

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

LUIS FERNANDO MUNARETTI DA ROSA

**TOWARD EXPLICATING AND MODELLING EPISTEMIC  
RATIONALITY**

Porto Alegre

2014

LUIS FERNANDO MUNARETTI DA ROSA

**TOWARD EXPLICATING AND MODELLING EPISTEMIC  
RATIONALITY**

Tese apresentada como requisito para a obtenção de grau de Doutor pelo Programa de Pós-Graduação em Filosofia da Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Dr. Cláudio de Almeida

Co-orientador: Dr. Peter Klein

Porto Alegre

2014

LUIS FERNANDO MUNARETTI DA ROSA

**TOWARD EXPLICATING AND MODELLING EPISTEMIC  
RATIONALITY**

Tese apresentada como requisito para a obtenção de grau de Doutor pelo Programa de Pós-Graduação em Filosofia da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovada em: \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

**BANCA EXAMINADORA:**

---

Prof. Dr. Claudio Gonçalves de Almeida (Orientador)

---

Prof. Dr. Felipe de Matos Müller

---

Prof. Dr. Peter Klein - Rutgers University / EUA

---

Prof. Dr. Tito Alencar Flores - UNILA

---

Prof. Dr. Rogel Esteves de Oliveira

Porto Alegre

2014

I dedicate this work to any human being who is on the business of trying to understand human rationality.

## ACKNOWLEDGMENTS

First and foremost, I would like to thank CAPES for granting me both, the fellowship to my Ph.D. course at PUCRS and the Sandwich fellowship at Rutgers (as a Visiting Ph.D. Student). I hope I can contribute to the Brazilian academia in order to make their investment worth it.

To my advisor (and partner in philosophical adventures) Cláudio de Almeida and to my co-advisor Peter Klein (Rutgers). It is a big honor for me to have my work scrutinized by these illustrious philosophers and to get their judicious feedback.

To John Turri, Alvin Goldman, Rodrigo Borges, Kurt Sylvan, Min Xu, Durga Laxmi, Justin Sharber, Danilo Dantas, Lucas Roisenberg, Marcelo Fischborn and two anonymous referees for reading and commenting parts of what is written in the present work (presented in various forms). Also to Leonardo Ruivo, Luiz Paulo, Rafael Santin, Katia Martins, Eros de Carvalho and Felipe Muller for their feedback and stimulating ideas in our study group on knowledge-how at *PUCRS*.

To the postgraduate program in Philosophy at *PUCRS* for their excellent working conditions. My best wishes of success to this program in the years to come!

To my family for their support and their warm receptions in the weekends. To my friends and band partners (*Greek van Peixe*) for similar reasons.

Finally, to Stéphane Dias for her tenderness, her patience and for being really nice to me during all this time. Without you, I would not be *that* peaceful to write whatever is written here.

*“Wiggle your big toe”*. (The Bride in *Kill Bill Vol. 1*, by  
Quentin Tarantino)

## RESUMO

Na primeira parte deste trabalho, analisamos o conceito de *racionalidade epistêmica* e oferecemos uma teoria sobre as condições que precisam ser satisfeitas para que uma determinada atitude doxástica seja (*ex ante*) epistemicamente racional. Na segunda parte, consideramos e desenvolvemos um tipo de semântica formal para atribuições de racionalidade.

**Palavras-chave:** Epistemologia; Justificação Epistêmica; Conhecimento Procedural; Racionalidade Epistêmica; Inferência; Razões; Semântica de Modelos.

## ABSTRACT

In the first part of this work, we analyze the concept of *epistemic rationality* and we present a theory about the conditions that need to be satisfied in order for a doxastic attitude to be (*ex ante*) epistemically rational for someone. In the second part, we develop a type of formal semantics for attributions of epistemic rationality.

**Keywords:** Epistemology; Epistemic Justification; Procedural Knowledge; Epistemic Rationality; Inference; Reasons; Model-Theoretic Semantics.



## LIST OF SYMBOLS

- $\alpha, \beta, \gamma$  – Variables for inferential schemata  
 $\varphi, \psi, \chi$  – Variables for propositions  
 $B, D$  – Doxastic operators (*belief* and *doubt* respectively)  
 $R( )$  – Absolute-rationality operator  
 $R( | )$  – Relative-rationality operator  
 $p, q, r$  – Constants for particular propositions  
 $S$  – Variable for subjects  
 $t$  – Variable for particular times

# CONTENTS

<b>Introduction</b> .....	10
 <b>PART 1 – Explicating epistemic rationality</b>	
<b>1. Setting the stage</b> .....	15
1.1 Preliminary remarks and philosophical assumptions .....	15
1.2 <i>Ex ante</i> rationality as a function of reasons (and reasons only) .....	23
<b>2. Knowing how to reason</b> .....	29
2.1 Inference .....	29
2.2 Inferential schemata ( <i>i-schemata</i> ) .....	32
2.3 Optimality .....	37
2.4 Instantiation and availability of <i>i-schemata</i> .....	42
2.5 <i>Knowing how to reason</i> explicated .....	47
<b>3. <i>Ex ante</i> rationality</b> .....	52
3.1 <i>Ex ante</i> rationality explicated (a Rylean move) .....	52
3.2 Similar theories .....	62
3.3 A very brief section on <i>ex post</i> rationality .....	70
 <b>PART 2 – Modelling epistemic rationality</b>	
<b>4. Setting the stage again</b> .....	73
4.1 Language .....	73
4.2 Models – what we need to take into account .....	77
<b>5. Syntax and semantics</b> .....	86
5.1 A formal language .....	86
5.2 A possible-states semantics .....	89
5.2.1 Truth-conditions, consequence and validity .....	95
5.2.2 SM-valid formulas .....	99
5.3 Level of idealization and adequacy of simple models .....	112
5.4 Inconclusive conclusion about our <i>ps-semantics</i> .....	116
 <b>Conclusion</b> .....	 118
<b>References</b> .....	120

# Introduction

Imagine a futuristic device — call it ‘*The Rationalization Helmet*’ — that accurately determines what is epistemically rational for a person to believe, disbelieve or doubt at any given moment. The person wears the *Rationalization Helmet*, turns it on and waits for its output on a little screen. As soon as it is turned on, the helmet somehow starts ‘reading’ the person’s current thoughts and gaining access to her memories and background beliefs. As a result, the program running in the *Rationalization Helmet* builds a set of doxastic attitudes (in some specific formal or semi-formal language) that is supposed to represent the person’s available reasons during a certain time frame (from time  $t_1$  to time  $t_n$ ). Call this process ‘*Step 1*’.

After *Step 1*, the helmet starts a *Q&A* (*Question and Answer*) session with its user by means of a dialogical interface. The questions presented by the helmet are supposed to test the person’s reasoning abilities. The subject wearing the helmet is prompted to answer if something follows from something else, to select the correct answer to a given problem, to evaluate arguments, etc. Using the results gathered in the *Q&A* session, the program running in the *Rationalization Helmet* builds a profile of inferential schemata (in some specific formal or semi-formal language) that is supposed to represent the person’s reasoning abilities in a certain time frame. Call this process ‘*Step 2*’.

As soon as the *The Rationalization Helmet* finishes steps 1 and 2, it builds a *model* of the person’s cognitive state and starts calculating which doxastic attitudes are rational or justified for that person at that time. The helmet’s program will output, then, a set of sentences of the form: ‘It is rational for you to believe  $\phi$ ’, ‘It is rational for you to suspend judgment about  $\psi$ ’, etc. It could also output negative judgments of the form: ‘It is not rational for you to believe  $\phi$ ’, etc. The helmet’s user could, then, request an explanation why believing/suspending judgment about something is/is not rational for her. Accordingly, the program running in our helmet would maintain a track–record of its own derivations in its memory and it would be able to output, upon request, sentences of the form: ‘It is rational for you to suspend judgment about  $\phi$  because nothing you

---

rationally believe gives support to  $\phi$ , ‘It is not rational for you to believe  $\phi$  *because* you rationally believe  $\psi$  and  $\psi$  entails  $\neg\phi$ ’, etc. If the device turns out being accurate, one could advertise it as follows: ‘*Rationalization Helmet — the dream of every epistemologist*’.

There are interesting questions about what a person should do when reading the output in the *Rationalization Helmet*’s screen, specially if the person knows that the helmet is reliable. But we are not going to deal with these questions here. Rather, our interest is in both, the logical machinery that would be used to build the program that runs in our hypothetical helmet and its philosophical foundations. Accordingly, we aim to build a *model-theoretic semantics* that will validate attributions of epistemic rationality to doxastic attitudes for thinking subjects. Each model in our semantics will be a formal representation of the situation of a hypothetical agent in a certain cognitive state. Given the structure of a certain model, it will ‘make’ true or false certain formulas attributing/denying rationality to doxastic attitudes for the subject represented in that model.

It is important to emphasize, however, that we do not intend to determine if a certain model is an accurate representation of the cognitive state of a particular agent. Presumably, that would be a work for cognitive scientists and a necessary step before we start selling our helmet! What *we do intend*, however, is to offer a general semantic framework to ‘make’ attributions of rationality *true* or *false*. As elsewhere, our work is of an abstract, conditional sort: *if* a certain model is an accurate representation of a certain agent, *then* such-and-such doxastic attitude is rational for that agent. So we will not worry about the techniques that our *Rationalization Helmet* could eventually use to build a model of the cognitive state of a certain subject.

Logicians might not feel interested in the project we just described. In general, we want a model-theoretic semantics to be used as a validation structure for a particular *logic*, and we want our logics to have a certain degree of generality with respect to the domain of natural language they are supposed to formalize. In our semantical framework, however, each model is supposed to be a representation of the situation of a *single* (hypothetical) agent. As a result, there would be a ‘logic’ for each agent. But this is not all about the type of semantics that we envision. By means of certain relevant properties shared by different models (representing situations of different agents) we expect to work out the notion of *model families*. Our semantics can be used, then, as a validation structure for what we would call ‘*Rationality Logics*’ — intensional logics with doxastic as well as rationality operators. To each model family there would be a corresponding *Rationality Logic* with a certain degree of generality. As soon as we have our *Rationality Logics*, they can be put to work and we can use them to derive general epistemological consequences.

Our present work is a first step towards developing the relevant semantic framework and

its corresponding logics. But we are not going to develop the relevant *Rationality Logics* right here — we need to cover some philosophical groundwork until we get there! We do not want just to build another logical machinery and think about its applications. Our hypothetical device is supposed to be a *reliable indicator* of what is rational for someone to believe or doubt at any given moment: its epistemological judgments must be accurate. Accordingly, we want our semantic framework to validate the relevant formulas *in the right way*. It is the purpose of the present work, then, to cover the relevant philosophical groundwork. So let us make it clear what we will pursue in our investigation.

First and foremost, we need to determine *what are the crucial elements* that we need to take into account when judging if a certain doxastic attitude is epistemically rational for someone. Is rationality just a function of the reasons available to a person? If not, what else is important? So we are going to deal with necessary and sufficient conditions for a doxastic attitude to be epistemically rational for someone. We expect to flesh out a plausible theory about these conditions in *Part 1 — Explicating Epistemic Rationality*.

We will argue that, in order for a belief to be *rational* or *inferentially justified* for a certain person, she not only needs to have good reasons for holding the relevant belief — she also needs to have a certain kind of *procedural knowledge*. More specifically, being inferentially justified requires *knowing how to perform an inference*. The account of epistemic rationality fleshed out in *Part 1* is motivated by a problem with the view that believing  $\phi$  is rational for  $S$  solely in virtue of the reasons available to  $S$ . Further, we purport to show that our account is better than rival but similar accounts. So much for *Part 1*.

Second, we will consider how to *model* the crucial elements that we need to take into account when judging if a certain doxastic attitude is epistemically rational for someone. This is a task to be pursued in *Part 2 — Modelling Epistemic Rationality*. Our models will be abstract representations of situations that cognitive agents are in, and we will sketch both a formal language and a model-theoretic framework for attributions of epistemic rationality. In *Part 2* we start building our semantics and dealing with a number of problems involved in such a task. This part is supposed to be a transition to a bigger project: that of developing the *Rationality Logics* we mentioned above. As such, it is not a complete and exhaustive work on the semantics of attributions of rationality — just a first attempt to build such a semantics. What we have here is, admittedly, *unfinished business*.

As a whole, then, the present work offers a theory about epistemic rationality (*Part 1*) and it starts fleshing out a semantics for attributions of epistemic rationality (*Part 2*). Philosophers will probably have more interest in *Part 1*, since it is in this part that we offer arguments for the account of epistemic rationality that we defend. More technically—

---

minded researchers (formal semanticists and logicians) will probably have more interest in *Part 2*. But it is fair to say that the really substantial part of this work lies in *Part 1* — there is more to the philosopher than to the logician or formal semanticist here.

