, we have to qualify that[17]. Furthermore, if it is plausi-

[17] Indeed, two previous studies to Starmans & Friedman’s (2012) paper also found results with the majority of the participants not attributing knowledge to subjects in Gettier cases with authentic evidence. Jennifer Wright (2010), for instance, found that most participants gave the expected answer in Gettier cases when it followed a clear case of knowledge or ignorance, but that the same thing did not happen when it followed a less clear case, suggesting a ordering effect to these judgments. Simon Cullen (2010) found the expected answer to Gettier cases, but also found a sensitivity

ble that the tripartite structure really dramatizes the apparent luck involved in Gettier scenarios, then Turri's results support the thesis that laypeople are just less sensible to luck than philosophers.

3.2.1.5. A robust positive result

In a very well designed experiment of Jennifer Nagel, Valerie San Juan, and Raymond A. Mar (2013) found more evidence for the robustness of the Gettier intuition. In their experiment participants had to read and respond about 16 vignettes, randomly ordered in order to avoid potential ordering effects. Half of them were actually fillers, and the other 8 formed four types of experimental vignettes, including Gettier cases. Besides Gettier cases, which is what we are interested in this section, they also tested justified false belief cases, standard true belief cases, and skeptical pressure cases. The objective of Nagel et al. was to create an experiment that would allow the direct comparison of these four types of cases. Participants were also randomly assigned to different stories for each type of vignette, so that each vignette contained a type of story. This allows some control over the possibility of responses being due to other variables in the stories. This experiment also included a last vignette, which tried to replicate the results from Weinberg et al. (2001), along with a series of parameters to evaluate individual differences, which we will not discuss here.

As in other experiments, participants were asked comprehension questions before answering the parameters questions. In this experiment, however, it was included a question about how much the agent is justified – to be answered in a scale. Furthermore, in the knowledge ascription question participants had to choose between answering "yes, he/she knows", "no, he/she doesn't know", and "unclear – not enough information provided in the story". Interestingly, for those who answered "yes, he/she knows", a further question was asked about which of two sentences better describes their opinion, e.g., "(a) Emma knows that the stone

of those answers to the options of response. When the options were "really knows/only believes", instead of "knows/does not know", it raised the denial of knowledge. We will not discuss these studies, however, because although they call attention to important methodological issues, their overall results do not threaten the thesis that Gettier intuitions are robust. On the contrary, both studies attempt to answer another previous study which interprets the finding of an ordering effect as a highly problematic result to the use of intuitions (Swain et al. 2008).

is a diamond", or "(b) Emma feels like she knows that the stone is a diamond, but she doesn't actually know that it is". Answers that chosen (b) were classified as "delayed knowledge denial". This follow-up question serves to measure the robustness of participant's attribution of knowledge, especially in problematic cases such as gettiered ones, and controls the possibility of laypeople being using a non-literal sense of "know" in these cases, which is a factive verb in its literal and philosophical relevant sense. This is important because until this experiment there was no control over this variable.

In the Gettier vignette, what differentiates the two versions of the same story is that one of them is a case of authentic evidence, while the other a false lemma or apparent evidence, as Starmans & Friedman call it, case. Nagel et al. found that although there is a greater tendency to attribute knowledge in authentic evidence cases (2.28 times more likely), most participants (almost 65%) deny knowledge in these cases, being 41.1% of them "immediate knowledge denial" and 23.6% "delayed knowledge denial". Note that without the control of non-literal sense we would register a majority 59% attributing knowledge to the subject. The rating of justification attribution to these cases were also very high (M= 6.08, where 7 stands for "completely justified"). Furthermore, Nagel et al. were unable to find significant differences between ethnic groups while replicating the experiment from Weinberg et al. (2001) relating groups to the attribution of knowledge. Therefore, we have a strong positive result favoring the idea that Gettier intuitions are robust. That is, laypeople seems to have a strong tendency to judge that gettiered justified true beliefs do not constitute knowledge, be it an instance of authentic evidence or otherwise. And no well-designed experiment to date showed that there is significant difference between ethnic groups.

But we still have the negative result about authentic evidence cases from Starmans & Friedman (2012). What can we say about that? One first observation is that the experiment of Nagel et al. has an additional control condition over the non-literal sense of the verb "to know". This raises the possibility of Starmans & Friedman's results are constituted by this non-literal use. This is an open question. Another possibility, which we particularly think is more plausible, is that the pattern they found is the result of something peculiar to the stories they used. Positive results with authentic evidence used different stories, which may be more effective in making participants to consider the relevant features of Gettier cases. But what these features are? We think they are related to the chanciness and the significance of the event. Consider, for instance, the two following authentic evidence cases tested by Nagel et al.

> **Emma case:** Emma is shopping for jewelry. She goes into a nice-looking store, and selects a diamond necklace from a tray marked "Diamond Earrings and Pendants". "What a lovely diamond!" she says as she tries it on. Emma could not tell the difference between a real diamond and a cubic zirconium fake just by looking or touching. In fact, this particular store has a very dishonest employee who has been stealing real diamonds and replacing them with fakes; in the tray Emma chose almost all of the pendants had cubic zirconium stones rather than diamonds (but the one she chose happened to be real).

> **Wanda case:** Wanda is out for a weekend afternoon walk. As she passes near the train station, she wonders what time it is. She glances up at the clock on the train station wall and sees that it says 4:15 pm. What she doesn't realize is that this clock is broken and has been showing 4:15 pm for the last two days. But by sheer coincidence, it is in fact 4:15 pm just at the moment when she glances at the clock.

In both cases, we think it is safe to say, either the chanciness or both the chanciness and significance of the event are salient. This is may be the essential reason Nagel et al. (2013) and Turri (2013) found positive results with authentic evidence cases, but this is speculative and there are still no studies to confirm this hypothesis. Further experiments may convincingly clarify this question. Anyway, considering the best evidence we have, we have no reasons to think that the Gettier intuition is not a robust intuition. Philosophers seem to be right about the robustness of this intuitive judgment and that laypeople do not consider gettiered justified true beliefs to be instances of knowledge. As we will see in the next session, there are now important following questions about the philosophical significance of this *intuitive phenomenon*.

3.2.2. The error-effect

A number of factors can be used to motivate the general theory of *epistemic contextualism*. For instance, it gives a very insightful solution for the *skeptical paradox* or *skeptical puzzle* (Cohen 1986), one which conserves both the appeal of our ordinary attributions of knowledge and of the skeptical conclusion. The main source of evidence for contextualists, however, comes from the use of intuitive epistemic judgments. Roughly, the central thesis of contextualism is that the meaning of knowledge ascriptions changes from context to context, being determined by conversational factors. Evidence for this thesis, therefore, comes from differences in the apparent content of knowledge ascription sentences when they are used in distinct conversational contexts. So contextualism is a theory that clearly relies on the intuitive patterns coming from laypeople's epistemic judgments, which leaves no room for the idea that what matters are the intuitions of experts. DeRose says:

> The best grounds for accepting Contextualism concerning knowledge attributions come from how knowledge-attributing (and knowledge-denying) sentences are used in ordinary, non-philosophical talk: what ordinary speakers will count as 'knowledge' in some non-philosophical contexts they will deny is such in others. This type of basis in ordinary language provides (...) the best grounds we have for accepting contextualism concerning knowledge attributions. (2005, p. 172)

Particular contextualist theories differ about how conversational factors affect the content of knowledge ascriptions and different contextual rules were postulated, but there is one common factor to these theories and which is central to the defense of the general contextualist thesis, *viz.*, the *possibility of error*. The idea is that when error possibilities are made salient in a conversation the context becomes more stringent about the truth conditions of knowledge ascriptions. That is, the *epistemic standards* are raised (Lewis 1979). This is the basic mechanism that allows the contextualist answer to the skeptical paradox. For instance, skeptical conclusions would be true because they are made in a context with extremely high epistemic standards, raised by possibilities of error that seem to be not eliminable. But that fact would not affect the "lower" truth conditions of knowledge ascriptions made in ordinary contexts. Importantly, opponents of contextualism typically do not disagree about the intuition that salient possibilities of errors somehow make it more difficult to take a knowledge attribution as true, but they contest the interpretation that it implies different truth conditions. In other words, opponents of contextualism dispute the thesis that epistemic standards change through contexts.

But is the *error effect*, as we may call it, really a robust intuitive phenomenon? We already have sufficient empirical evidence to solve this matter. The first evidence we have was originated from parallel experiments trying to determine if laypeople epistemic intuitions are sensitive to *stakes*, which is another kind of important intuition whose robustness we will further evaluate. In addition to these, as we mentioned above, the experiment of Nagel et al. (2013) also included vignettes to test the effect of the inclusion of possibilities of error in the stories. Let us start with these experiments whose interest in the error effect was parallel.