\*\*\*

I have benefitted immensely from conversations with various professional philosophers when writing parts of this work. It is hard to imagine good philosophy without good conversation and feedback. Without assuming that what the reader will find here is really good philosophy, I would like to emphasize that the present work would not have its merits (if there are such merits) without the stimulus from the mentioned conversations (I am aware that this is a risky counterfactual).

The present work was developed in two main environments. The first one is the department of philosophy at *PUCRS* (*Pontifícia Universidade Católica do Rio Grande do Sul*). I am happy to note that this place is filled with academically excellent epistemological debates (all due to the wonderful work that our ‘Jedi Master’ Claudio de Almeida has been doing for a long time now). The second one is the department of philosophy at *Rutgers* (*The State University of New Jersey*). When it comes to philosophical activity, this is the most exciting place I have ever seen and I spent a hard-working, fruitful time there.

*PART 1*

*Explicating Epistemic Rationality*

# Chapter 1

## *Setting the stage*

In this chapter we will make some preliminary remarks about the concepts we are going to use and set forth a number of philosophical assumptions (*Section 1.1*). We will also present and criticize the account of epistemic rationality that we aim to substitute (*Section 1.2*).

### 1.1 Preliminary remarks and philosophical assumptions

In this work we are going to talk about *rationality* as a property ascribed to (actual or potential) doxastic attitudes for a certain agent at a certain time. For simplicity, we chose a *binary* typology of doxastic attitudes (beliefs and doubts)<sup>1</sup> instead of a *graded* typology (degrees of belief, or credences)<sup>2</sup>. A standard way of attributing rationality to a belief will be: ‘Believing  $\phi$  is rational for  $S$  at time  $t$ ’. Sometimes we will also use sentences like ‘The belief that  $\phi$  is rational for  $S$  at  $t$ ’ and ‘It is rational for  $S$  to believe that  $\phi$  at  $t$ ’ in order to avoid clumsiness in some particular contexts — but these sentences must be understood as synonymous with the first one, despite their grammatical differences.

The concept of *rationality* used here is supposed to denote what is sometimes called ‘inferential justification’ in the contemporary literature<sup>3</sup>. To say that forming a certain belief is rational for  $S$  comes to the same thing as saying that forming a certain belief is inferentially justified for  $S$ . Rationality or inferential justification is an epistemic status that certain doxastic attitudes have, for a certain person  $S$  at a certain time  $t$ , in virtue of the fact that there is a set of reasons available to  $S$  at  $t$  (maybe *not only* in virtue of those reasons, though, as we will see). The reasons themselves are other doxastic attitudes

---

<sup>1</sup>In the binary typology that we are assuming, a disbelief towards  $\phi$  is just a belief towards  $\neg\phi$  (or not- $\phi$ ).

<sup>2</sup>Christensen (2004) argues that the graded typology is preferable when it comes to making sense of logical constraints on epistemic rationality. The theoretical points we are going to make in this work, however, are independent of the choice regarding which typology is more appropriate.

<sup>3</sup>See Audi (1993).



that are rationally held by the agent<sup>4</sup>. So we will use the concepts of *justification* and *rationality* interchangeably here, but the reader must keep in mind that we are always talking about *inferential* justification, not about *non-inferential* justification<sup>5</sup>.

It is important to notice that from the fact that a certain doxastic attitude is rational for a certain agent at a certain time it does not follow that the agent holds the relevant doxastic attitude at that time. It can be rational for  $S$  to believe  $\phi$  at  $t$  even though  $S$  does not believe  $\phi$  at  $t$ . So we have to make a distinction that is similar to the distinction between *propositional* justification and *doxastic* justification. The relevant distinction is simply the distinction between a doxastic attitude *being rational* for  $S$  at  $t$  and a doxastic attitude *being rationally held* by  $S$  at  $t$ . Call the former property ‘*ex ante* rationality’ (or ‘*ex ante* inferential justification’) and the latter property ‘*ex post* rationality’ (or ‘*ex post* inferential justification’)<sup>6</sup>.

Whenever a belief is rationally held by  $S$ , that belief is rational for her, but not vice-versa (whenever a belief is *ex post* justified, it is *ex ante* justified, but not vice-versa). Our primary investigation is about doxastic attitudes that are rational for an agent — not about doxastic attitudes that are rationally held by an agent. From the fact that our primary focus is on *ex ante* rationality, however, it does not follow that it is not our overall aim to clarify and explicate the concept of *ex post* rationality as well<sup>7</sup>. We think that informative truth-conditions for the proposition expressed by ‘ $S$  rationally believes that  $\phi$  at  $t$ ’ are pretty straightforward after the establishment of informative truth-conditions for the proposition expressed by ‘It is rational for  $S$  to believe that  $\phi$  at  $t$ ’, as we will see in *Chapter 3*.

We have to say something about what is involved in an attribution of rationality to a doxastic attitude for a certain agent. We are talking about *epistemic* rationality, not about *practical* rationality, here. Believing  $\phi$  is taken to be epistemically rational for a certain subject  $S$  only when, on the assumption that the propositions making up the contents of the reasons available to  $S$  are true<sup>8</sup>,  $S$  would maximize the epistemic goal of having true beliefs and not having false ones by forming a belief in  $\phi$ <sup>9</sup>. That is, believing  $\phi$  is

<sup>4</sup>We will talk about the ontology of reasons that is assumed here in a moment.

<sup>5</sup>We do not believe that there is such a thing as *non-inferential* justification, but our conclusions here will not depend on this assumption.

<sup>6</sup>Here we follow Goldman’s (1979) use of the *ex ante/ex post* qualifiers in order to talk about justification. We avoid using the distinction between *propositional* and *doxastic* justification because we will reject the former type of justification as a sufficient condition for what we call ‘*ex ante* rationality’ in the next section (*Section 1.2*).

<sup>7</sup>Here we use the term ‘explication’ in Carnap’s sense (1962). For discussion, see Maher (2007).

<sup>8</sup>This is the assumption that we have to make when the reasons available to  $S$  are constituted by *beliefs*. Sometimes attitudes of *doubt* are also part of one’s reasons, and such cases require a separate treatment. We will get back to this in *Chapter 2*.

<sup>9</sup>Whenever we talk about ‘the’ epistemic goal we have in mind this goal: having true beliefs and avoiding false ones. The goal of believing truths and avoiding falsehoods is supposed to be a *general*

epistemically rational for a certain subject  $S$  only when  $S$  would *conditionally* maximize the epistemic goal by believing  $\phi$  (conditionally on the truth of propositions that are already rationally believed by  $S$ )<sup>10</sup>. There are at least three important observations here.

First, when we assume that the propositions already believed by  $S$  are true (in order to determine if something else would be true or probably true) we are not committed to the claim that reasons are always beliefs in true propositions. Suppose  $S$ 's available reasons at  $t$  are  $R = \{B\psi_1, \dots, B\psi_n\}$ <sup>11</sup>. When we are trying to judge if believing  $\phi$  is epistemically rational for  $S$  at  $t$ , we need to check if  $\phi$  is true or probably true on the assumption that the members of  $\{\psi_1, \dots, \psi_n\}$  are all true. If  $\phi$  is neither true nor probably true conditional on the truth of the members of  $\{\psi_1, \dots, \psi_n\}$ , then  $S$  would not conditionally maximize the epistemic goal by believing  $\phi$ . And if there is no such conditional maximization of the epistemic goal, there is no epistemic rationality (at least in the sense that interests us). That does not mean, however, that in order for a belief in  $\phi$  to be justified for  $S$  in virtue of reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  all the members of  $\{\psi_1, \dots, \psi_n\}$  need to be true.

Second, one could point out that our maximization-of-the-epistemic-goal condition trivializes rationality for beliefs. If the reasons available to  $S$  are beliefs in propositions that together form an *inconsistent* set of propositions, then it is not possible that all propositions making up the contents of  $S$ 's available reasons are true. In this case the conditional (*if* the propositions making up the contents of  $S$ 's available reasons are true, *then*  $S$  would maximize the epistemic goal of having true beliefs and not having false ones by forming a belief in  $\phi$ ) would be true for any  $\phi$  in virtue of the falsity of its antecedent (at least for a classical interpretation of the relevant *if-then* structure as a *material conditional*). But it is a mistake to think that, if that is really the case, then we are trivializing rationality for beliefs: the condition we are stating is just a *necessary* condition for a belief to be rational for someone, not a *sufficient* one<sup>12</sup>. Therefore, we are

---

epistemic goal, but we will leave it open if the value of further epistemic goals (like having knowledge) is derived from the value of this general goal. See Foley's (1993, p. 94) description of the general epistemic goal as 'having an accurate and comprehensive belief system'.

<sup>10</sup>When we say that ' $S$  would *conditionally* maximize the epistemic goal by believing  $\phi$ ' we mean this: that, conditional on the assumption that the propositions making up the contents of  $S$ 's reasons are true,  $S$  would maximize the epistemic goal by believing  $\phi$ .

<sup>11</sup>We will use the operator  $B$  to represent states of *belief* and the operator  $D$  to represent states of *doubt*. So ' $B\phi$ ' stands for a state of belief towards  $\phi$  and ' $D\phi$ ' stands for a state of doubt towards  $\phi$ . This use of the operators  $B$  and  $D$  should not be confused with the use of the same operators with an index ' $S$ ' to attribute beliefs to subjects. ' $B_s\phi$ ' says that  $S$  believes  $\phi$  and ' $D_s\phi$ ' says that  $S$  doubts/suspends judgment about  $\phi$ . The difference can be explained as follows: ' $B\phi$ ' and ' $D\phi$ ' do not have truth-values, because these symbols denote particular states and they do not constitute complete sentences (they behave like proper names), while ' $B_s\phi$ ' and ' $D_s\phi$ ' *do have* truth-values and are complete (assertoric) sentences. We will use the  $B$  and  $D$  operators with indexes to subjects in our semantic framework later (*Chapters 4 and 5*).

<sup>12</sup>Of course, from the fact that the condition of conditional maximization is only necessary for rationality it does not follow that it is not trivially fulfilled in the case of inconsistent reasons.

not committed to attributing rationality to *any* belief for an agent when she has beliefs in inconsistent propositions.

Third, *conditional* maximization of the epistemic goal does not entail, and differs from, *actual* maximization of the epistemic goal. It may be that the content of  $S$ 's reasons gives support to a false proposition. In such a situation, believing the relevant proposition would not *actually* maximize the epistemic goal — but doing so would *conditionally* maximize the epistemic goal. So the idea of conditional maximization of the epistemic goal leaves open the possibility that subjects in skeptical scenarios (such as vatted-brain scenarios and Evil Demon worlds) have perfectly rational beliefs.

Here is a good way to make sense of the claim that being epistemically rational (i.e., inferentially justified) is a matter of conditionally maximizing the epistemic goal: if believing  $\phi$  is rational for  $S$  in virtue of  $S$ 's reasons  $R = \{B\psi_1, \dots, B\psi_n\}$ , then  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ <sup>13</sup>. Likewise, if suspending judgment about  $\phi$  is rational for  $S$  in virtue of  $S$ 's reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  at  $t$ , then  $\{\psi_1, \dots, \psi_n\}$  gives support neither to  $\phi$  nor to  $\neg\phi$ . Either way, the notion of *support* is used to establish a necessary condition for the rationality of a doxastic attitude. But what is it for a set of propositions to maintain a relation of support with a further proposition?

One way to answer this question is to interpret the support relation as a *confirmation relation* —  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  when  $\{\psi_1, \dots, \psi_n\}$  *confirms*  $\phi$ . Further, confirmation relations can be explicated by means of *probability functions*<sup>14</sup>. We will adopt this interpretation. There are, however, at least two types of confirmation: *incremental* confirmation and *absolute* confirmation (in both cases probability functions ( $Pr$ ) are used to determine what confirms what)<sup>15</sup>. Roughly,  $\{\psi_1, \dots, \psi_n\}$  *incrementally* confirms  $\phi$  when and only when  $\{\psi_1 \wedge \dots \wedge \psi_n\}$  raises the prior probability of  $\phi$ , and  $\{\psi_1, \dots, \psi_n\}$  *absolutely* confirms  $\phi$  when and only when  $\phi$  is more probable conditional on  $\{\psi_1 \wedge \dots \wedge \psi_n\}$  than  $\neg\phi$  is. Formally,  $\{\psi_1, \dots, \psi_n\}$  *incrementally* confirms  $\phi$  when and only when  $Pr(\phi | \psi_1 \wedge \dots \wedge \psi_n) > Pr(\phi)$ , for some relevant probability function  $Pr$ , while  $\{\psi_1, \dots, \psi_n\}$  *absolutely* confirms  $\phi$  when and only when  $Pr(\phi | \psi_1 \wedge \dots \wedge \psi_n) > Pr(\neg\phi | \psi_1 \wedge \dots \wedge \psi_n)$ ,

<sup>13</sup>Notice, again, that we are not saying that being epistemically rational is *only* a matter of conditionally maximizing the epistemic goal. We are saying, rather, that being epistemically rational is *also but maybe not only* a matter of conditionally maximizing the epistemic goal.

<sup>14</sup>Probability functions are functions obeying the axioms of the *Probability Calculus*. For a standard axiomatization of the probability calculus, see Earman (1992, p. 36). As far as we know, probabilistic theories of confirmation are our best theoretical choice, in view of the well known *paradoxes of confirmation* (compare Bayesian probabilistic theories of confirmation with, for example, hypothetico-deductivism). For discussion, see Crupi (2013).

<sup>15</sup>See Maher (2005). Carnap (1962, p. xviii) uses the notion of *confirmation as increase in firmness* to talk about incremental confirmation and the notion of *confirmation as firmness* to talk about absolute confirmation.

for some relevant probability function  $Pr$ <sup>16</sup>.

One might ask: which one is the *relevant* probability function here? We are said that both types of confirmation (*incremental* and *absolute* confirmation) are determined by probabilities, but we are not said where these probabilities come from or how they are ultimately determined. There are at least two options here. First, when we are trying to determine if the content of  $S$ 's reasons,  $\{\psi_1, \dots, \psi_n\}$ , gives support to  $\phi$  we can take  $Pr$  to be determined by  $S$ 's *priors* (initial ascriptions of subjective probabilities) over the relevant propositions<sup>17</sup>. In this case, the 'gives support to' relation would always be indexed to a certain doxastic agent. Second, when we are trying to determine if  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  we can take  $Pr$  to be determined by two 'objective' (i.e., not indexed to a particular agent) constraints:  $Pr$  must be such that  $Pr(\psi_1 \wedge \dots \wedge \psi_n) = 1$  and such that it distributes *evenly* over the state-descriptions where  $(\psi_1 \wedge \dots \wedge \psi_n)$  is true (state-descriptions are conjunctions whose conjuncts are either atomic formulas or negations of atomic formulas)<sup>18</sup>. We need not delve into the details of each of these approaches here. Even if we do not decide which one is the *relevant* probability function our merely structural points will suffice for the present investigation.

It is clear that *incremental* confirmation is not sufficient for the type of support relation that interests us. That I bought a single ticket from a fair lottery with a million tickets may plausibly *raise the probability* that I will win the lottery — but it can hardly be said that the former proposition *gives support* to the latter. Even if it is true that I bought a single ticket from the relevant lottery, it is quite unlikely that I will win. So maybe it is the notion of *absolute* confirmation that we need to use in order to explicate the relevant support relation. But there are problems here as well: there are cases where  $Pr(\phi | \psi) > Pr(\neg\phi | \psi)$  but  $\psi$  actually *decreases* the probability of  $\phi$ <sup>19</sup>. This may happen when  $\phi$  is already highly probable, and conditioning it on  $\psi$  do not put its probability over the threshold 0.5.

We can avoid both problems (that *incremental* confirmation alone is not sufficient for the support relation and that *absolute* confirmation alone is not sufficient for the support relation) by making the following requirement:  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  when  $\{\psi_1, \dots, \psi_n\}$  *absolutely* confirms  $\phi$  and it is not the case that  $\{\psi_1, \dots, \psi_n\}$  *incrementally* confirms  $\neg\phi$ . On the one side, we avoid the claim that  $\{\psi_1, \dots, \psi_n\}$  can give support to  $\phi$  even when the probability of  $\phi$  is lower than the probability of  $\neg\phi$  conditional on  $(\psi_1 \wedge \dots \wedge \psi_n)$ . On the other, we avoid the claim that  $\{\psi_1, \dots, \psi_n\}$  can give support to  $\phi$

<sup>16</sup>In standard *Probability Calculus*, the formula  $Pr(\phi | \psi_1 \wedge \dots \wedge \psi_n) > Pr(\neg\phi | \psi_1 \wedge \dots \wedge \psi_n)$  is equivalent to  $Pr(\phi | \psi_1 \wedge \dots \wedge \psi_n) > 0.5$ .

<sup>17</sup>See Talbott (2001).

<sup>18</sup>See Adams (1998, p. 63-64).

<sup>19</sup>For discussion, see *Chapter 4* of Achinstein (2001).

even when  $(\psi_1 \wedge \dots \wedge \psi_n)$  decreases the probability of  $\phi$ . That will be our interpretation about the support relation from now on: we interpret it as *absolute* confirmation *plus* lack of *negative incremental* confirmation. We will reconsider this particular interpretation about the support relation only if any of the epistemological problems we are going to deal with can be solved by abandoning it and finding a better interpretation.

We just explained one important thing that is involved in an attribution of rationality or inferential justification here: the conditional maximization of the epistemic goal. Believing  $\phi$  is rational for  $S$  in virtue of  $S$ 's reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  at  $t$  only if  $S$  would conditionally maximize the epistemic goal by believing  $\phi$ . We made sense of this requirement by means of the notion of support: believing  $\phi$  is rational for  $S$  in virtue of  $S$ 's reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  at  $t$  only if  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ . But there is another important implication involved in such attributions: when we say that having a certain doxastic attitude is rational for someone, we are implicitly attributing a *set of reasons* to that person. If the reasons  $R$  are not possessed by  $S$ , then it cannot be rational for  $S$  to believe  $\phi$  in virtue of reasons  $R$ . We have been assuming that reasons are doxastic attitudes, but one could object that reasons are *propositions* – not doxastic attitudes. Call the former view '*statism*'<sup>20</sup> and the latter '*propositionalism*'<sup>21</sup>. We want to briefly explain why we chose *statism* over *propositionalism*.

We chose *statism* about reasons because reasons are items that necessarily have a certain *modality*. In the typology of doxastic attitudes adopted here, the proposition *Today is Tuesday* can be present in  $S$ 's cognition in two relevant doxastic modes: it can be believed or it can be doubted<sup>22</sup>. When  $S$  (rationally) believes that proposition (that is, when the modality towards that proposition is the *belief* modality),  $S$  *has a reason to believe* that *Tomorrow is Wednesday* (assuming, as we are, that  $S$  knows that *Tuesday is the day before Wednesday* and that  $S$  knows how to use the indexicals 'today' and 'tomorrow'). When  $S$  (rationally) suspends judgment about or doubts that proposition

<sup>20</sup>There are at least two versions of *statism*. The first one only counts doxastic attitudes as evidence or reasons (Davidson (2001), for example, argues that only doxastic attitudes are reasons for beliefs). One can then decide how to classify such doxastic attitudes — by way of *type* (belief, disbelief, suspending judgment) or by way of *degrees* (credences, real-valued functions mapping to the unit interval). The second version of *statism* counts as evidence not only doxastic attitudes, but also experiences. One can then decide what to count as *experience*: perceptual experiences, mnemonic experiences, introspective experiences and intuitions are candidates. The details need not bother us at this moment. For the sake of simplicity we will mostly talk of evidence or reasons, from the point of view of *statism*, as consisting of doxastic attitudes, since in both versions of the view doxastic attitudes are included in the class of things that count as evidence. Some *statist* accounts of evidence include those in Lewis (1996), Conee and Feldman (2008) and Alston (2005). For a defense of *statism* over *propositionalism*, see Turri (2009).

<sup>21</sup>We take it that *factualism* about evidence (the view that evidence consists of facts) is a type of *propositionalism* — one that says that evidence consists of *true* propositions. Some *propositionalist* accounts include those in Williamson (2002), Dougherty (2011) and Neta (2008).

<sup>22</sup>As we pointed out before, disbelieving a proposition is just believing its negation. So disbelief is not a further doxastic modality.

(that is, when the modality towards that proposition is the *doubt* modality), *S* has a reason to suspend judgment about whether *Tomorrow is Wednesday*. Using ‘*p*’ to represent the proposition that *Today is Tuesday* and ‘*q*’ to represent the proposition that *Tomorrow is Wednesday*, we can say that *S*’s belief that *p* (*Bp*) is a reason for *S* to believe *q*. We can also say that *S*’s doubt towards *p* (*Dp*) is a reason for *S* to doubt *q*.

If we drop these doxastic modalities to talk about reasons a problem emerges<sup>23</sup>. Both *Bp* and *Dp* have the same propositional content (*p*), but the reasons that one has when one is in a state that includes the former doxastic attitude are clearly different from the reasons that one has when one is in a state that includes the latter doxastic attitude<sup>24</sup>. How can a propositionalist theory of reasons take into account the fact that *S* has a reason to doubt that *Tomorrow is Wednesday* when *S* doubts that *Today is Tuesday*? Can she defend the idea that there is a *proposition* that constitutes *S*’s reason here? Well, it certainly cannot be the case that the relevant proposition is *p* itself: that *Today is Tuesday* cannot be a reason for *S* to doubt that *Tomorrow is Wednesday*!

One way the propositionalist could go would be to claim that the reason *S* has to suspend judgment about whether *Tomorrow is Wednesday* is the proposition *I (S) do not know if today is Tuesday*. In general, reasons to suspend judgment are propositions about what one does not know or about one’s uncertainties, etc. But it is simply not true that whenever we have reasons to suspend judgment about something we have a second-order belief in a proposition like that — we can have reasons to doubt things without necessarily going through any higher-order thought (any appearance to the contrary may be due to the fact that states of doubt usually stimulate us to reflection). Further, *when we use* our reasons to suspend judgment about something (and we do so rationally) we do not necessarily use propositions about what we do not know. *S* is in a state of doubt about if *Today is Tuesday*. She considers the proposition *Tomorrow is Wednesday* and suspends judgment about it on the basis of her state of doubt towards the proposition *Today is Tuesday*. *S* clearly does not need to perform any higher-order inference when she uses her reasons to (rationally) suspend judgment about if *Tomorrow is Wednesday*<sup>25</sup>.

Maybe the propositionalist could claim that there is no such a thing as a reason for suspending judgment in the type of case we are dealing with. What happens is just that *S* lacks a reason to believe that *Tomorrow is Wednesday*. But this cannot be right either. To be sure, it is true that *S* lacks such a reason, but it is also true that it is possible for *S*

<sup>23</sup>Notice that reasons must always be *believed* according to propositionalism (or *known* in some versions), although the reasons themselves are not the relevant states of belief.

<sup>24</sup>John Turri (2009) makes the same point that we are making against *propositionalism* here (where we use ‘*propositionalism*’ Turri uses ‘*abstractionism*’), and calls this problem ‘The Problem of Withholding’.

<sup>25</sup>This is not to say that *S* does not perform *any* kind of inference when she uses her reasons to (rationally) suspend judgment about if *Tomorrow is Wednesday*.

to suspend judgment about if *Tomorrow is Wednesday* on the basis of her state of doubt towards the proposition *Today is Tuesday* (plus the relevant knowledge he has about the days of the week) in a rational way. That *S* may do so in a rational way means that *that* on the basis of which she suspends judgment is *a reason for* her to do so.

None of this is a problem for *statism*, though. We said above that, according to *statism*, reasons always have a certain modality. So two reasons with different modalities towards the same proposition have different epistemic properties, in such a way that we do not need to ‘put more propositions in *S*’s head’ to explain what is made rational for her when she doubts a certain proposition. Any ontology of epistemic reasons must make sense of the claim that sometimes one’s *reasons for* suspending judgment about a certain proposition are further states of doubt one has towards other propositions. Also, any ontology of reasons must make sense of the claim that sometimes the *reason why* one rationally suspends judgment about a certain proposition is a further state of doubt one has towards another proposition (*S* doubts *q* because *S* doubts *p*). *Statism* clearly makes sense of these facts. In conclusion, we chose statism as our ontology of epistemic reasons.