3.2.2.1. Error possibilities in bank cases

To motivate contextualism DeRose (1992) used the now famous *bank cases*. DeRose described two versions of the same basic story where he goes to the bank with his wife on a Friday afternoon in order to make a deposit and after facing long lines inside the bank decides to return in the next morning. His wife reminds him that a lot of banks close on Saturdays, but he believes that this bank will be open in the Saturday morning because he was on this bank on the Saturday two weeks ago. The subject is said to have the same evidence and confidence in both cases, but two big differences are present in the second case. First, now is very important that he make the deposit before Monday, otherwise a very high value check will bounce, and, second, his wife raises the possibility of the bank changing its hours: "Banks do change their hours. Do you know the bank will open tomorrow?" (p. 913). In both cases the belief that "the bank will open tomorrow" is true. The putative intuition here is that when evaluating if the subject knows this proposition, we are much less willing to assign knowledge in the second case than in the first one. The only different things here are the stakes, which DeRose takes to be a conversational factor, and the mentioned possibility of error, so they must be what is causing this effect. To determine whether the possibility of error causes an effect by itself we should test it separately.

Wesley Buckwalter (2010) found some negative evidence in an experiment using versions of the bank case. Buckwalter tested three conditions, one version of the story where the subject (Bruno) is in a situation with high stakes, one version where Bruno is not in a high stakes situation, but a possibility of the bank changing hours is mentioned, and a more standard situation to comparison, where Bruno is neither in a high stakes nor an error pressure case. Participants were randomly assigned just one of these conditions, and asked to score a five-point scale (going from "strongly disagree" to "strongly agree" and 3 being a neutral point) about how they agree with the knowledge statement made by Bruno within the story. Buckwalter found that, as expected, most participants agreed with the knowledge statement in standard condition (74.3%), but that most participants also agreed the knowledge statement in the error pressure condition (66.1%). Furthermore, he did not find a statistical significance between the group means.

In a similar design, another experiment from by Joshua May and colleagues (May et al. 2010) also found some negative result when comparing cases with and without the presence of error possibilities. They used four conditions of the same story, two low stakes, and two high stakes, where one low stakes conditions and one high stakes condition had mentioned possibilities of error. They found that mean judgments attributed knowledge in the four conditions and no significant effect of the error possibility raising the denial of knowledge. It is doubtful with we can conclude something from both of these experiments, however. The problem is simply that they lack what now seems to be indispensable controls, such as comprehension questions, control over non-literal sense of "know", and distinct stories to avoid hidden variables. The lack of positive results is not sufficient to empirically deny the robustness of the error effect, we must have reliable negative results. And we have no reasons to trust these experiments.

In fact, another study points to the unreliability of these results. Jonathan Schaffer and Joshua Knobe (2010) suspected that the possibility of error was not salient enough in these experiments to generate an effect. They speculated if they could manipulate the salience by presenting the possibility “in a concrete and vivid fashion” and tested it with versions of the bank cases. They used a control condition (a clear case of knowledge) and a “salient contrast condition” where one of the subjects in the conversation says:

> “Well, banks do change their hours sometimes. My brother Leon once got into trouble when the bank changed hours on him and closed on Saturday. How frustrating! Just imagine driving here tomorrow and finding the door locked.”

Shaffer & Knobe found that while the mean rating of the group that received the control condition agreed with the knowledge attribution to the subject (5.54 out of 7), most participants in the “salient contrast condition” disagreed with it (3.05 out of 7), and that this difference was statistically significant. This, contradicting previous studies, suggests the existence of a strong error effect, one that changes the character of one’s epistemic judgment. This effect was replicated by Buckwalter in a further experiment (Forthcoming).

3.2.2.2. Forgetting about the bank cases

As we mentioned above, the experiment by Nagel et al. (2013) tested four types of cases. Between them, there were cases including error possibilities, which they called “skeptical pressure cases”, to be compared with Gettier cases, justified false belief cases, and standard cases of knowledge. They used the same four basic stories, none of them being a version of DeRose’s bank cases. One of the stories tested in the skeptical pressure condition was a version of Wanda case:

> **Wanda skeptical pressure case:** Wanda is out for a weekend afternoon walk near the train station and wonders what time it is. She glances up at the clock on the train station wall and sees that it says 4:15 pm. It is in fact 4:15 pm at that moment. The station clock is in fact working, but it has no second hand, and Wanda only looks at it for a moment, so she would not be able to tell if the clock were stopped.

Nagel et al. found that just a minority of people attributed knowledge to the subject in this type of condition (39.8%) compared with 72% of control condition, and that this was a significant difference. This result supports the philosophers' expectation about this sort of case and, contrary to the other experiments discussed, it has an exemplary design. Therefore, just like the Gettier intuition, when we weigh the available evidence, we find good reasons to believe that the error effect is an authentic intuitive phenomenon.

Chapter 4

The cognitive bases of intuitions and epistemological theorization

So we have evidence of the robustness of at least two important kinds of intuitions in the epistemological literature. This, however, is not the end for philosophers engaged in the line of investigation we are calling the cognitive turn. The fact that an intuition is robust only leaves the door open for theories that use it in their favor. They were right about its robustness after all and now should not worry with the threats from experimental epistemologists, but it is unclear what the intuition by itself means. Indeed, a single intuition can be used to support a number of different theories. Take the case of the error effect. The fact that the consideration of a possibility of error makes us much less inclined to attribute knowledge to a subject, even he not being aware of that possibility, is used to support, for example, skepticism, epistemic contextualism, and it is accepted by many non-skeptical who are not friendly to contextualism. Skeptics claim that it is the more stringent resultant judgment that reflects the correct standard for knowledge, so once we cannot reach such a high standard we should also deny attributions of knowledge in cases where possibilities of error are not being considered. Contextualists accept that a sentence denying knowledge to subjects in these cases is true, but they claim that the standards for knowledge are not fixed and can vary from context to context. Non-skeptical invariantists, on the other hand, at least those who do not want to be at odds with the consensual intuition, have to find a plausible explanation for this effect in a way that does not imply skepticism or variant standards.

One may then observe that there is much more to say about the nature of an intuitive judgment. Statistical data shows that in fact there are strong relations between certain features of the tested situations and our epistemic judgments, but it just cannot tell us the full story of what explains these outcomes. It shows the existence of causal relations, but the processes that explain them can only be elucidated by other kinds of investigation. So we know that Gettier situations are generally judged as not constituting knowledge, and that the consideration of error possibilities makes people deny knowledge to other subjects which are not considering them. But what explain these judgments? In other words, what is happening in the mental level? This question has the potential to show that some theories are better supported by the intuitions in play than others. For example, it may be the case that the cognitive bases of certain intuitions directly support only one of the theories in dispute or directly disfavor one of them. Indeed, as we will see, some philosophers already tried arguments of this kind by providing psychological explanations for some of the intuitions we are dealing here.

This kind of argument is not without difficulties, however. Many philosophers might have reservations, to say the least, about one using a descriptive psychological analysis in epistemological theorization. For instance, how is that looking to the cognitive basis of an intuition would help to determine the *correct* theory of knowledge? How a descriptive analysis would provide normative content for theorization? In this chapter, we will try to clear up this issue. There are different metaepistemological views about the proper way of doing epistemology or the proper way of interpreting intuitive data, and those views make different claims about the relevance of empirical data, but we cannot really directly discuss them here. Instead, what we will do in this chapter is just to briefly defend one way to make sense of these psychological arguments present in this cognitive turn of epistemology by using the cases of the Gettier intuition and the error-effect. First, we will introduce two attempts to dismiss the error-effect as a relevant intuitive phenomenon for epistemological theory. Then, we will focus on Jennifer Nagel's psychological explanation of the error-effect. An initial problem of her account, recognized by herself, helps us to problematize the use of psychological descriptions in this kind of argument. Lastly, we will argue that we can make sense of her argument by the use of the notion of *reflective equilibrium.*

4.1. THE AVAILABILITY HEURISTIC HYPOTHESIS

Although Williamson is not sympathetic to Hawthorne's subject-sensitive invariantism, they both are critics of the contextualist thesis that epistemic standards vary according the context of the ascriber, and they provided a similar psychological explanation for why the ascriber considering a possibility of error tends to deny knowledge to a subject who is not considering it (Hawthorne 2004; Williamson 2005). Their explanation is based in what is known in social psychology as *availability heuristic* (Tversky & Kahneman 1973). Roughly, the availability heuristic is a cognitive mechanism which makes us to judge the likelihood of an event according the ease with each we can remember or imagine an event of the same type. This is a well-known phenomenon and can explain, for example, why people tend to overestimate the frequency of violent deaths compared to the death from some common diseases, or why someone who heard a remarkable story of a drunken man falling off a fifteen-story building on a car and surviving unscathed may highly overestimate the probability of drunk people surviving falling off buildings.