Let us sum up our basic assumptions. We use the concepts of *rationality* and *inferential justification* interchangeably. The typology of doxastic attitudes of our choice is a *binary* typology. Our primary focus is on *ex ante* rationality, but we also aim to clarify the concept of *ex post* rationality. A doxastic attitude is taken to be epistemically rational for *S* in virtue of the fact that *S* would conditionally maximize the epistemic goal (believing truths, not believing falsehoods) by forming that doxastic attitude (i.e. if, conditional on the assumption that the propositions making up the content of *S*’s reasons are true, *S* would maximize the epistemic goal). That means that believing  $\phi$  is rational for *S* only if the propositional content of *S*’s reasons gives support to  $\phi$ . We interpret the support relation as *absolute* confirmation plus lack of *negative incremental* confirmation. Finally, reasons are taken to be doxastic attitudes or doxastic states.

Are there objections to our basic assumptions? There are, for sure, but we cannot address all of them here<sup>26</sup>. If it turns out that giving up any of our basic assumptions is a plausible way of solving any of the problems we are going to deal with, we will get back to these assumptions and question them.

---

<sup>26</sup>For example, one might claim that *statism* is at odds with some common linguistic practices. More specifically, there are contexts where we talk about *reasons* as if they were *propositions* or *facts*. We already offered some arguments for *statism* about reasons and we think those arguments are good ones, so we will not try to address this objection. We are pretty sure, however, that the statist can offer a plausible *error theory* to deal with sentences that talk about reasons as if they were propositions or facts in ordinary, scientific and legal contexts.

## 1.2 *Ex ante* rationality as a function of reasons

When we say that believing  $\phi$  is rational for  $S$  we are saying that it is rational for  $S$  to believe  $\phi$ . That is, we are attributing the property of *ex ante* rationality to a certain belief for a certain agent<sup>27</sup>. *Ex ante* rationality is a property attributed to doxastic attitudes for a certain agent in virtue of the *reasons* available to that agent. As we saw in our preliminary remarks, when we say that believing  $\phi$  is rational or inferentially justified for  $S$  we are also saying that the content of  $S$ 's reasons gives support to  $\phi$ . Of course, we say *more* about  $S$ 's reasons than that. Presumably, we are also saying that the epistemic status that the relevant reasons available to  $S$  confer to a certain belief is not counterbalanced by whatever further reasons that are also available to  $S$ .

Given that much, one could advance an analysis of *ex ante* rationality purely in terms of available reasons. The idea is that nothing other than available reasons needs to be taken into account in order for us to make accurate judgments about *ex ante* rationality. Call this the '*ex ante* rationality as propositional justification' account. It can be formulated more precisely as follows:

(PJ) Believing  $\phi$  is rational or inferentially justified for  $S$  at  $t$  if and only if there is a set of undefeated reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  available to  $S$  at  $t$  such that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ .

In other words (PJ) says that believing  $\phi$  is rational for  $S$  at  $t$  when and only when there is a set of undefeated reasons  $R$  available to  $S$  at  $t$  such that the propositional content of the members of  $R$  gives support to  $\phi$ . The proviso 'undefeated' in 'undefeated reasons' is part of (PJ) because the reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  may be available to  $S$  while further reasons  $R' = \{B\chi_1, \dots, B\chi_n\}$  are available to  $S$  such that  $\{\chi_1, \dots, \chi_n\}$  gives support to  $\neg\phi$  or such that  $\{\psi_1, \dots, \psi_n, \chi_1, \dots, \chi_n\}$  gives support neither to  $\phi$  nor to  $\neg\phi$  (that is,  $\{\psi_1, \dots, \psi_n, \chi_1, \dots, \chi_n\}$  is *neutral* with respect to  $\phi$ ). If both situations are avoided, we say that  $S$ 's reasons  $R$  for believing  $\phi$  are 'undefeated'. (One complication could be added: if we admit of something like *degrees of justification*, we would need to express the necessary quantitative or comparative relations between reasons *pro* and *con* believing  $\phi$ , but let us put this complication aside for now).

There are at least two crucial questions about (PJ): What it is for a set of reasons  $R$  to be available to  $S$  at  $t$ ? What it is for a set of propositions to give support to a further proposition?<sup>28</sup> When it comes to the latter, we are assuming the interpretation

<sup>27</sup>A more precise way of talking, perhaps, would be to say that *ex ante* rationality is a property attributed to the *formation of a doxastic attitude*. To be sure, we attribute *ex ante* rationality to doxastic attitudes that have not been formed yet.

<sup>28</sup>(PJ) looks like an evidentialist theory, of the type defended by Conee and Feldman (1985. p. 83):



about the support relation that we fleshed out in the previous section —  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  when  $\{\psi_1, \dots, \psi_n\}$  *absolutely* confirms  $\phi$  and it is not the case that  $\{\psi_1, \dots, \psi_n\}$  *incrementally* confirms  $\neg\phi$  (of course, this is not a ‘complete’ account of the support relation, since we only established necessary conditions for this relation to hold)<sup>29</sup>. An alternative interpretation (one that we will reject) will be considered in *Chapter 3*.

When it comes to the former question, let us assume that both, the beliefs whose contents  $S$  is consciously entertaining at  $t$  and the ones whose contents  $S$  can ‘easily access’ at  $t$  count as available reasons to  $S$  at  $t$ . The latter class of beliefs should include beliefs that are retrievable via memory without too much effort on the part of  $S$ . It should also include some beliefs such that, although their content is not consciously entertained by  $S$  at  $t$ , they still play a functional role in  $S$ ’s cognition at  $t$  (for example, by being part of the reasons why  $S$  holds further beliefs). We will not explore alternative definitions of the relevant notion of *easy accessibility* and try to decide which one is better here (this is not within the scope of the present work). That much will suffice for our present investigation<sup>30</sup>.

(PJ) purports to establish a necessary and sufficient condition for a belief to be rational or inferentially justified for any subject  $S$ . It takes into account the consideration that *ex ante* rationality is a function of available reasons. More than that, however, it says that *ex ante* rationality is a function of available reasons *only*.

When  $S$  satisfies the right-side of (PJ) with respect to a certain (actual or otherwise) state of belief that  $\phi$  (that is, when  $S$  has undefeated reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  such that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ ), we will say that  $S$  is *propositionally justified* in believing  $\phi$  (where justification is, remember, inferential justification). Now we want to suggest that in order for a belief to be rational or inferentially justified for  $S$  it is

---

‘Doxastic attitude  $D$  toward proposition  $p$  is epistemically justified for  $S$  at  $t$  if and only if having  $D$  toward  $p$  fits the evidence  $S$  has at  $t$ ’. Accordingly, evidentialism about justification in general faces challenges that are similar to the ones listed here. There are at least two differences, however, between (PJ) and Conee and Feldman’s evidentialism. First (PJ) is about *inferential* justification only. Second, where Conee and Feldman’s thesis talks about a *fitting relation* that holds between the relevant doxastic attitudes, (PJ) talks about a *support relation* that holds between the propositional contents of the relevant doxastic attitudes — and these relations may not be co-extensive.

<sup>29</sup>We gave reasons for choosing that interpretation *among the ones that use the concept of confirmation* to explicate the support relation. We gave no reasons, however, for accepting the idea that the support relation is properly explicated by the concept of *confirmation* (which is, in turn, explicated by probability functions). We just rely on this time-honored idea.

<sup>30</sup>Peter Klein (1981, p. 46) includes more types of beliefs than the ones we mentioned above in the set of *available* reasons to a subject  $S$  (although Klein talks about available *propositions* because, presumably, he is a propositionalist about evidence or reasons). Some beliefs whose content gets support from what is already subscribed to by  $S$  would also count as available reasons, that is, some beliefs that are not actually held by  $S$  but are inferentially justified for  $S$  can be said to be available to  $S$ . Maybe that is true, but we are not committed to the idea that the types of beliefs mentioned above are the *only ones* that can count as available reasons.

not sufficient that  $S$  is propositionally justified in believing  $\phi$ . There are cases where  $S$  is propositionally justified in believing  $\phi$  but, still, believing  $\phi$  is not rational for  $S$ : something else is required. It turns out, then, that (PJ) is false — or so we aim to show in what follows.

Let us introduce a function  $\lceil \cdot \rceil$  that takes a set of *beliefs* as input and returns the *propositional contents* of those beliefs as output. So when  $R = \{B\psi_1, \dots, B\psi_n\}$  it follows that  $\lceil R \rceil = \{\psi_1, \dots, \psi_n\}$ . Now we can make our point by noticing that for each set of reasons  $R$  available to any subject  $S$  there are *infinite* propositions that gain support from  $\lceil R \rceil$ . Some of these propositions are logical consequences of  $\lceil R \rceil$ <sup>31</sup> and some of them are propositions that receive non-deductive support from  $\lceil R \rceil$  (inductive, abductive or probabilistic support). It is pretty obvious that there is a subset of the propositions that gain support from  $\lceil R \rceil$  such that  $S$  does not infer them to be true (or probably true) on the basis of her reasons  $R$ <sup>32</sup>. More than that, there is a subset of the propositions that gain support from  $\lceil R \rceil$  such that  $S$  is *not even able* to infer them to be true (or probably true) on the basis of her reasons  $R$ . That may be so because  $S$  is not able to notice that these propositions receive support from  $\lceil R \rceil$ , or because the inferential path one would need to go through in order to infer them to be true on the basis of  $R$  is too complex, or because these propositions themselves are highly complex, etc. Call the beliefs that  $S$  is not able to competently form on the basis of her reasons at  $t$  the ‘*unreachable*’ beliefs to  $S$  at  $t$ .

It is intuitive to think, then, that it is not rational for  $S$  (at  $t$ ) to form those beliefs that are unreachable to her at  $t$ <sup>33</sup>. If our intuition is accurate, then the following thesis is true<sup>34</sup>:

(Ab) Believing  $\phi$  is rational for  $S$  at  $t$  in virtue of her reasons  $R$  only if  $S$  is able to infer that  $\phi$  on the basis of her reasons  $R$ .

---

<sup>31</sup>The fact that  $\lceil R \rceil$  has logical consequences is what entitles us to say that there are infinite propositions that gain support from  $\lceil R \rceil$ .

<sup>32</sup>One might point out here that we should write ‘ $S$  does not infer them to be true (or probably true) on the basis of  $\lceil R \rceil$ ’. But it is hard to make sense of the claim that  $S$  infers/fails to infer that a certain proposition is true on the basis of further propositions. If  $S$  infers that a certain proposition is true (that is, if  $S$  forms an inferential belief towards a certain proposition), then  $S$  does so on the basis of other intensional attitudes she has towards further propositions. Reasons, as we have argued in our preliminary remarks, always have a certain modality. Here is another way to put it: while propositions are the *relata* in arguments, intensional attitudes are the *relata* in inferences. We are talking about inference here — not about argument.

<sup>33</sup>Similar points have been made elsewhere in the literature. See Goldman (1986, p. 84) and Harman (1986, p. 17).

<sup>34</sup>Goldman (1979) advances a similar thesis when stating what would be a reliabilist analysis of *ex ante* justification. Turri (2010) advances a similar thesis as well when stating his theory about the relationship between propositional and doxastic justification. Although these theses entail the one stated above, they are *stronger* than it. We will get back to Goldman’s and Turri’s theses later (*Chapter 3*) and explain how they differ from our own.

Of course, we need to explain what it is for someone *to be able to* perform a certain inference, in such a way as to make sense of our intuitive judgment about the epistemic status of unreachable beliefs. To be able to perform an inference is to have a certain cognitive ability: the ability to reason in such-and-such a way<sup>35</sup>. More informatively, *to be able to perform an inference is to know how to perform an inference*. So being able to perform an inference is not just a matter of having a way (any way) of coming to form the inferential belief on the basis of the pre-inferential beliefs. When we say that one is not able to perform a certain inference we are saying that one is not able to *competently* form the inferential belief on the basis of the pre-inferential ones.

We will try to explicate the relevant type of procedural knowledge (knowing how to reason) in a more precise way in *Chapter 2*. Before we get there, however, let us present some examples that seem to give support to (Ab) and try to explain why this is so. This will make our case stronger, for what we have so far is an intuition that it is not rational for someone to form those unreachable beliefs. We want to show that the relevant intuition is not mistaken.

Consider the following cases:

**Nocond's case:**

Nocond rationally believes that  $p$  at  $t$ . As one can check through basic propositional logic,  $p$  entails  $((q \rightarrow \neg p) \rightarrow \neg q)$ <sup>36</sup>. Unfortunately, Nocond is not able to infer that  $((q \rightarrow \neg p) \rightarrow \neg q)$  from his belief that  $p$  — he does not know how to perform this kind of inference. Further, Nocond has no reasons to disbelieve or doubt that  $((q \rightarrow \neg p) \rightarrow \neg q)$ , and he has no other reasons for believing that proposition (such as the testimony from someone else).

**Noind's case:**

Noind rationally believes that *only 0.01% of the Fs that are Hs are also Gs* and that *a is an F and an H*. The set of propositions  $\{\textit{only 0.01\% of the Fs that are Hs are also Gs, a is an F and an H}\}$  gives inductive support to *a is not G*. However, Noind is not able to infer that *a is not G* from his beliefs that *only 0.01% of the Fs that are Hs are also Gs* and that *a is an F and an H*. He does not know how to perform this kind of inference. Further, Noind has no reasons to disbelieve or doubt that *a is not G*, and he has no further reasons for believing that proposition.

---

<sup>35</sup>One should not confuse it being merely *possible* for one to perform an inference with one being able to do so. As Maier (2010, *Section 2.1*) puts it when advancing his extensional constraints on a theory of abilities: 'there are many actions that it is metaphysically possible for someone to perform that he lacks the ability to perform'. The same applies to logical and nomological possibility.

<sup>36</sup>' $p$ ' and ' $q$ ' are used here to name particular propositions. The arrow ' $\rightarrow$ ' expresses the material conditional.

These cases are structurally similar. Both cases involve subjects with a set of undefeated reasons  $R = \{B\psi_1, \dots, B\psi_n\}$ , where  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ , such that they do not know how to infer that  $\phi$  from  $R$ . Now let us ask: Is it rational for Nocond to believe that  $((q \rightarrow \neg p) \rightarrow \neg q)$  at  $t$  and for Noind to believe that  $a$  is not  $G$  at  $t$ ?

Consider answering ‘yes’ to our question. That would imply that there are situations where it is rational for  $S$  to believe  $\phi$  (given  $S$ ’s reasons  $R$ ) at  $t$  but  $S$  is not able to infer that  $\phi$  from  $R$  at  $t$ . Such an answer appears to have unwelcome consequences. In order to see why, let us try to figure out what would happen if Nocond were to believe that  $((q \rightarrow \neg p) \rightarrow \neg q)$  given his cognitive state at time  $t$ . Since Nocond has no reasons for believing that  $((q \rightarrow \neg p) \rightarrow \neg q)$  other than his belief that  $p$  at  $t$  and since he does not know how to infer that  $((q \rightarrow \neg p) \rightarrow \neg q)$  from his belief that  $p$  at  $t$ , it follows that *if* Nocond were to believe that proposition at  $t$  he would do so in an epistemically non–approvable manner (for example, by guessing) — he would be *accidentally* forming a belief on the basis of good reasons<sup>37</sup>.

That is, Nocond would not believe  $((q \rightarrow \neg p) \rightarrow \neg q)$  as a result of exercising his ability to perform a certain epistemically approvable type of inference<sup>38</sup> — he cannot exercise an ability he does not have. He would be luckily getting things right, in the sense that he would believe what in fact gets support from his reasons. Of course, Nocond could learn how to perform the relevant inference later — but what matters here is our assumption that *at  $t$*  Nocond does not know how to perform such an inference.

Given these counterfactual considerations, we are entitled to say that, at  $t$ , Nocond *is not in a position* to infer in the right way that  $((q \rightarrow \neg p) \rightarrow \neg q)$  is true from the reasons available to him (similar points apply to Noind’s case). But being inferentially justified in believing  $\phi$  in virtue of reasons  $R$  is, among other things, *being in a position* to infer that  $\phi$  from  $R$  in the right way<sup>39</sup>. If the concept of *rationality* or *inferential justification* does not have that implication, we are not sure *why* one would use it in paradigmatic cases of rational belief in the first place. What would be the point of saying that it is epistemically rational or justified for  $S$  to form a certain belief on the basis of certain reasons without the implication that  $S$  *is in a position* to form the relevant inferential belief in the right way? We have no idea. It would seem that when one says: ‘It is justified for  $S$  to believe

<sup>37</sup>By ‘ $S$  *accidentally* forms a belief on the basis of good reasons’ we mean that it is not the case that  $S$  forms a belief on the basis of good reasons *because*  $S$  knows how to perform an epistemically approvable inference.

<sup>38</sup>An epistemically approvable type of inference is what we will call an ‘*optimal inferential schema*’ later. Roughly, an inferential schema is (epistemically) *optimal* when it conditionally maximizes the epistemic goal of believing truths and not believing falsehoods. The distinction between *optimal* and *non–optimal* inferential schemata will be made in the next chapter.

<sup>39</sup>That does not entail, however, that being inferentially justified in believing  $\phi$  in virtue of reasons  $R$  entails forming a *doxastically justified* belief in  $\phi$  on the basis of  $R$  in a counterfactual situation. We will get back to this in *Chapter 3*.

$\phi$  in virtue of  $S$ 's reasons  $R$ , but  $S$  is not in a position to form a belief in  $\phi$  on the basis of  $R$  in the right way', one gives with one hand what one takes away with the other.

Nocond and Noind, however, are in no such position: they are not able to (or they do not know how to) perform the relevant inferences<sup>40</sup>. And since these subjects are not in a position to form the relevant inferential beliefs in the right way, it is not rational for them to form those beliefs. Both Nocond and Noind, however, satisfy the right-side of the biconditional in (PJ): they both have *propositional justification* for the relevant beliefs. It turns out, then, that (PJ) is false<sup>41</sup>. Further, given what we just fleshed out about the concept of *rationality* or *inferential justification*, our intuition about the epistemological status of unreachable beliefs appears to be right.

We want to fix (PJ) and find a suitable explication of the concept of *ex ante* rationality. Since we will use the notion of *knowing how to perform an inference* in our proposal (see *Chapter 3*), we will also try to explicate what it is for someone to know how to perform a certain inference in the next chapter (*Chapter 2*). Of course, we could just use the results in the present chapter to offer a substitute for (PJ) and leave the notion of knowing how to perform an inference unexplicated. But our developments in the next chapter will serve both to distinguish our own theory about epistemic rationality from similar theories and to present an important feature of our model-theoretic semantics: *inferential schemata* (or *i-schemata*).

---

<sup>40</sup>We are assuming a tight relation between the relevant type of procedural knowledge (knowing how to perform an inference) and a certain kind of *ability* here — not only a certain kind of *competence*. In some sense, at  $t$  Nocond may have the competence to perform the inference described: he can learn how to perform it (he has the competence to do this), or it is not altogether impossible for him to perform it, etc. It seems that in neither case, though, Nocond knows how to perform the relevant inference at  $t$ . If we think in terms of Ernest Sosa's (2010) triad of competence (involving *constitution*, *condition* and *situation*), maybe we could say that we are requiring rational agents not only to have the appropriate *constitution* of a reasoner, but also to be in a certain *condition*. Also, notice that one does not have the cognitive ability to perform an inference when one has only *non-approvable* ways (epistemically speaking) of forming the inferential belief on the basis of the pre-inferential ones.

<sup>41</sup>We are not making the trivial point that (PJ) is false *as an analysis of doxastic justification*. We are not even talking about *doxastic justification* (or *ex post* rationality) at this point. We are saying, rather, that (PJ) is false as an analysis or explication of what it is for  $S$  to have *justification* to believe something, or of what it is for a belief to be *ex ante* rational for  $S$ . We thank Claudio de Almeida and Rogel de Oliveira for pressing on this point.

# Chapter 2

## *Knowing how to reason*

In this chapter we will try to explicate what it is for someone *to know how to perform an inference*. In order to do so, we will first explicate what it is for an inferential schema (or *i-schema*) to be available to someone at a certain time *t*. These notions will be used to offer a substitute for (PJ) in the next chapter, and they are part of our theory about *ex ante* rationality. We also aim to explicate what it is for someone to instantiate an inferential schema.

### 2.1 Inference

What is an inference? While that may look like a philosophical question, answering the (different) question ‘*How do we perform inferences?*’ is a task for cognitive psychologists. There are at least three main psychological theories about how we reason<sup>42</sup>. According to one such theory, the *Mental Logic* theory, subjects reason by applying abstract, general-purpose, rules of inference to received sentential inputs<sup>43</sup>. A rival theory, the *Mental Models* theory, says that subjects reason by building mental models (roughly, image-like representations) of sentential inputs and manipulating them in order to draw conclusions<sup>44</sup>. Finally, the *Availability* theory says that humans reason in a more case-by-case basis, using a variety of *heuristic procedures*<sup>45</sup>.

The details need not bother us here. The important thing to notice is that the arguments *pro* and *con* these theories are mostly based on data gathered from problem-solving experiments and *Q&A* tests with humans, and it is not the purpose of the present work to assess the relevant empirical data and to determine which one of these psychological theories is better supported by it<sup>46</sup>. We intend to stay as neutral as possible when it comes

---

<sup>42</sup>See *Chapter 5* of Hanna (2006).

<sup>43</sup>This kind of approach can be found in Rips’ theory (1994) about deductive reasoning.

<sup>44</sup>See Johnson-Laird (2013).

<sup>45</sup>See Tversky and Kahneman’s (1973) theory about probabilistic reasoning.

<sup>46</sup>For discussion, see Cohen (1981).

to different psychological theories about how we reason. But we have to make at least some substantial assumptions about inferential processes. Let us briefly consider these assumptions.

We will talk about inference as some kind of *state–transition*. For example, we may say that ‘*S* reasons *from* a state that includes *R* to a state that includes *Bφ*’ or, shortly, that ‘*S* reasons *from R to Bφ*’. By this ‘*from...to...*’ structure we are not implying that *S* actually follows rules of inference when *S* reasons<sup>47</sup>. In particular, one should not assume that we are taking the *Mental Logic* theory about reasoning for granted here. It may be that agents build models in order to reason, or that they use a number of heuristic procedures to reason, etc. In any case, there is a transition between mental states, and the only additional information that we are assuming the state–transition attributions to express is that the doxastic attitudes arrived at in such transitions are formed *on the basis of* the initial doxastic attitudes.

Not all reasoned state–transitions consist in adding beliefs to one’s cognitive state, though. Reasoners can also doubt things by means of some sort of inferential process, for example, by performing a state–transition from a state that includes an attitude of doubt *Dφ* to a state that includes a further attitude of doubt *Dψ*<sup>48</sup>. So, in general, here is what is involved in our state–transition attributions: the attribution of *pre–inferential* doxastic attitudes (the reasons - sets of beliefs or doubts) to the reasoner, the attribution of *inferential* doxastic attitudes (beliefs or doubts) to the reasoner and the assumption that the inferential doxastic attitudes are *based on* the pre–inferential ones.

But what is it for an inferential belief *to be based on* pre–inferential beliefs in a reasoned state–transition? Much in the recent literature has been discussed about the so–called ‘*basing relation*’<sup>49</sup>. A great number of epistemologists take the basing relation to be a *causal relation*<sup>50</sup>, but some do not<sup>51</sup>. Briefly, according to the causal interpretation, if a belief *Bφ* is based on reasons *R* it is caused in an appropriate way by reasons *R*. We do not intend to take a stand here, but that much seems to be uncontroversial: when inferential beliefs are based on pre–inferential beliefs the occurrence/presence of the latter ones *explains* (or *is part of the explanation of*) the occurrence/presence of the former ones. When *S* infers

---

<sup>47</sup>That *may* be the case (that *S* follows rules of inference when *S* reasons), but this is not implied by our state–transition talk.

<sup>48</sup>So we also understand inference as ‘reasoned change in view’ as Harman (1986) and Boghossian (2012) do. It might appear to some that it is wrong to call the relevant type of state–transition (from doubt to doubt) as an ‘inference’, presumably because inference would always be carried on under an *assertive mode*, so to say. We are just conceiving the concept of *inference* in a broader sense, however, in order to avoid coining a new term for processes underlying state–transitions from states of doubt to states of doubt. We thank Claudio de Almeida for this observation.

<sup>49</sup>See Korcz (2002).

<sup>50</sup>See Goldman (1979), Audi (1993), Pollock and Cruz (1999).

<sup>51</sup>See Lehrer (1971), Tolliver (1982) and Kvanvig (2003).

that  $\phi$  from  $\{B\psi_1, \dots, B\psi_n\}$ ,  $S$  believes  $\phi$  *because*  $S$  believes each of  $\{\psi_1, \dots, \psi_n\}$  — that is, the *reason why*  $S$  believes  $\phi$  is that  $S$  believes each of  $\{\psi_1, \dots, \psi_n\}$ <sup>52</sup>. Although this much does not give us an account of the *basing relation*, it is sufficient for our present purposes.