Applying the heuristics to epistemology, Hawthorne claims that "when certain non-knowledge-destroying counterpossibilities are made salient, we overestimate their real danger; as a result, we may find ourselves inclined to deny knowledge to others in cases where there is in fact no real danger of error" (2004, p. 164). Similarly, Williamson wonders if the lurid possibilities that epistemologists are always raising in their imaginary cases are not creating an illusion of danger. For example, when presenting his idea he questions if "an illusion of epistemic danger result from exposure to lurid stories about brains in vats, evil demons, painted mules, or gamblers who bet the farm?" (2005, p. 226). If they are right, then maybe the fact we are intentionally trying to make ascribers to consider these possibilities causes them to mistakenly think there is a real danger in the agent's situation. Therefore, the error effect could be dismissed as an illusion or a widely shared cognitive bias resulting from this heuristic.

Of course, Williamson and Hawthorne were not trying to directly explain the empirical results we saw above, which had not even been achieved; instead they had in mind the appeal of skeptical arguments that typically appears in epistemology classes. Nagel (2010), however, calls attention to what she thinks constitutes empirical problems for their account. First, when defending their idea, Williamson and Hawthorne seem to overlook the fact that the availability heuristic works both ways, overestimating and underestimating the likelihood of an event, and it is debatable if the error probabilities that appear in imaginary cases are "ease to come to mind". For example, cases involving "brains in vats, evil demons, painted mules, or gamblers who bet the farm" are not ordinary. Most of these, in fact, are very atypical things to be imagined and it seems more reasonable to expect most people to "underestimate" the chances of these events in such a way that they would still attribute knowledge to the agents. So, contrary to what Williamson and Hawthorn suggest, it is doubtful that the availability heuristics can really explain an error effect in cases involving far-fetched possibilities. But what about the data we saw?

Given our discussion about the reasons for the initial failures to detect an error effect, however, there is some sense in talking about the availability heuristic. Remember that experiments began to detect an effect only when Schaffer & Knobe intentionally tried to introduce the possibility of error "in a concrete and vivid fashion". Indeed, even the possibilities used by Nagel in her further study (Nagel et al., 2013) were not so far-fetched, e.g., the possibility of Emma buying a falsification instead of a diamond, and the possibility of Wanda looking to a broken clock. Therefore, given the development of the experiments we saw, it is plausible to think that the positive results we found are at least in part due to the availability of the stories used. We do not have, to date, no evidence of far-fetched possibilities of error causing an error effect. Furthermore, the hypothesis that the availability of the used possibilities caused the detected error effect is perfectly compatible with the empirical evidence we have.

However, another possible empirical problem found by Nagel (2010) is that there is also a well-documented spontaneous discounting that would cause the "underestimation" of the just mentioned possibility of error. Some studies show that we tend to overcompensate when we sense that there is an alternative explanation for the increased availability of what comes to our mind. Particularly, this may be the case when something has just been mentioned to us. For example, in a suggestive experiment, when asked to judge the frequency of the name "Ashcroft" in comparison to "Digby", which is in fact less common, 69.7% of one group of participants correctly judged "Ashcroft" to be more common than "Digby", but in a second group which just read an article mentioning Attorney General John D. Ashcroft's name in the first line, only 42.4% of the subjects judged "Ashcroft" to be more common, a significant drop (Oppenheimer 2004). So even if the mentioning of a possibility of error increases its availability, we have empirical reasons to expect people to spontaneously discount this availability, what would cancel the bias which supposedly constitutes the error effect. This is not what we see, so Williamson and Hawthorne need a further argument to defend that is not the mentioning of the possibilities of error that causes their availability.

One last problem pointed by Nagel is that most heuristics tends to be easily cancelled in certain conditions, such as when we are especially motivated to be accurate, or when we are self-conscious about our judgments. In such conditions we tend to abandon automatic processes and assume a more systematic processing instead. Thinking in the context of epistemology, she says that it "is awkward to posit the activation of a relatively fragile heuristic exactly in conditions that would ordinarily suppress heuristic cognition" (p. 298). In particular, this seems problematic because stringent epistemic judgments do not seem equally cancelable. For instance, the participants of all the experiments we saw in the last chapter were told about the real situation of the agent. The bank was actually open, the stone was really a diamond, the clock was not really broken, etc. If their judgments are the result of the availability heuristic, why this information is so readily ignored? Furthermore, when one goes from an ordinary knowledge attribution to a knowledge denial after being presented with error alternatives, the philosophical consensus is that the natural tendency is to rethink the more relaxed judgment instead of the more stringent one. That is why the skeptical paradox is appealing. The point is that, in general, the reflective reasoning one use to try to justify keeping the knowledge does not cancel the intuition. However, correct information, self-consciousness about one's own reasoning, and motivation for accuracy, for example, are typically enough for cancelling a bias.

So, in conclusion, although there is some sense in thinking about availability explaining the empirical results, the hypothesis of the availability heuristic presents some serious empirical challenges. We have reasons to expect discounting of availability, what would generate results which are different from what we found, and the defeasibleness of heuristics is at odds with the apparent strength of the error effect. Let us look to a more promising psychological explanation

4.2. THE EGOCENTRIC BIAS HYPOTHESIS

After criticizing the availability heuristic hypothesis, Nagel provides a sketch of her own explanation of the error effect, which she later develops in more detail (Nagel 2012). In a similar vein, she also proposes that we can dismiss this effect as illusory because consists in a cognitive bias, but differently from Williamson and Hawthorne, her thesis blames a bias that results from limitations of the processes we use to attribute mental states to other. The problem is not about overestimating epistemic danger, but to judge the knowledge of others from the inputs we have as *mindreaders*. It is obvious that we do not have access to the private thoughts of others, so a basic feature of our mindreading is that it depends on the inputs we have about the evidential position of others. We are especially good, for example, in tracking the perceptual inputs and attention of others in order to calculate the evidence they possess (Baron-Cohen 1994), but we also have limitations which in certain conditions invariably result in cognitive biases.

As we saw in the second chapter, one general type of bias that affects our attributions to others is what is called the *egocentric bias*, a group of deviations patterns characterized by the tendency of making self-serving assessments of oneself or of others (Goethals 1986). A known bias in self-attribution is the *hindsight bias*, sometimes also called the "knew-it-all-along effect", which make us to commonly think that our memory or knowledge of a recently known event is better than it really was in the recent past (Hawkins & Hastie 1990). Attribution to others, in the other hand, can be affected by forms of *epistemic egocentrism* (Royzman et al. 2003), which is a robust tendency to not suppress privileged information when evaluating the mental state of other who are in a more naïve situation. The bias that Nagel (2012) calls attention to is an instance of epistemic egocentric bias. In particular, she develops her argument by pointing to the existence of different strategies of reasoning and how they affect our mindreading processes.

We have different cognitive strategies to deal with problems and they can vary with respect the effort they expend. One common division separates "low" strategies, which are heuristic and effortless in character, and "high" strategies, which are more effortful, sequential and conscious in character, typically involving the consideration of alternatives. This division follows the *dual process theory*, a general psychological theory that says there are at least two distinct types of processes underling a number of different psychological processes (Frankish & Evans 2009), including social judging, categorization, mindreading, probability assessment, etc. Importantly, two distinct processes can produce different outcomes for the same problem, e.g., a subject can make a certain probability assessment through a heuristic processes, and deliver a different outcome through a more slow and conscious reasoning. These processes vary in accuracy according the conditions they are triggered. Also, different types of processes can interact with each other, as the outcome of one serves as the input of other.

Nagel claims that the conditions in which we judge skeptical pressure cases causes us a high level process of judgment. We are told a possibility of error and induced to think about the epistemic consequences of this alternative situation, so we assume a sequential reasoning. One possible explanation, therefore, is that from this reasoning state we intuitively *misrepresent* the subject as having the same concerns as us and as failing to meet them. That is, they should not form a rightful belief in those cases. Thus, we are committing a form of epistemic bias. Nagel (2012) however, latter rejects this hypothesis. The problem is that it seems very implausible since, as she and her colleagues showed (Nagel et al., 2013), we judge subjects in skeptical pressure cases to be justified in their beliefs. So there is little sense in saying that we represent them as thinking just like us, and somehow failing in doing what they should, when in fact we judge them to be justified in their belief.

Nagel then proposes that we not really represent subjects in skeptical pressure cases as thinking like us, with more elaborated strategy, but that we use this higher strategy as "benchmark" to evaluate the subject's situation. Citing the case of Wanda, she says:

> [W]e don't have to ascribe this higher strategy to the observed subject (even implicitly) in order to feel that she is falling short of knowing: if we intuitively take the appropriateness of our own cognitive strategy for granted, then rather than representing Wanda as attempting but failing at our more complex way of thinking, we could more simply be sanctioning Wanda for her failure to adopt either our cognitive strategy or

> the range of evidence we now find intuitively necessary, given the strategy we have adopted. (2012, p. 186)

In other words, error possibilities induce us to think what would be necessary to know in those alternative situations and we use that reasoning as a benchmark to evaluate the epistemic situation of others[18]. Contextualists are in perfect agreement with this. However, the idea is that once we realize the psychological mechanism underling this judgment, we can dismiss it as a case of epistemic egocentrism. For instance, note that the subject himself would adopt a different strategy to judge his epistemic situation in these cases. As Nagel says, "the default style of judgment in these circumstances is routine and automatic: if she has no particular reason to worry, Wanda would naturally go from looking at the clock to forming a belief about the time without any personal-level reflection on the basis of her judgment" (p. 179). We attribute knowledge to ourselves and to others through this kind of judgment in countless situations, in cases where possibilities of error are not mentioned, and we only evaluate Wanda or any other subject in a skeptical pressure case as we do, because we impose our perspective. This is a distortion like any other egocentric bias.

Note that this psychological analysis is in perfect tune with a simulationist account of mindreading. Nagel herself tries to remain neutral about the dispute between TT and ST, but we are no longer in position to do the same. The error-effect is naturally explained as an instance of simulative mindreading. Roughly, we evaluate Wanda by "trying" to put ourselves in her shoes, but being called attention to the possibility of the clock being broken, we adopt a different reasoning strategy than Wanda's and somehow we intuitively "disapprove" either her mental state or her reasoning strategy. We still do not have a complete account to fill this very rough description, but the process obviously has a simulative flavor. Furthermore, to explicitly interpret the error-effect as a case of simulative mindreading motivates the idea that it is an instance of a bias. If this is the correct psychological analysis of the error-effect, then we have a solid argument against the use of this intuition as a valid evidence for epistemological theories.

[18] This is better explained by the psychological notion of *epistemic anxiety*, the subjective feeling of how much evidence one needs to achieve a rightful belief. For a detailed discussion, see Nagel (2010a).

One important implication of such a psychological analysis arises here, however. If we are willing to call the error-effect a cognitive bias, then we have to say the same of the Gettier intuition. In Gettier cases we are equally induced to a higher strategy by an error possibility and we are equally making a judgment available only from our perspective. It is because we know that the necklace chosen by Emma is the only jewelry with real diamonds that we are led to adopt a different reasoning strategy when thinking about the case – maybe in the process of simulating her state – and to disapprove something in her mental states. The problem is that most epistemologists are not willing to call the Gettier intuition a cognitive bias. Gettier cases caused a great impact in epistemology and much of the development of the literature is due to the relative consensus that they constituted a serious problem to the traditional analysis. However, if the intuition is just an explicable bias, then defenders of the traditional analysis should not worry about these cases anymore. And since this intuition actually has the same cognitive basis of another intuition that we are willing to call a bias, this seems the reasonable conclusion. Nagel recognizes this point and considers it a problem because she is part of the majority of epistemologists who want to say that people in Gettier situations do not know and that people in skeptical pressure cases are in knowledge states.

4.3. THE METHODOLOGY QUESTION

So, must we simply accept that we should dismiss Gettier intuitions along with the error-effect? Should we conclude this by this psychological argument alone? Despite the strong naturalistic tone of our investigation so far, we think that the answer is a negative one. More exactly, our answer is a timorous “not exactly”. We do not think that the proper way of doing epistemology forces us to reject the intuition. But, more generally, what is the proper way of providing a theory of knowledge? Of course, there is no obvious or consensual answer here. There are different characterizations of the epistemological project and philosophers’ answers depend on the project they are carrying on. If one, for example, is convinced that there is no way of achieving the correct theory of knowledge by looking to our intuitions then a psychological argument like this would only be example of the futility of consulting imaginary cases. Hilary Kornblith (2002) advocates a view like that as he claims that knowledge is a natural kind and that the analysis of knowledge could only reveal the ordinary concept of knowledge. Instead of looking for the ordinary concept we should look to the natural kind. On the other hand, one may criticize the very idea of intuitions being evidence of conceptual matters. Timothy Williamson (2007) seems to do that when he criticizes what he calls the "psychologization of evidence". For him, "few philosophical questions are conceptual questions in any distinctive sense" (p. 3), so we should not consider evidence which actually concerns our concepts. However, very distant from a project like Kornblith's, Williamson's intention is to defend the *a priori* methodology for epistemology. That is, he argues that one can achieve the correct theory of knowledge by a priori methodology alone. So, again, there is dispute about what is the proper way of seeking for a theory of knowledge.

We cannot directly confront these views here, but we must say that we are not inclined to neither of these views. For instance, we have trouble believing that one can characterize *a priori* methodology in such a way that it runs free of any psychological consideration. We do not oppose the idea of *a priori* methodology or knowledge itself, not at all, but as far as we believe they have a natural basis, we cannot see how these notions are immune to psychological considerations. We cannot see, for example, how such a methodology could not involve attributions of knowledge (Brown 2012), and it should be clear now how psychological considerations about our processes of knowledge attribution can be relevant for the analysis of knowledge. Is it not possible that there are different kinds of reflective categorizations or that reflective reasoning can be unreliable in certain circumstances? It is not possible, then, that these possible problems affect that very *a priori* methodology? If this is true, then either every relevant psychological consideration should be *a priori* or we should at least implement this methodology with empirical investigation.

On the other hand, Kornblith claims that in order to learn about knowledge as a natural kind we should look to what ethologists and other cognitive scientists describe as knowledge in their field, instead of looking to our intuitions. We do not think this is an inconsistent proposal, but we have trouble believing it does justice to the traditional project. For instance, one essential aspect of the epistemological project is its intuitive methodology. Epistemologists are interested in imaginary cases and make extensive use of intuitions. A project that has no place for intuitions and is exclusively empirical seems to distance itself too much from the philosophical practice (Goldman 2007). Kornblith can bite the bullet and assert that the traditional project is really flawed, but what sustains this conclusion? Nothing we saw so far, even the negative conclusion of the second chapter, determines the failure of the project, at least not if we minimally characterize it as the attempt of achieving the *correct* theory of knowledge. The conclusion from the structural question is that it seems very unlikely that we succeed in finding an intuitively satisfactory definition of knowledge. But that does not mean that we cannot achieve a satisfactory theory of knowledge or the most satisfactory theory possible. If we can do this without mischaracterizing the project too much, e.g., by transforming it in a pure empirical project, and also accommodating the psychological considerations we are seeing, then we can justify the traditional appeal to intuitions in epistemology and at the same time have a naturalistically respectable project – even if we have to abandon the initial pretension of the analysis of knowledge for that. Do we have a way to do this? We think so.

4.3.1. Reflective equilibrium

We can make sense of a methodology that uses both intuitive ascriptions and psychological considerations through the well-known notion of *reflective equilibrium.* This notion was first proposed by Nelson Goodman (1955) when dealing with the justification of our rules of inductive and deductive rules and gained notoriety when John Rawls (1971) articulated and applied it to political philosophy. Roughly, the idea is that certain activities involve adjustments between the judgments or intuitions one is inclined to make and the principles that one believes govern these judgments. To use the example of the domain of ethics, it is plausible to say that a moral agent starts only with some set of initial, maybe spontaneous, moral judgments and progressively form more considered moral judgments. From this, a moral agent can to develop more explicit principles or a theory which explain these judgments. When faced with a certain case, however, one may discover that he is naturally inclined to make a moral judgment *a*, but also that the explicit principles or theory that he thinks determine what is morally right or wrong dictates that this case should be considered *b*. So the agent is in a situation where either his judgment or what he explicitly believes must be modified if he does not want to continue in an inconsistent state. The reflective equilibrium, therefore, would be such a state where one is successful in finding a coherent point between one's judgments and principles.

The idea is not that the point of equilibrium be achieved anyhow, but that it is the end of rational process. For this, empirical, theoretical and reflective considerations can have an essential role, with the possible consequence of beliefs being rejected and new ones being acquired. In a way, it is a matter of having more reasons to keep the judgments or intuitions, or to keep the beliefs, principles, etc., which makes reflective equilibrium not only an ending state, but a proper methodology for justifying the judgments and principles or theory one has. Also, this methodology does not refer only to the mutual adjustment between judgments and a particular theory, but also to the confrontation of rival theories, what is called *wide reflective equilibrium* (Rawls 1971). For example, to test our moral judgments with the claims of distinct moral theories is a straight way of trying to determine the better of them. Of course, reflective equilibrium is not restricted to ethics, but applies to a number of fields. Relevant for us, it can be applied to the theoretical practices of epistemologists and, more specifically, give a proper role to intuitions and the psychological considerations we are discussing.

We can make sense of their theoretical practice by noting first that intuitive ascriptions have a fundamental role in it. Intuitive ascriptions work as maybe the most important observational basis from which theoretical principles or specific analyses are proposed. That is why a statistical discovery showing that philosophers are wrong about the intuitive consensus would be a very impactful result. Second, these intuitive judgments are not the only factor taken in consideration in theorization. Other factors like logical considerations and explanatory criterion are often used in argumentation along with intuitions. For example, one can motivates one's position by arguing it accommodates a very plausible logical principal like *epistemic closure* (Pritchard 2005), or one can argue that one's theory does not use any concept that is not acceptable from the scientific point of view (Goldman

1986). Third, what is very relevant here, intuitive judgments themselves, despite all the confidence we assign to them, are not sovereign. You can do considerations that alter the status of an individual intuition or an intuitive pattern. For instance, you can find a reason to think that a certain intuition is flawed. This is the case of the psychological arguments we saw in this chapter. If we conclude that a certain intuitive pattern consists in a psychological bias, then we have a very strong empirical reason to disfavor it and the theories that sustain themselves on it, which gives psychological considerations a great importance in epistemology.