We have been using the verb ‘to infer’ with a ‘that’ clause in *Chapter 1*. Examples of sentences with this feature are ‘ $S$  is able to infer *that*  $\phi$  from  $R$ ’, ‘ $S$  infers *that*  $\phi$  from  $R$ ’, etc. Notice that the first sentence is synonymous to ‘ $S$  is able to infer *that*  $\phi$  *is true* from  $R$ ’ and the second one is synonymous to ‘ $S$  infers *that*  $\phi$  *is true* from  $R$ ’. Accordingly, the truth of the first sentence entails that  $S$  is able to form an inferential belief in  $\phi$  on the basis of  $R$ , and the truth of the second one entails that  $S$  actually believes that  $\phi$  on the basis of  $R$ . That  $S$  *infers that* something is true entails that  $S$  *believes that* something is true (to *infer that...* is to *inferentially believe that...*)<sup>53</sup>.

But there is a further use of the concept of *inference* where no doxastic attitude formation towards what is said to be inferred is entailed, as when one says ‘ $S$  infers  $\phi$  from  $\{\psi_1, \dots, \psi_n\}$ ’. The latter sentence can be used as a synonymous to ‘ $S$  performs a derivation from  $\{\psi_1, \dots, \psi_n\}$  to  $\phi$ ’, ‘ $S$  does a proof from  $\{\psi_1, \dots, \psi_n\}$  to  $\phi$ ’, ‘ $S$  builds an argument whose conclusion is  $\phi$  and whose premises are  $\{\psi_1, \dots, \psi_n\}$ ’, etc. So, the truth-conditions for ‘ $S$  infers that  $\phi$  from  $\{B\psi_1, \dots, B\psi_n\}$ ’ and ‘ $S$  infers  $\phi$  from  $\{\psi_1, \dots, \psi_n\}$ ’ are different. While the truth of the former entails that  $S$  *believes that*  $\phi$  on the basis of  $S$ ’s beliefs  $\{B\psi_1, \dots, B\psi_n\}$ , the truth of the latter does not (in fact, the latter says nothing about  $S$ ’s attitude towards  $\phi$ )<sup>54</sup>.

We will not use the concept of *inference* in the latter sense. Instead of saying ‘ $S$  infers  $\phi$  from  $\{\psi_1, \dots, \psi_n\}$ ’, where no specific doxastic attitude towards  $\phi$  on the part of  $S$  is implied, we will say ‘ $S$  performs a derivation from  $\{\psi_1, \dots, \psi_n\}$  to  $\phi$ ’ or ‘ $S$  builds an

---

<sup>52</sup>We are not assuming, again, that the relevant explanation is a *causal* explanation (see Mayes (2001) for different theories about explanation). Here is an alternative account of explanation for the issue at hand (a *behaviorist account*):  $S$  believes  $\phi$  *because* she has reasons  $R$  when, if  $S$  were asked why she believes  $\phi$  she would explain this by citing her reasons  $R$  (or something like that). Here is another alternative (a *counterfactual dependence account*):  $S$  believes  $\phi$  *because* she has reasons  $R$  when, if it were true that  $S$  did not have reasons  $R$ , she would not believe  $\phi$  (or something like that). We thank Peter Klein for suggesting this point. We refrain from assuming the causal account of the basing relation mainly because we agree with an observation made by Klein in his (2012). Klein argues that causal explanations are not good explanations of how we acquire knowledge, for we do not know enough about the *actual causes* of our beliefs. Our claim, for example, that beliefs constituting inferential knowledge are *caused by* the reasons we have for those beliefs is highly speculative and prone to empirical disconfirmation (for all we know, those beliefs can be caused by many other things). The same observation applies to the causal account of the basing relation: given our poor knowledge about mental causation, causal explanations are not good explanations of how beliefs are *based on* reasons. Although this does not render the causal account false, it makes it problematic.

<sup>53</sup>We said before that some state-transitions from doubts to doubts also count as inference. In this case, instead of saying that ‘ $S$  infers that  $\phi$ ’ we say that ‘ $S$  inferentially doubts that  $\phi$ ’.

<sup>54</sup>It may be natural in the latter sense, unnatural in the former one, to say that  $S$  *infers a contradiction* from a certain set of premises.



argument whose conclusion is  $\phi$  and whose premises are  $\{\psi_1, \dots, \psi_n\}$ , etc. In general, we must distinguish *argument* from *inference*. An argument is a structure where some set of propositions or sentences constitute the premises and a certain proposition or sentence constitute the conclusion (and agents in some sense ‘build’ arguments). An inference, however, is a type of cognitive process whose *relata* are doxastic attitudes towards propositions — not the propositions themselves.

To sum up, an inference is assumed to be a type of cognitive process whose *relata* are doxastic attitudes. By means of processes of this type cognitive agents perform state-transitions, where the inferential doxastic attitudes (the ones ‘arrived at’) are *based on* the pre-inferential doxastic attitudes (the ones from which one ‘departed’). We will proceed with these minimal assumptions about inference and, if necessary, we will get back to the question: *What is an inference?*

## 2.2 Inferential schemata (*i*-schemata)

Before we explicate what it is for someone to know how to perform an inference, some technical details are in order. We have to introduce the concepts of *inferential schema* (or *i*-schema) and *substitution instance* of the input/output-variable of an *i*-schema, as well as the formal apparatus that we will use to represent the objects denoted by these concepts.

Let us begin by introducing the concept of *inferential schema*. Roughly, *i*-schemata are descriptive of *ways* of performing inferences. These schemata are structures of the following type:

$$\begin{array}{ll} \text{IF} & \{B\psi_1, \dots, B\psi_n\} \\ \text{THEN} & B\phi, \end{array}$$

where  $\{B\psi_1, \dots, B\psi_n\}$  is a variable for the input to the *i*-schema and  $B\phi$  is a variable for its output (we will use the term ‘pre-inferential belief’ to refer to the input of an *i*-schema and the term ‘inferential belief’ to refer to its output). If we name the above schema using the Greek letter ‘ $\alpha$ ’, we can represent it as a function in the following way:  $\alpha(B\psi_1, \dots, B\psi_n) = B\phi$ . That is, we can represent it as a function that takes a set of doxastic attitudes as argument and returns a doxastic attitude as value (where  $R = \{B\psi_1, \dots, B\psi_n\}$ , we can also use ‘ $\alpha(R) = B\phi$ ’ to represent the *i*-schema above). We will use this formal representation from now on<sup>55</sup>. Extensionally, each inferential schema corresponds to a set of (ordered) pairs of states where certain doxastic attitudes towards

<sup>55</sup>This is what we will call an ‘intensional representation’ of inferential schemata in *Chapter 5*.

propositions represented in a certain language  $\mathcal{L}$  hold (this will be made clearer in *Chapter 5*).

As an example, consider an *i-schema* by means of which one could form a belief in the consequent of a conditional on the basis of one's beliefs in both, the relevant conditional and its antecedent – call it '*dmp*' for '*doxastic modus ponens*'. It can be represented as follows:

*dmp*:  
 IF  $\{B(\phi \rightarrow \psi), B\phi\}$   
 THEN  $B\psi$ .

It can also be represented in our functional notation by:  $dmp(B(\phi \rightarrow \psi), B\phi) = B\psi$ .

The input/output of an *i-schema* can contain not only beliefs, but also doubts or attitudes of suspension of judgment (we use '*D* $\phi$ ', remember, to represent the state of suspension of judgment about  $\phi$ ). The following *i-schema* (where '*dc*' stands for '*doubt over conjunction*'), for example, represents a way of doubting a conjunction on the basis of attitudes of doubt towards each of its conjuncts:

*dc*:  
 IF  $\{D\phi, D\psi\}$   
 THEN  $D(\phi \wedge \psi)$ .

Or in our functional notation:  $dc(D\phi, D\psi) = D(\phi \wedge \psi)$ .

So far we gave no precise interpretation of the 'IF... THEN...' structure in our *i-schemata*<sup>56</sup>. Briefly, there are at least three ways of interpreting them. First, one may point out that our *i-schemata* look like *procedural rules*, of the type cognitive psychologists usually talk about<sup>57</sup>. The relevant procedural rules are analogous to conditionals in programming languages (for example, *C* and *Python*): the 'IF' part contains a certain *condition* and the 'THEN' part contains a *command* for a certain action. In this case, if the condition is satisfied by a certain agent (if she has beliefs  $B\psi_1, \dots, B\psi_n$ ), then she is commanded, in some sense, to perform the relevant action (where the 'action' here is to believe  $\phi$ ). According to this interpretation — call it the '*imperativist interpretation*' — the 'THEN' part hides an imperative: IF this-and-that condition is satisfied, THEN *do* such-and-such (where the command '*do*' is interpreted as '*believe*').

Second, one may suggest that our *i-schemata* are conditionals whose consequent is

<sup>56</sup>It is clear that, as it stands, the 'IF... THEN...' structure cannot be formalized by a material conditional: neither what follows the 'IF' part is a complete assertoric sentence nor what follows the 'THEN' part is a sentence of this type.

<sup>57</sup>See *Chapter 3* of Thaghard (2005).

modalized with a *deontic operator*: IF this-and-that condition is satisfied THEN such-and-such is *permitted*, or: IF this-and-that condition is satisfied THEN such-and-such is *obligatory*. Call this the ‘*deontic interpretation*’ about the nature of *i-schemata*<sup>58</sup>.

Finally, according to the third interpretation — call it the ‘*process-type interpretation*’ — inferential schemata are general patterns of inference, or types of inference, and they contain no hidden imperatives or deontic operators. In this case, a more accurate representation of the relevant schemas would be: IF *input is* so-and-so THEN *output is* such-and-such. In this case, using the notion of *i-schemata* to explicate inferential justification would consist in advancing a process-reliabilist theory about inferential justification<sup>59</sup>.

We will not argue in favor of any of these interpretations here — this is not the purpose of the present work. Whatever lurks behind the ‘IF... THEN...’ structure of *i-schemata*, the important thing to keep in mind is that these structures represent ways of performing inferences. There are two questions we will need to answer involving *i-schemata* (the question about the nature of these structures not being one of them). The first one is: What is it for *S* to *instantiate* a particular *i-schema* at a particular time? The second one is: What is it for an *i-schema* to be *available* to a certain agent at a certain time? We will try to answer both questions in *Section 2.4*.

We still need to make some important points about the nature of inferential schemata. First, we are not assuming that *i-schemata* are ‘internally represented’<sup>60</sup>. That is to say that, when we claim that a certain subject *S* reasons in accordance with an *i-schema*, we are not implying that *S* has any intensional attitude towards the *i-schema* itself, or that *S* entertains it as a rule. Inferential schemata are not assumed to be ‘in the heads’ of reasoners.

Second, we can attribute the possession and use of these *i-schemata* in order to *explain* cognitive performance. Mary believes that *it is raining* and she believes that *if it is raining then there are clouds in the sky*. On the basis of these beliefs she infers that *there are clouds in the sky*. We can explain Mary’s cognitive performance by saying that an inferential schema like *dmp* was available to her and that she reasoned in accordance with it. We need not commit ourselves to the idea that *i-schemata* are actually part of the cognitive architecture of reasoners, however, as if they were constitutive of something like a language of thought<sup>61</sup>. The notion of inferential schema (that which is *instantiated* by particular reasoners) can be seen here, rather, as a useful device for explanations of

<sup>58</sup>There is a similarity between these first two interpretations about the nature of *i-schemata* and two views about epistemic rules discussed by Boghossian in his (2008): rules as imperatives and rules as norms.

<sup>59</sup>The *locus classicus* of process-reliabilism about justification is Goldman (1979).

<sup>60</sup>See Pylyshyn (1991) for a discussion (in the context of cognitive psychology) about rules being ‘internally represented’.

<sup>61</sup>About the language of thought hypothesis, see Fodor (2008).

cognitive performance from a third–person point of view.

Third, we should not equate *i–schemata* with *epistemic norms*. Epistemic norms are supposed to express what is epistemically correct, rational or justified for one to believe in such–and–such situation or, when they do not address specific types of situations, they state general requirements for epistemic rationality (like ‘It is not rational to hold beliefs in contradictions’, etc)<sup>62</sup>. Inferential schemata, however, express no such thing. One can reason in accordance with an *i–schema* and still fail to conform to any epistemic norm. Further, no reference to what is epistemically correct or permitted is made in our inferential schemata (unless, of course, we choose the *deontic interpretation* presented above — but we are not taking this interpretation for granted).