We can now return to the more specific question we saw above: So must we simply accept that we should dismiss Gettier intuitions? Should we rescue the traditional analysis from epistemology's limbo? Again, not exactly. The problem is that a psychological analysis that says that a certain cognitive pattern is a bias consists is an epistemological consideration itself. To call a response pattern a bias is to say it is epistemically bad; it is an evaluative or normative judgment. So, like any other principle or theoretical consideration, this evaluation is itself subject to adjustment. In particular, it can be an instance of *narrow reflective equilibrium*. That is, we can focus only in this particular assessment and the intuitive judgments it is supposed to explain – in contrast of opposing the different interpretations of the intuition by distinct epistemological theories of knowledge. If we have reasons to keep it, or no reason to modify it, then reflective equilibrium dictates that the intuitive judgment that denies knowledge to subjects in Gettier and skeptical pressure cases should be disadvantaged relative to this psychological assessment. What is the reason to call the error-effect and the Gettier intuition instances of egocentric bias? The idea is that there are different mindreading processes, some of them attributing knowledge to the subjects, and that the ones denying knowledge have this outcome only because there is a failure in inhibiting one's own mental states when mindreading (probably simulating) others. An ideal mindreading process, one without possibilities of error being mentioned, for example, would not have such an outcome. One can, however, dispute this very assessment. For instance, one may argue that actually there is something appropriate in the intuitive outcome of these cases, or that there is something appropriate in one of them. If someone can present a reason for this then one may favor the intuition in relation to such a psychological assessment. Indeed, being an invariantist about knowledge and willing to deny knowledge to subjects in Gettier situations, this is what Nagel tries to do.

Nagel does not deny that the same cognitive basis is response for the judgments in Gettier and skeptical pressure cases, but she claims that a big difference appears *on reflection*:

> In skeptical pressure cases we can appreciate on reflection that the agent is succeeding at a cognitive task that can be performed very simply. In Gettier cases we can appreciate on reflection that the agent is failing to execute the more complex type of thinking that would be needed for knowledge in her environment. (2012, p. 186)

This can be further defended if we look at the consequences in case the subjects were to search for more evidence in the two types of situations. In a skeptical pressure case, the subject would confirm his belief and no qualitative difference is apparent in comparison to the more automatic judgment that he would have done otherwise, e.g., Wanda would see that the clock is really working. In Gettier cases, in contrast, the search for more evidence would result in a different outcome and the subject would need more reasons to secure his knowledge, e.g., Wanda would see that that clock is broken, and would need to discover the hours elsewhere. Therefore, the privileged information we have when judging agents in gettiered scenarios do not form an instance of cognitive bias, but really say something about the subject's epistemic state.

Thus, Nagel finds a way of keeping both the intuition from Gettier cases and her assessment that the error effect is actually an instance of epistemic egocentrism. More importantly here, although she not explicitly characterize her reasoning this way, we have a clear example of how reflective equilibrium can make sense of psychological considerations from of what we are calling cognitive turn. It is not only because there are distinct psychological processes, some of which attribute knowledge to those subjects, that knowledge denial became a cognitive bias. And it is not only because the intuitions from two different kinds of cases have the same cognitive bases that both have the same status in epistemological theorization. The psychological and epistemological aspects of an intuition are much more intertwined than it might initially appear. What determine their status for epistemology is a reflective methodology which assesses both the reasons for using an intuitive pattern in favor of particular theories and the reasons for keeping a theory or its interpretation of the intuition. This can be done through wide reflective equilibrium, as, for example, we use statistical discoveries or make psychological arguments to favor one theory's interpretation of the intuition, or through narrow reflective equilibrium, as, for example, we reflect on the very reasons to consider one intuition a cognitive bias. In conclusion, therefore, we can make sense of the cognitive turn with the notion of reflective equilibrium, wide and narrow. That allows one to defend a particular theory of knowledge in face of cognitive considerations and, more generally, defend the maintaining of the project of the analysis of knowledge. Maybe we cannot achieve a definition of knowledge that is intuitively satisfactory, but we may achieve the more satisfactory theory possible.

Conclusão

Alcançamos quatro conclusões nesta tese. No primeiro capítulo, após discutirmos a plausibilidade de hipóteses estruturais básicas da literatura em psicologia de conceitos aplicadas ao caso de CONHECIMENTO, concluímos, contrariamente à suposição da visão ortodoxa da filosofia, que o conceito ordinário de conhecimento não só é um conceito de um estado mental, como também não é um conceito parcialmente composto por CRENÇA, sendo ordinariamente visto como um estado mental próprio. A avaliação dessas hipóteses básicas também tornou claro a importância de uma teoria sobre a estrutura de CONHECIMENTO ser capaz de responder (Q1) porque epistemólogos tem tido grande dificuldade de gerar uma definição de conhecimento intuitivamente satisfatória e (Q2) como esse conceito é adquirido. No segundo capítulo respondemos a questão que intitula esta tese. Nos vendo obrigados a adentrar na extensa literatura sobre a teoria teoria e a teoria simulacionista, derivamos duas hipóteses sobre a estrutura de CONHECIMENTO oriundas dessas teorias, incluindo uma surpreendente hipótese de que o conceito ordinário de conhecimento é de fato primitivo. Após avaliarmos a melhor evidência disponível para cada uma dessas hipóteses, assim como seu poder explicativo, concluímos que a resposta mais razoável é uma posição híbrida no qual categorizações intuitivas de conhecimento podem ser realizadas tanto por instâncias de processos simulativos como por instâncias de inferências teóricas. Ambas as teorias apresentam evidência empírica positiva, *viz.*, inclinações egocêntricas e erros sistemáticos, e podem responder (Q1) e (Q2). Isto implica que CONHECIMENTO possui tanto um conteúdo (possivelmente primitivo) que permite identificação introspectiva, como armazena generalizações teóricas, i.e., armazena informação causal, ou modal, ou funcional, etc. Uma lição importante para a epistemologia, além das razões positivas para duvidar da possibilidade de gerarmos uma definição de conhecimento intuitivamente satisfatória, é que atribuições intuitivas de conhecimento podem sofrer de dois tipos de erro: egocentrismo e generalizações corretas.

Na segunda parte da tese tratamos de outros tipos argumentos presentes na linha de investigação cognitivista que estamos interessados. No terceiro capítulo, após revisar uma série de trabalhos experimentais sobre a intuição Gettier e a intuição a partir de casos de pressão cética, concluímos que estas intuições podem ser consideradas robustas para o uso na teorização epistemológicas. No quarto capítulo, no entanto, mostramos que descobertas estatísticas positivas por si só não garante um status positivo para intuições. Em particular, considerações sobre as bases cognitivas de uma intuição podem ser usadas para alterar seu status evidencial para uma teoria epistemológica. Apresentamos a explicação psicológica de Nagel paras as duas intuições e como ela gera um problema para epistemólogos que pretendem manter o valor de face da intuição Gettier, mas rejeitar a intuição que surge a partir de casos de pressão cética. Por fim, apresentamos como podemos dar sentido a esta linha cognitivista de argumentos através da noção de equilíbrio reflexivo. Isto nos permite mostrar tanto que argumentos cognitivistas podem eles mesmos ser revisados, como podemos argumentar a favor da manutenção do projeto tradicional da análise do conhecimento.

Referências bibliográficas

Aikhenvald, A. (2004). *Evidentiality*. New York, NY: Oxford UP.

Apperly, I. (2011). *Mindreaders*: *The Cognitive Basis of "Theory of Mind"*. Hove and New York: Psychology Press.

Apperly, I., Robinson, E. (2003). "When can children handle referential opacity? Evidence for systematic variation in 5-and 6-year-old children's reasoning about beliefs and belief reports". *Journal of Experimental Child Psychology*, 85(4), 297-311.

Baron, J. & Hershey, J. (1988). "Outcome bias in decision evaluation". *Journal of Personality and Social Psychology,* 54, 569-579.

Baron-Cohen, S. (1994). "The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology". In C. Moore & P. Dunham (Eds.), *The Role of Joint Attention in Development* (pp. 41-59). Mahwah, NJ: Erlbaum.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). "Does the autistic child have a 'theory of mind'?". *Cognition* 21:37-46.

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). "Recognition of mental state terms. Clinical findings in children with autism and a functional neuroimaging study of normal adults." *The British Journal of Psychiatry*, 165(5), 640.

Bartsch, K., Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.

Bealer, G. (1998). "Intuition and the Autonomy of Philosophy". In DePaul, M. and Ramsey, W. (eds.) *Rethinking Intuition*. Oxford: Rowman & Littlefield Publishers, Inc., 201-39.

Bealer, G. (2002). "Modal epistemology and the rationalist renaissance". In: Szabo Gendler T., Hawthorne J. (eds), *Conceivability and possibility*. Oxford University Press, Oxford.

Birch, S., & Bloom, P. (2003). "Children are cursed: an asymmetric bias in mental-state attribution". *Psychological Science*. 14, 283–286.

Birch, S. & Bloom, P. (2007). "The curse of knowledge in reasoning about false beliefs". *Psychological Science*, 18(5), 382–386.

Blank, H., & Nestler, S. (2007). “Cognitive Process Models of Hinsight Bias”. *Social cognition*, 25(1), 132-147.

Block, N. (1986). “Advertisement for a semantics for psychology”. In P. A. French, T. E. Uehling Jr. and H. K. Wettstein (eds.), *Midwest studies in philosophy X: Studies in the philosophy of mind*. Minneapolis: University of Minnesota Press.

Bonjour, L. (1980). “Externalist Theories of Empirical Knowledge,” *Midwest Studies in Philosophy* 5: 53-73.

Brandt, M. (1978). “Relations between cognitive role-taking performance and age”. *Developmental Psychology*, 11, 206-213.

Bräuer, J., Call, J., Tomasello, M. (2008). “Chimpanzees do not take into account what others can hear in a competitive situation”. *Animal Cognition*, 11, 175-178.

Brooks, L. (1978). “Nonanalytic concept formation and memory for instances”. In *Cognition and concepts*, ed. E. Rosch and B. B. Lloyd, 169–211. Hillsdale, NJ: Erlbaum.

Brown, J. (2012). “Words, concepts, and epistemology”. In Brown and Gerken (eds.), *Knowledge Ascriptions*, OUP: Oxford, 32-54.

Brown, J., & Gerken, M. (2012). “Introduction” in Brown, J., & Gerken, M. (eds), *Knowledge Ascriptions*. OUP: Oxford, 1-30.

Brueckner, A. (2002). “Williamson on the primeness of knowing”. *Analysis*, 62 (275), 197-02.

Buckwalter, W. (2010). “Knowledge Isn’t Closed on Saturdays,” *Review of Philosophy and Psychology*, 1 (3):395-406.

Buckwalter, W. (Forthcoming). “The Mystery of Stakes and Error in Ascriber Intuitions”. In James Beebe (ed.), *Advances in Experimental Epistemology*, Continuum.

Buckwalter, W., & Stich, S. (2010). “Gender and Philosophical Intuition”. Available at SSRN: http://ssrn.com/abstract=1683066

Call, J., Tomasello, M. (2008). “Do chimpanzees have a theory of mind: 30 years later.” *Trends in Cognitive Science*, 12, 187-192.

Cappelen, H. (2012). *Philosophy Without Intuitions*, Oxford: Oxford University Press.

Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge: MIT Press.

Carey, S. (1995). On the origin of causal understanding . In D. Sperber , D. Premack and A.J. Premack (eds.), *Causal Cognition: A Multidisciplinary Debate*. Oxford: Clarendon.

Carlson, S., Moses, L., & Hix, H. (1998). "The role of inhibitory processes in young children's difficulties with deception and false belief". *Child Development*, 69, 672-691.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Churchland, P. M., (1991). Folk psychology and the explanation of human behavior. In: J. D. Greenwood (Ed.). *The future of folk psychology*. Cambridge: Cambridge University Press. 51–69.

Churchland, P. M. (1991). Folk psychology and the explanation of human behavior. In: J. D. Greenwood (Ed.). *The future of folk psychology*. Cambridge: Cambridge University Press. 51–69.

Clements, W. A., & Perner, J. (1994). "Implicit understanding of belief." *Cognitive Development*, 9, 377-397.

Cohen, S. (1986). "Knowledge and Context", *The Journal of Philosophy*, 83: 574–583.

Cohen, S. (1988). "How to be a Fallibilist", *Philosophical Perspectives* 2, 91-123.

Colaco, D., Buckwalter, W., and Stich, S. (Forthcoming). "Epistemic intuitions in fake-barn thought experiments". To appear in *Episteme*.

Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology*, 1, 275-296.

Cummins, R. (1998). "Reflection on Reflective Equilibrium". In DePaul, M. and Ramsey, W. (eds.), *Rethinking Intuition*. Oxford: Rowman & Littlefield Publishers, Inc, pp. 113-27.

Currie, G. & Ravenscroft, I. (2002). *Recreative Minds*. Oxford: Oxford University Press.

Davidson, D. (1963). "Actions, reasons, and causes". *The Journal of Philosophy*, 685-700.

De Villiers, J. (2007). “The interface of language and Theory of Mind.” *Lingua,* 117(11), 1858-1878.

Decety, J. & Greze, J. (2006). “The power of simulation: Imagining one's own and other's behavior”. *Brain Research* 1079: 4-14.

Dennett, D. (1971). “Intentional systems”. *The Journal of Philosophy*, 68(4), 87-106.

DeRose, K., (1992). “Contextualism and Knowledge Attributions”, *Philosophy and Phenomenological Research*, 52(4): 913–929.

DeRose, K. (2005). “The ordinary language basis for contextualism and the new invariantism”. *The Philosophical Quarterly*, 55, 219 pp. 172-198.

DeRose, K. (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context*: Vol. 1, Oxford: Clarendon Press.

Dummett, M. (1993). *Seas of Language*, Oxford: Oxford University Press.

Dunn, J., & Dale, N. (1984). I a Daddy: Two year old's collaboration in joint pretend with sibling and with mother. In I. Bretherton (Ed.) *Symbolic play and the development of social understanding* (pp. 131- 158). New York: Academic Press.

Elman, J. L., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*, Cambridge, MA: MIT Press.

Fiengo, R. (2003). “Linguistic Intuitions”, *Philosophical Forum*, 34, 253–66.

Fischhoff, B. (1975). “Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty”. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299

Fletcher, L., Carruthers, P. (2012). “Behavior-Reading versus Mentalizing in Animals”. In J. Metcalfe & H. Terrace (Eds.), *Agency and Joint Attention*. New York: Oxford.

Fodor, J. (1994). “Concepts: A potboiler.” *Cognition*, 50, 95-113.

Fodor, J. (1998). ‘There are no recognitional concepts – not even RED’ in *In Critical Condition*, Cambridge, MA: MIT Press, pp. 35–47.

Frankish, K., & Evans, J. (2009). “The duality of mind: an historical perspective”. In K. Frankish & J. Evans (Eds.), *In Two Minds: Dual Process Theory and Beyond* (pp. 1–29). Oxford: Oxford University Press.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of Royal Society B: Biology*.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences* 2: 493-501.

Gallese, V., Keysers, C. & Rizzolatti, G. (2004). "A unifying view of the basis of social cognition." *Trends in Cognitive Sciences* 8: 396–403.

Garnham, W. A., & Perner, J. (2001). "Actions really do speak louder than words – but only implicitly: Young children's understanding of false belief in action. *British Journal of Experimental Psychology*, 19 (3), 413-432.

Gettier, E. (1963). "Is Justified True Belief Knowledge?". *Analysis* 23: 121-123.

German, T., & Leslie, A. (2001). "Children's inferences from knowing to pretending and believing". *British Journal of Developmental Psychology*, 19, 59–83.

Glucksberg, S., Krauss, R. M., & Higgins, E. (1975). The development of referential communication skills". In F. D. Horowitz, E. M. Hetherington, S. Scarr-Salapek, & G. M. Siegel (Eds.), *Review of child development research* (Vol. 4, pp. 305-345). Chicago: University of Chicago Press.

Goethals, G. R. (1986). "Fabricating and ignoring social reality: Self-serving estimates of consensus". In J. Olson, C. P. Herman, & M. P. Zanna (Eds.), *Relative deprivation and social comparison: The Ontario Symposium on Social Cognition* (Vol. IV, pp. 135-157). Hillsdale, NJ: Lawrence Erlbaum

Goldman, A. (1967). "A causal theory of knowing." *Journal of Philosophy*, 64, 357–372.

Goldman, A. (1976). "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73: 771-791.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. (1992). "Epistemic folkways and scientific epistemology". In *Liaisons: Philoso-phy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.

Goldman, A. (1995). "In defense of the simulation theory." In M. Davies and T. Stone (eds.), *Folk Psychology*. Oxford, Blackwell, pp. 191–206.

Goldman, A. (2006). *Simulating Minds.* Oxford, Oxford University Press.

Goldman, A. (2007). "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status," *Grazer Philosophische Studien*, 74, 1–26.

Goldman, A., & Pust, J. (1998). "Philosophical Theory and Intuitional Evidence" in M. DePaul and W. Ramsey, eds., *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, pp. 179-197, Rowman & Littlefield.