The other concept relevant to our discussion is that of a *substitution instance*. There are substitution instances of *single* doxastic attitude–variables and substitution instances of *sets* of doxastic attitude–variables. A substitution instance of a *single* doxastic attitude–variable whose content placeholder is represented by a formula with *variables* of  $\mathcal{L}$  is a doxastic attitude whose content is represented by a formula with *constants* of  $\mathcal{L}$  — constants that uniformly and rightly substitute those variables ( $\mathcal{L}$  is called a ‘parameter–language’ in these situations). A substitution instance of a *set* of doxastic attitude–variables is simply a set of substitution instances of those doxastic attitude–variables, where for each doxastic attitude–variable in the former set there is one and only one substitution instance for it in the latter one.

For example, consider the language of propositional logic ( $\mathcal{PL}$ ) where  $\phi, \psi$  are variables for atomic formulas (and atomic formulas *only*) and the constants  $p, q, r, s$  are atomic formulas. Here, both  $Bp$  and  $Bq$  are substitution instances of  $B\phi$  when the parameter–language is  $\mathcal{PL}$ . Further, all of the following are substitution instances of the set  $\{B\neg\phi, B(\phi \vee \psi)\}$  when the parameter–language is  $\mathcal{PL}$ :  $\{B\neg p, B(p \vee q)\}$ ,  $\{B\neg q, B(q \vee p)\}$ ,  $\{B\neg s, B(s \vee p)\}$ ,  $\{B\neg r, B(r \vee s)\}$ . Given that  $\phi$  and  $\psi$  are variables for *atomic* formulas, the following *is not* a substitution instance of  $\{B\neg\phi, B(\phi \vee \psi)\}$  when the parameter–language is  $\mathcal{PL}$ :  $\{B\neg(p \wedge q), B((p \wedge q) \vee r)\}$ . That is because  $(p \wedge q)$  is not an *atomic* formula and, therefore,  $B\neg(p \wedge q)$  is not a substitution instance of  $B\neg\phi$  (we are assuming here that, in  $\mathcal{PL}$ ,  $\phi$  is a variable for atomic formulas only).

In general, when  $B\psi$  is a substitution instance of  $B\phi$  and the parameter–language is  $\mathcal{L}$ ,  $\phi$  and  $\psi$  should have the same syntactic complexity in  $\mathcal{L}$ . Substitution instances must uniformly replace variables with constants, and they should not add or remove logical constants. We do not want to say, for example, that  $B(p \wedge \neg p)$  is a substitution instance of  $B(\phi \wedge \psi)$  when the parameter–language is  $\mathcal{PL}$ .

---

<sup>62</sup>Pollock and Cruz (1999, p. 123), for example, conceive epistemic norms as ‘norms describing when it is epistemically permissible to hold various beliefs’.

Similar considerations apply to other, more sophisticated languages. As an example, consider a language  $\mathcal{LM}$  where:  $x, y$  are variables for objects in a certain domain  $D$ ;  $a, b$  are constants denoting objects in  $D$ ;  $P, Q$  are unary (or *one-place*) predicates whose extensions are subsets of  $D$ . Finally,  $\mathcal{LM}$  contains a restricted quantifier represented by ‘90%’. The formula  $(90\% x \in P)Qx$ , for example, reads as ‘ninety percent of the objects that are  $P$  are also  $Q$ ’. Now let us represent the contents of doxastic attitudes with  $\mathcal{LM}$  and form the relevant sets accordingly. Here, both sets of doxastic attitudes  $\{B(Pa), B((90\% x \in P)Qx)\}$  and  $\{B(Pb), B((90\% x \in P)Qx)\}$  are substitution instances of the set  $\{B(Px), B((90\% x \in P)Qx)\}$  when the parameter-language is  $\mathcal{LM}$ . The variable  $x$  in the quantified formula  $(90\% x \in P)Qx$  is not replaced by a constant in these cases because the rules of  $\mathcal{LM}$  do not allow for such replacement (in that context,  $x$  is a *binded variable*). It will also depend on the rules of  $\mathcal{LM}$  if two distinct free variables can be replaced by the same constant. That is, whether or not  $\{B(Pa), B(Qa)\}$  counts as a substitution instance of  $\{B(Px), B(Qy)\}$  when the parameter-language is  $\mathcal{LM}$  will be determined by the very rules of  $\mathcal{LM}$ .

Substitution instances are always indexed to a relevant language with a particular set of rules. In order to represent the set of substitution instances of a certain set of doxastic attitude-variables whose parameter-language is  $\mathcal{L}$  we use a function  $si^{\mathcal{L}}$ . So  $si^{\mathcal{L}}(B\psi_1, \dots, B\psi_n)$  is the set of all substitution instances of  $\{B\psi_1, \dots, B\psi_n\}$  when the parameter-language is  $\mathcal{L}$  (where  $R = \{B\psi_1, \dots, B\psi_n\}$  we can represent the set of substitution instances of  $\{B\psi_1, \dots, B\psi_n\}$  parameterized with  $\mathcal{L}$  using ‘ $si^{\mathcal{L}}(R)$ ’ as well). Similarly,  $si^{\mathcal{L}}(B\phi)$  is the set of all substitution instances of  $B\phi$  when the parameter-language is  $\mathcal{L}$ <sup>63</sup>.

To pinpoint *particular* substitution instances of a doxastic attitude-variable or set of doxastic attitude-variables whose parameter-language is  $\mathcal{L}$ , we will also use natural numbers ( $n$ ) as indexes to the function  $si^{\mathcal{L}}$ . The values of two applications of a  $si_n^{\mathcal{L}}$  to two different arguments must be *uniform* with each other. To illustrate how that works, let us consider again the set  $\{B\neg\phi, B(\phi \vee \psi)\}$  and its substitution instances in  $si^{\mathcal{PL}}$  (that is, its substitution instances when the parameter-language is  $\mathcal{PL}$ ), but this time let us consider it alongside  $B(\psi)$ . In this case, if  $si_1^{\mathcal{PL}}(B\neg\phi, B(\phi \vee \psi)) = \{B\neg p, B(p \vee q)\}$ , then  $si_1^{\mathcal{PL}}(B\psi) = Bq$ . That is because  $si_1^{\mathcal{PL}}$  mapped the variable  $\psi$  to the constant  $q$  in the first case and it has to do the same thing in the second one.

So, roughly, this is how the concepts of *inferential schema* and *substitution instance* must be understood here (although we still need to spell out some important formal properties of *i-schemata*). In *Section 2.4* below, we will use these concepts to explicate both, what it is for a subject to instantiate an *i-schema* and what it is for an *i-schema* to be

---

<sup>63</sup>Sometimes, when it is clear from the context that the function  $si$  should be used with a *variable* for a language  $\mathcal{L}$  as index (as in ‘ $si^{\mathcal{L}}$ ’), we will just use ‘ $si$ ’.

available to someone. Before doing the explication, however, let us consider an important distinction between *optimal* and *non-optimal* inferential schemata.

## 2.3 Optimality

In *Chapter 1* we used the concept of *epistemically approvable types of inference*, or *epistemically approvable ways of performing inferences*. Here, we will distinguish the approvable ways of performing inferences from the non-approvable ones by distinguishing *optimal* from *non-optimal i-schemata*. More precisely, an approvable way of performing an inference is an *optimal i-schema* and a non-approvable way of performing an inference is a *non-optimal i-schema*. How should we distinguish *optimal* from *non-optimal* inferential schemata?

Roughly, an *optimal* inferential schema is descriptive of a *right way* of performing an inference. Optimality assessments are supposed to capture epistemically correct ways for one to reason — they consist in procedural evaluations. The optimality value of an *i-schema* is a function of the fact that one would/would not conditionally maximize the epistemic goal by forming doxastic attitudes that constitute the output of the *i-schema* on the basis of doxastic attitudes that constitute the input of the *i-schema* across a relevant range of cases<sup>64</sup> (we will not try to define what exactly is the range of cases where conditional maximization of the epistemic goal is supposed to occur). The ‘cases’ here are actually substitution instances of the input/output variable of the *i-schema*. When  $\alpha(R) = B\phi$  is an *optimal* inferential schema, an agent whose reasons constitute some particular substitution instances of  $R$  (and those substitution instances alone)<sup>65</sup> would conditionally maximize the epistemic goal by forming a belief that is a certain substitution instance (uniform with the first one) of  $B\phi$ . It may be that *some* particular substitution instances  $si_n(R)$  and  $si_n(B\phi)$  are such that  $S$  would not conditionally maximize the epistemic goal by forming  $si_n(B\phi)$  on the basis of  $si_n(R)$ , but that does not mean that  $\alpha(R) = B\phi$  is *non-optimal*.

---

<sup>64</sup>The ‘epistemic goal in our context, remember, is the goal of believing truths and not believing falsehoods. We should note that the view that truth is the primary epistemic goal has been challenged — see Kvanving (2014). Further, *we are not* assuming that truth is the primary epistemic goal. We are only assuming that epistemic rationality, in the sense that interests us, requires conditional maximization of that epistemic goal.

<sup>65</sup>The proviso ‘and those substitution instances alone’ is added here because an agent could have reasons  $si_n(R)$  and also further reasons  $R^*$  such that  $S$  would not conditionally maximize the epistemic goal by forming  $si_n(B\phi)$  (because the content of  $si_n(R)$  plus the content of  $R^*$  fails to give support to  $\phi$ ). We make this idealization only for the purposes of explicating the optimality of an *i-schema* here. As we will see in the next section, we also use this idealization to explicate what it is for an *i-schema* to be available to someone.

We can try to take these considerations into account by making reference to *valid* or *inductively strong* arguments whose premises have the forms (in some particular schematic language) of the contents in the *i-schema*'s pre-inferential beliefs and whose conclusion has the form of the content in the *i-schema*'s inferential belief<sup>66</sup>. Let us use the ordered pair  $\langle \{\psi_1, \dots, \psi_n\}, \phi \rangle$  to represent an argument whose premises are  $\{\psi_1, \dots, \psi_n\}$  and whose conclusion is  $\phi$ . Where  $R = \{B\psi_1, \dots, B\psi_n\}$ , an inferential schema  $\alpha(R) = B\phi$  will be said to be *optimal* when  $\langle \{\psi_1, \dots, \psi_n\}, \phi \rangle$  is a *valid* or *inductively strong* argument.

There are several details that need to be worked out here, but we cannot deal with all of them. Still, we need to make some important points about this way of defining optimality. First, we are using the notion of 'inductively strong argument' (or 'inductively strong *argument form*') as an umbrella term for several kinds of non-deductive arguments (or non-deductive *argument forms*), including: probabilistically valid arguments<sup>67</sup>, arguments whose premises maintain a relation of partial entailment with the conclusion<sup>68</sup>, arguments to the best explanation, statistical syllogisms<sup>69</sup>, enumerative induction, statistical generalization, induction by analogy, etc.

Second, the concept of *validity* (and its accompanying concept of *entailment* or *logical consequence*)<sup>70</sup> used for the purposes of defining optimality must be non-classical. There are two features of classical validity that is a good idea for us to avoid. The first one is that classical validity does not require the premises to be *relevant* to the conclusion. An argument with  $q$  as premise and  $(p \vee \neg p)$  as conclusion is classically valid. But we do not want an *optimal i-schema* to conditionally maximize the epistemic goal by delivering outputs whose contents are disconnected from the contents of the input (in general, we do not want *optimal i-schemata* to maximize the epistemic goal *at any costs* — maximization of the epistemic goal is just a *necessary* condition for optimality). That means that we want to use a concept of *relevant validity* — a concept used in *Relevance Logics*<sup>71</sup>. The second one is that classical validity allows for the principle of 'explosion'<sup>72</sup> or 'trivialization'<sup>73</sup>: anything follows from a contradiction. We do not want to regard an *i-schema* that spits anything as output given a contradictory input as *optimal*. That means that we want to

---

<sup>66</sup>One might ask: Why do not we take that into account by using, again, the notion of *support relation*? The idea would be to say that an *i-schema*  $\alpha(B\psi_1, \dots, B\psi_n) = B\phi$  is *optimal* when  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ . But that would require a further notion of *support relation* — one that is supposed to hold between *propositional forms* (the notion of *support relation* we have been using is not *schematic*: it holds between actual *propositions*, not between *propositional forms*). We avoid that complication by using the notions of *validity* and *inductive strength*, as these notions may be taken to apply to argument forms.

<sup>67</sup>See Adams (1996), Hailperin (1996).

<sup>68</sup>See Adams (1996).

<sup>69</sup>See Pollock (1995), *Chapter 2, Section 4*.

<sup>70</sup>See Beall and Restall (2005).

<sup>71</sup>See Dunn (1986).

<sup>72</sup>See Priest (2002).

<sup>73</sup>See Brown (2002).

use a concept of *paraconsistent validity* — a concept used in *Paraconsistent Logics*<sup>74</sup>.