Goldman, A. & Sripada, C. (2005). Simulationist models of face-based emotion recognition. *Cognition* 94: 193-213.

Goodman, N. (1955) *Fact, Fiction, and Forecast, Cambridge*, MA: Harvard University Press.

Gopnik, A. (1996). "The scientist as child." *Philosophy of Science*, 63, 485-514.

Gopnik, A., Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA, MIT Press.

Gopnik, A., Schulz, L. (2004). "Mechanisms of theory-formation in young children." *Trends in Cognitive Sciences* 8(8): 371–377.

Gopnik, A., Wellman, H. (1992). "Why the child's theory of mind really is a theory". *Mind and Language* 7: 145-171.

Gopnik, A. (1996). "The scientist as child." *Philosophy of Science*, 63, 485-514.

Gordon, R. (1986). "Folk psychology as simulation." *Mind and Language*, 1: 158–171;

Gordon, R. (1995). "Simulation without introspection or inference from me to you." In M. Davies and T. Stone (eds.), *Mental simulation: Evaluations and Applications*. Oxford, Blackwell, pp. 53–67.

Gordon, R. (1996). "Radical simulationism." In P. Carruthers and P. Smith (eds.), *Theories of theories of mind*. Cambridge, Cambridge University Press, pp. 11–21.

Grèzes, J., Decety, J. (2001). "Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis". *Human Brain Mapping* 12:1–19.

Hampton, J. (1981). "An investigation of the nature of abstract concepts." *Memory and Cognition* 9 (2).

Harris, P. (1992). 'From Simulation to Folk Psychology: The Case for Development', *Mind and Language* 7: 120–44;

Hare, B., Call, J., Tomasello, M. (2001). "Do chimpanzees know what conspecifics know?". *Animal Behaviour*, 61 (1), 139–151.

Hare, B., Tomasello, M. (2004). "Chimpanzees are more skillful in competitive than in cooperative cognitive tasks". *Animal Behaviour*. 68, 571-581.

Hawkins, S. & Hastie, R. (1990). "Hindsight: Biased judgments of past events after the outcomes are known". *Psychological Bulletin*, 107, 311-327.

Hawthorne, J. (2004). *Knowledge and Lotteries*, New York and Oxford: Oxford University Press.

Hawthorne, J. (2004a) Replies to Commentators. *Philosophical Issues* 14, 510-523.

Heal, J. (1986). 'Replication and Functionalism', in J. Butterfield (ed.), *Language, Mind and Logic*, Cambridge: Cambridge University Press, 135–50.

Heyes, C. (1998). "Theory of mind in nonhuman primates". *Behavioral and Brain Sciences*, 21 (1), 101–134.

Hintikka, J. (1999). "The Emperor's New Intuitions", Journal of Philosophy 96(3), 127–147.

Jackman, H. (2009). "Semantic Intuitions, Conceptual Analysis and Cross-Cultural Variation". *Philosophical Studies*, vol. 146, n. 2, 159-177.

Jackson, F. (1994). "Armchair Metaphysics". In F. Jackson (1998) *Mind, Method, and Conditionals: Selected Essays*. London: Routledge, 154-76.

Jackson, F.. (1998). *From Ethics to Metaphysics*. Oxford: Oxford University Press.

Kahneman, D., Tversky, A. (1982). The simulation heuristic. In: D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty*. Cambridge: Cambridge University Press. 201–208.

Kaminski, J., Call, J., Tomasello, M. (2008). "Chimpanzees know what others know, but not what they believe." *Cognition*, 109(2), 224-234.

Kauppinem, A. (2007). "The rise and fall of experimental philosophy". In *Philosophical Explorations* 10 (2):95 – 118.

Kornblith, H. (2002). *Knowledge and its Place in Nature*. Oxford University Press.

Kornblith, H. (2007). "Naturalism and Intuitions" in *Grazer Philosophische Studien* 74, 27–49.

Kurdek, L. (1977). "Structural components and intellectual correlates of cognitive perspective taking in first- through fourth-grade children". *Child Development*, 48, 1503-1511.

Kyburg, Henry. (1961) *Probability and the Logic of Rational Belief*, Middletown: Wesleyan University Press.

Laurence, S., & Margolis, E. (1999). “Concepts and Cognitive Science”. In E. Margolis & S. Laurence (eds.) *Concepts: Core Readings* (p. 3-81). Cambridge, MA: MIT Press.

Laurence, S., & Margolis, E. (2003). “Concepts and Conceptual Analysis”. *Philosophy and Phenomenological Research* 67 (2):253-282.

Lehrer, K. (1990). *Theory of Knowledge*. Boulder, Co: Westview.

Leslie, A. (1987). “Pretense and representation: The origins of ‘theory of mind’”. *Psychology Review* 94:412-426.

Leslie, A. (2000). ‘How to acquire a representational theory of mind’. In D. Sperber, & S. Davies (Eds.), *Metarepresentation*. Oxford: Oxford University Press.

Lewis, D., (1979), “Scorekeeping in a Language Game”, *Journal of Philosophical Logic*, 8: 339-359.

Lewis, D. (1994). “Reduction of Mind”. In D. Lewis (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, 291-324.

Lewis, C., & Osborne, A. (1990). “Three-year-olds' problems with false belief: conceptual deficit or linguistic artifact?”. *Child Development*, 61, 1514-1519.

Lewis, C, Freeman, N., Hagestadt C., & Douglas H. (1994). “Narrative access and production in preschoolers’ false belief reasoning”. *Cognitive Development*, 9, 397-424.

Lord, C., Lepper, M. & Preston, E. (1984), “Considering the opposite: a corrective strategy for social judgment”. *Journal of Personality and Social Psychology*, 47, 1231-1243.

Machery, E. (2009). *Doing Without Concepts*, New York: Oxford University Press.

Margolis, E. (1995). “What is conceptual glue?” *Minds and Machines*, 9, 241-255.

May, J., Sinnott-Armstrong, W., Hull, J. G., and Zimmerman, A. (2010). “Practical Interests, Relevant Alternatives, and Knowledge Attributions: An Empirical Study,” *Review of Philosophy and Psychology*. 1, 265-273.

McCloskey, M. (1983a). Naive theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 75-98). Hillsdale, NJ: Lawrence Erlbaum.

McCloskey, M. (1983b). “Intuitive physics”. *Scientific American*. 248:122-130

McCloskey, M., Caramazza, A., & Green, B. (1980). “Curvilinear Motion in the Absence of External Forces: Naïve Beliefs about the Motion of Objects”. *Science*, 210, 1139-41.

McCloskey, M., Glucksberg. S. (1979). “Decision processes in verifying category membership statements: Implications for models of semantic memory”. *Cognitive Psychology*, 11, 1-37.

McCloskey M., Washburn A. & Felch L. (1983). “Intuitive Physics: The Straight Down Belief and Its Origin”, *Journal of Experimental Psychology:* Learning, Memory and Cognition, Vol 9 (4), 636-649.

Medin, D. L., Schaffer, M. M. (1978). “Context theory of classification learning”. *Psychological Review* 85: 207–238.

Meltzoff, A., Moore, M. (1999). “A new foundation for cognitive development in infancy: The birth of the representational infant”. In E. Scholnick, K. Nelson, P. Miller, & S. Gelman (Eds.), *Conceptual development: Piaget’s legacy* (pp. 53–78). Mahwah, NJ: Erlbaum Press.

Mitchell, P., & Lacohée, H. (1991). “Children's early understanding of false belief”. *Cognition*, 39, 107-127.

Moll, H., & Tomasello, M. (2007). “How 14-and 18-month-olds know what others have experienced.” *Developmental Psychology*, 43 (2), 309-317.

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). “Conflicting desire and the child’s theory of mind”. *Cognitive Development*, 10, 467– 482.

Morton, A. (1980). *Frames of Mind*. Oxford: Oxford University Press.

Murphy, G. (2002). *The Big Book of Concepts*, Cambridge, MA: MIT Press.

Murphy, G., & Medin, D. (1985). “The role of theories in conceptual coherence”. *Psychological Review* 92: 289–316.

Nagel, J. (2008). “Knowledge Ascriptions and the Psychological Consequences of Changing Stakes”, *Australasian Journal of Philosophy* 86, 279-294.

Nagel, J. (2010). “Knowledge Ascriptions and the Psychological Consequences of Thinking about Error”. *Philosophical Quarterly* 60:239 (2010), 286-306.

Nagel, J. (2010a). “Epistemic Anxiety and Adaptive Invariantism,” *Philosophical Perspectives* 24, 407-435.

Nagel, J. (2012) “Mindreading in Gettier Cases and Skeptical Pressure Cases”, in *Knowledge Ascription: New Essays*, Jessica Brown and Mikkel Gerken, eds. (Oxford University Press, 2012), 171-191.

Nagel, J. (2012a) “Intuitions and Experiments: A Defence of the Case Method in Epistemology,” *Philosophy and Phenomenological Research* 85:3.

Nagel, J. (2013). “Knowledge as a Mental State,” *Oxford Studies in Epistemology* 4, 275-310.

Nagel, J., San Juan, V. & Mar, R. (2013) “Lay Denial of Knowledge for Justified True Beliefs,” to appear in *Cognition*.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). “Surveying freedom: folk intuitions about free will and moral responsibility”. *Philosophical Psychology* 18: 561-584.

Nichols, S., Stich, S., & Weinberg, J. (2001). “Normativity and Epistemic Intuitions”, *Philosophical Topics* 29 (1–2), 429–460.

O'Neill, D. K. (1996). “Two-year-old children's sensitivity to aparent's knowledge state when making requests.” *Child Development*, 67, 659-677.

O’Neill, D., Astington, J., & Flavell, J. (1992). “Young children’s understanding of the role that sensory experiences play in knowledge acquisition”. *Child Development*, 63, 474-490.

Oppenheimer, D. (2004). “Spontaneous Discounting of Availability in Frequency Judgment Tasks”. *Psychological Science* 15 (2), 100–105.

Osherson, D., Smith, E. (1981). “On the adequacy of prototype theory as a theory of concepts”. *Cognition*, 9:35–58.

Papafragou, A., Li, P., Choi, Y and Han, C. (2007). “Evidentiality in Language and Cognition”. *Cognition* 103: 253–99.

Peacocke, C. (1992). *A Study of Concepts*, Cambridge, MA: MIT Press.

Pearl, J. (2000). *Causality.* New York: Oxford University Press.

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Phillips, M., Young, A., Senior, C., Brammer, M., Andrew, C., Calder, A., Bullmore, E., Perrett, D., Rowland, D., Williams, S., Gray, J., & David, S. (1997). “A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389:495–498.

Piaget, J. (1926). *The language and thought of the child* (M. Warden, Trans.). New York: Harcourt, Brace.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: MIT Press.

Pohl, R., & Hell, W. (1996). "No reduction in Hindsight Bias after Complete Information and repeated Testing". *Organizational Behaviour and Human Decision Processes*, 67(1), 49-58.

Pollock, J. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. Cambridge, MA: MIT Press.

Povinelli, D., Eddy, T. (1996). "What young chimpanzees know about seeing". *Monographs of the Society for Research in Child Development*, 61 (3), 1–152.

Povinelli, D., Vonk, J. (2003). "Chimpanzee minds: Suspiciously human?" *Trends in Cognitive Sciences*, 7 (4), 157–160.

Povinelli, D., Rulf, A., Bierschwale D. (1994). "Absence of knowledge attribution and self-recognition in young chimpanzees (Pantroglodytes)". *Journal of Comparative Psychology*, 108(1), 74–80.

Premack, D., Woodruff, G. (1978). "Does the chimpanzee have a theory of mind?". *Behavioral and Brain Sciences*, 1 (4), 515–526.

Pritchard, D. (2005). *Epistemic Luck*, Oxford: Oxford University Press.

Ramsey, W. (1992). "Prototypes and Conceptual Analysis." *Topoi* 11:59–70.

Rawls, J. (1971). *A Theory of Justice*, 2nd Edition 1999, Cambridge, MA: Harvard University Press.

Rehder, B .(2003). "Categorization as causal reasoning". *Cognitive Science*, 27: 709–748.

Rehder, B. (2003a). "A causal-model theory of conceptual representation and categorization". *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition* 29: 1141–1159.

Rips, L. (1995). "The Current Status of Research on Concept Combination". *Mind & Language* 10: 72–104.

Rosch, E. (1973). “On the internal structure of perceptual and semantic categories”, in *Cognitive Development and the Acquisition of Language*, T. E. Moore (ed.), Academic Press, New York.

Rosch, E. (1975). “Cognitive representation of semantic categories”, *Journal of Experimental Psychology*: General 104,192-233.

Rosch, E. (1978). “Principles of categorization”, in *Cognition and Categorization*, E. Rosch and B. Lloyd (eds.), Lawrence Erlbaum, Hillsdale, New Jersey, pp. 27-48.

Rosch, E., Mervis, C. (1975). “Family resemblances: studies in the internal structure of categories”, *Cognitive Psychology* 8, 382-439.

Rosch, E., Simpson, C., and Miller, R. S. (1976). “Structural bases of typicality effects”, *Journal of Experimental Psychology: Human Perception and Performance* 2, 491-502.

Royzman, E., Cassidy, K., and Baron, J. (2003). “‘I Know, You Know’: Epistemic Egocentrism in Children and Adults”. *Review of General Psychology* 7, 38-65.

Ruffman, T. (1996). “Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference”. *Mind and Language* 11,387–414.

Saxe, R. (2005). “Against simulation: the argument from error.” *Trends in cognitive sciences*, 9(4), 174-179.

Schaffer, J, and Knobe, J. (2010). “Contrastivism Surveyed”. *Noûs*. Online Publication, 15 DEC. DOI: 10.1111/j.1468-0068.2010.00795.x.

Shatz, M., Wellman, H., Silber, S. (1983). “The acquisition of mental verbs: A systematic investigation of the first reference to mental state.” *Cognition*, 14(3), 301-321.

Smith, E. Medin. D. and Rips L. (1984) “A psychological approach to concepts; Comments on Rey's 'Concepts and Stereotypes”, *Cognition* 17, 265-274.

Sodian, B., Thoermer, C., & Dietrich, N. (2006). “Two-to four-year-old children's differentiation of knowing and guessing in a non-verbal task.” *European Journal of Developmental Psychology*, 3(3), 222-237.

Southgate, V., Senju, A., & Csibra, G. (2007). “Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, 18, 587-592

Stanley, J. (2005). *Knowledge and Practical Interests*, New York and Oxford: Oxford University Press.

Stanley, J. (2005). *Knowledge and Practical Interests*, New York and Oxford: Oxford University Press.

Starmans, C., & Friedman, O. (2012). "The folk conception of knowledge". *Cognition*, 124 (3), 272–283.

Stich, S., Weinberg, J. (2001). "Jackson's Empirical Assumptions". *Philosophy and Phenomenological Research*, 62:3, 637-643.

Sosa, E. (1999). "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 141-54. A discussion of safety in the context of skepticism.

Sosa, E. (2008). "Experimental Philosophy & Philosophical Intuition. In *Experimental Philosophy*. Eds. Knobe, J., & Nichols, S. Oxford University Press.

Sosa, E. (2009). "A defense of the use of Intuitions in Philosophy". In *Stich and his Critics*. Eds. Murphy, D. & Bishop, M. Wiley-Blackwell.

Swain, S., Alexander, J., & Weinberg, J. (2008). "The instability of philosophical intuitions: Running hot and cold on true-temp". *Philosophy and Phenomenological Research*, 76(1), 138–155.

Tardif, T., & Wellman, H. (2000). "Acquisition of mental state language in Mandarin-and Cantonese-speaking children". *Developmental Psychology*, 36(1), 25.

Tomasello, M., Call, J. (1997). *Primate cognition*. New York, NY, USA: Oxford University Press.

Turri, J. (2011). "Manifest failure: The Gettier problem solved." *Philosophers' Imprint*, 11, 1–11.

Turri, J. (2013). "A conspicuous art: putting Gettier to the Test". To appear in *Philosophers' imprint*.

Weinberg, J., Nichols, S. and Stich, S. (2001). "Normativity and Epistemic Intuitions." Philosophical Topics, 29, 429-460.

Weiskopf, D. (2009). "The Plurality of Concepts", *Synthese*, 169: 145–173.

Wellman, H., Cross, D., and Watson, J. (2001). "Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, 72(3), 655-684.

Wellman, H., Liu, D. (2004). "Scaling of Theory of Mind Tasks. *Child Development*, 75 (2), 523-541.

Williamson, T. (1995). "Is knowing a state of mind?". *Mind*, 104 (415), 533.

Williamson, T. (2000). *Knowledge and its Limits.* New York: Oxford University Press.

Williamson, T. (2005). “Contextualism, subject-sensitive invariantism and knowledge of knowledge”. *The Philosophical Quarterly* 55, 213-235.

Williamson, T. (2007). *Philosophy without intuitions.* Oxford: Blackwell.

Williamson, T. (2009). “Replies to Critics”. In P. Greenough & D. Pritchard (Eds.), *Williamson on Knowledge* (pp. 279-284). New York: Oxford University Press.

Wimmer H., Perner J. (1983). “Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception”. *Cognition* 13 (1): 103–128.

Wittgenstein, Ludwig (1953/2001). *Philosophical Investigations*. Blackwell Publishing.

Wright, J. C. (2010). “On intuitional stability: The clear, the strong, and the paradigmatic”. *Cognition*, 115(3) ,491–503.

Zaitchik, D. (1991). “Is only seeing really believing? Sources of true belief in the false belief task”. *Cognitive Development*, 6, 91-103.

Zagzebski, L. (1994). ”The inescapability of Gettier problems.” *The Philosophical Quarterly*, 44, 65–73.