What our brief considerations reveal is that optimality *is not just* a function of maximizing the epistemic goal: considerations about the appropriate relations between the contents of pre-inferential and inferential beliefs are also important. The fine-grained details need not be of our concern now — the basic idea can be captured from what we have just said. From now on, let us assume that a *good argument* is an argument that is either valid or inductively strong. We can define optimality for inferential schemata with beliefs as input and a belief as output, then, as follows (where ‘*BB*’ reads as ‘from beliefs to beliefs’)<sup>75</sup>:

(BB) Where  $R$  is a set of beliefs, an inferential schema  $\alpha(R) = B\phi$  is *optimal* when and only when  $\langle [R], \phi \rangle$  is a good argument.

So, still under the assumption that  $R$  is a set of beliefs, an inferential schema  $\alpha(R) = B\phi$  is *non-optimal* when and only when  $\langle [R], \phi \rangle$  is *not* a good argument. So far so good for *i-schemata* whose input is a set of beliefs and whose output is a single belief (that is, *i-schemata* of the *BB-type*). But what about *i-schemata* whose input is a set of beliefs and whose output is an attitude of doubt (or *i-schemata* of the *BD-type*)?

Let us assume that the proposition *There is a 50/50 chance that  $\phi$  is true* is accurately represented by ‘ $Pr(\phi) = 0.5$ ’. A simple example of an optimal *i-schema* with a state of doubt as output would be (where ‘*ff*’ is short for ‘*fifty-fifty*’):

*ff*:  
 IF  $B(Pr(\phi) = 0.5)$   
 THEN  $D\phi$ .

Notice that, despite the fact that *ff* seems to be *optimal*, the definition of optimality present in (BB) does not allow us to judge if *ff* is optimal. That is because the output of *ff* is not a belief. Here is a definition of optimality for inferential schemata of the *BD-type*:

(BD) Where  $R$  is a set of beliefs, an inferential schema  $\alpha(R) = D\phi$  is *optimal* when and only when neither  $\langle [R], \phi \rangle$  nor  $\langle [R], \neg\phi \rangle$  is a good argument.

Where neither  $\langle [R], \phi \rangle$  nor  $\langle [R], \neg\phi \rangle$  is a good argument, an epistemically optimal inferential schema with  $R$  as input will not deliver  $B\phi$  or  $B\neg\phi$  — it will rather deliver  $D\phi$ .

<sup>74</sup>See Priest, Tanaka and Weber (1996).

<sup>75</sup>The function  $\lceil \ ]$ , remember, maps from a set of *doxastic attitudes* to a set of *propositions* (those constituting the contents of the relevant doxastic attitudes). So if  $R = \{B\psi_1, \dots, B\psi_n\}$ , then  $\lceil R \rceil = \{\psi_1, \dots, \psi_n\}$ .



We saw in *Chapter 1* that there are cases where one's reason for suspending judgment about something is a further attitude of doubt that one has. That means that we also have to take into account inferential schemata of the *DD-type*, that is, inferential schemata with doubts as input and a doubt as output. As an example, consider the following inferential schema (where 'dd' stands for 'doubt the disjunction'):

*dd*:  
 IF  $B\{D\phi, D\psi\}$   
 THEN  $D(\phi \vee \psi)$ .

It is clear that this is an epistemically approvable state–transition: as a result of suspending judgment about two particular propositions, one would do the right thing by suspending judgment about their disjunction. So, *dd* looks like an *optimal* inferential schema. One could think that this is so because  $\langle \{\phi, \psi\}, \phi \vee \psi \rangle$  is a good argument: if the premises were *neutral* about the conclusion, attitudes of doubt over the former would not make doubting the latter epistemically approvable. That would lead us to the following definition: where  $R$  contains only doubts, an *i-schema*  $\alpha(R) = D\phi$  is optimal when and only when  $\langle [R], \phi \rangle$  is a good argument. But that is wrong, for we want to say that an inferential schema  $\alpha(R) = D\phi$  (where  $R$  contains only attitudes of doubt) is *optimal* even when  $\langle [R], \phi \rangle$  is not a good argument. Suppose Amanda doubts that *Claudio is a philosopher*. Given that much (and only that much) would she conditionally maximize the epistemic goal by believing that *The skateboard was born in California*? Of course not, and neither would she conditionally maximize the epistemic goal by disbelieving (or believing the negation of) that proposition: the epistemically correct thing for her to do is to suspend judgment again. That would give us the following:

(DD) Where  $R$  is a set of doubts, an inferential schema  $\alpha(R) = D\phi$  is *always optimal*, unless  $\phi$  is a tautology.

A tautology is understood here as a proposition which is true as a matter of form alone. We make this *proviso* because optimality is supposed to be a function of conditional maximization of the epistemic goal accross a relevant range of cases — but if  $\phi$  is a tautology *there is no case* where one conditionally maximizes the epistemic goal by forming a substitution instance of  $D\phi$ .

One could think that a similar *proviso* must apply to:

(DB) Where  $R$  is a set of doubts, an inferential schema  $\alpha(R) = B\phi$  is *never optimal*,

because one would always (conditionally or not) maximize the epistemic goal by believing a tautology. We reject this suggestion, though. The reason is similar to the one we

offered for not choosing classical validity to define optimality for inferential schemata of the *BB-type*. We saw above that optimality for inferential schemata of the *BB-type* is not just a function of maximizing the epistemic goal, but also of there having an appropriate connection between the contents of the input and the content of the output (a *relevance* connection). The same can be said about (DB) for cases where  $B\phi$  is a belief in a tautology:  $R$  would need to contain a belief in something that is properly (relevantly) connected with  $\phi$  (a theorem, or an axiom, another tautology, etc). But  $R$  is a set of doubts in an *i-schema* of the *DB-type*: there is no belief in it. Therefore, there is no belief in it whose content is properly connected with  $\phi$ <sup>76</sup>.

Would that mean that *there is no optimal* inferential schema that outputs beliefs in tautologies, except for the ones whose inputs contain beliefs in contents properly connected with those tautologies? It may be assumed that one does not need reasons (understood as doxastic attitudes) to form rational beliefs in at least some tautologies, either because these truths can be rationally believed on the basis of *rational insight* or because one can reach these truths by reasoning from *assumptions* (for example, by performing *reductios*).

In the first case, this should not be a concern for us, since we are dealing with *inference* only, not with other sources of belief such as perception, rational insight, memory, etc. When it comes to the second one, we would need to allow for inferential schemata with *assumptions* as input as well. We have been assuming that the domain of the functions representing *i-schemata* is constituted by sets of beliefs and doubts, but it is clear that a general theory of reasoning should also include assumptions in the relevant domain. We welcome this suggestion — we have been working solely with beliefs and doubts just for the sake of simplicity, in order to offer a more easy-going exposition<sup>77</sup>. But the present point is that *i-schemata* with assumptions as input would not be *i-schemata* of the *DB-type* anymore and, therefore, this is not a problem for (DB). As long as  $R$  contains only doubts,  $\alpha(R) = B\phi$  is not an approvable way of performing an inference.

We just presented one way of defining optimality for *i-schemata* of the *BB*, *BD*, *DD* and *DB-types* — but we will mostly occupy ourselves with inferential schemata of the *BB-type* (conclusions about *i-schemata* of the other types will be more or less straightforward given what we will say about *i-schemata* of the *BB-type*). As we said before, there are details that still need to be worked out here, but one can get the general picture from our definitions. The notion of optimality will be used to explicate what it is for someone

<sup>76</sup>One may get the impression that it is wrong to say that an inferential schema  $\alpha(R) = B\phi$  of the *DB-type* is never *optimal* because one conceives of bodies of reasons that contain not only doubts, but also beliefs. Nevertheless, when we are trying to judge whether a certain *i-schema* of this type is optimal we must consider *only* its input (a set of doubts) and then try to judge if forming a certain belief on the basis of those doubts (and only those doubts) is epistemically approvable or not.

<sup>77</sup>We will not try to define optimality for inferential schemata with assumptions as input — this is a job for future work.

to know how to perform an inference later (*Section 2.5*).

## 2.4 Instantiation and availability of *i*-schemata

With our technical apparatus and definitions in place, we can now explain what it is for someone *to instantiate an i-schema* and what it is for an *i-schema to be available to* someone. Let us begin with instantiation.

It is noteworthy that there is a similarity between the question ‘What does it take for a subject to instantiate an *i-schema*?’ and the question ‘What does it take for a subject to follow an epistemic rule?’<sup>78</sup>. For example, we take it that it is not sufficient for *S* to instantiate an inferential schema  $\alpha(R) = B\phi$  that *S* forms a doxastic attitude that is a substitution instance of  $B\phi$ ,  $si_n(B\phi)$ , on the basis of reasons that constitute a substitution instance of  $R$ ,  $si_n(R)$ <sup>79</sup>. It may be that *S* ‘deviantly’ forms a belief  $si_n(B\phi)$  on the basis of reasons  $si_n(R)$ , in a way unrelated to the availability of the optimal inferential schema  $\alpha$ . Similarly, one can do what is recommended by a rule in a ‘deviant’ way — not *because* one followed the rule.

But we are not assuming that to instantiate an *i-schema* is to follow a rule. As we briefly saw above, there are at least three interpretations about the nature of *i-schemata*. According to one of them (the ‘*process-type interpretation*’: inferential schemata are general patterns or *types* of inference), *i-schemata* are not rules. In this case, to instantiate an *i-schema* is just to instantiate a type of inferential process. As we did not decide which one is the best interpretation about the nature of inferential schemata, we cannot decide if the instantiation of an inferential schema entails rule-following or not. Further, we are assuming that in order for an agent to instantiate an *i-schema* she does not need to realize that there is a match between the IF part of the *i-schema* and her pre-inferential beliefs, as is sometimes assumed in psychological explanations about what it is to follow a rule in reasoning<sup>80</sup>.

Be that as it may, this is how we explicate the instantiation of an *i-schema* (where  $R$  is a set of beliefs):

- (I) *S* instantiates an inferential schema  $\alpha(R) = B\phi$  at  $t$  if and only if (1) *S* forms a doxastic attitude  $si_n(B\phi)$  on the basis of *S*’s reasons  $si_n(R)$  at  $t$ , and (2)  $si_n(B\phi)$  is formed by *S* on the basis of  $si_n(R)$  *because* the inferential schema  $\alpha$  was available to *S* at  $t$ .

<sup>78</sup>About following an epistemic rule, see Boghossian (2008).

<sup>79</sup>Notice that we attach the same index ( $n$ ) to the function *si* in the two situations, which means that these substitution instances are uniform with each other.

<sup>80</sup>See Smith, E., Langston, C., Nisbett, R. (1992, p. 3).

Notice that the mere correspondence or matching between the input/output-variables of  $\alpha$  and their substitution instances in  $S$ 's cognition is not sufficient for  $S$  to instantiate  $\alpha$ . The fact that the inferential schema is available to  $S$  must explain why  $S$  believed as she did. Of course, we should not expect that the availability of the inferential schema is the *only* explanation why  $S$  believed the relevant proposition — presumably such an explanation would also mention the fact that  $S$  used her reasons  $R$  to form the relevant belief. It is one thing to explain why  $S$  believes  $\phi$  (*simpliciter*) and another to explain why  $S$  believed  $\phi$  in the way she did. It is the latter explanation that is relevant to (I).

We can draw an analogy here between our thesis about the instantiation of *optimal* inferential schemata and Sosa's theory of knowledge (2007, p. 23). Roughly, Sosa requires for a belief to be a case of knowledge the following properties: *accuracy* (the belief is true), *adroitness* (the formation of the belief manifests competence) and *aptness* (the belief is accurate *because* it is competently formed). In a similar way, we require for an inference to be an instantiation of an *optimal i-schema* that: (i) the content of the inferential belief receives support from the content of the pre-inferential beliefs (something analogous to *accuracy*), (ii) the formation of the inferential belief manifests availability of an *i-schema* (something analogous to *adroitness*), and (iii) the inferential belief is formed on the basis of the pre-inferential ones *because* the relevant inferential schema was available (something analogous to *aptness*). It is the third condition that excludes cases where one believes something on the basis of the right reasons but in the wrong way from the instantiations of an *optimal i-schema*.

As an example, suppose that Amanda forms a belief that *Claudio is a philosopher or Claudio is a happy man* on the basis of Amanda's belief that *Claudio is a philosopher*. It may appear that Amanda is instantiating an inferential schema like  $\alpha(B\phi) = B(\phi \vee \phi)$  here. However, let us suppose that Amanda would believe that *Claudio is a philosopher or Claudio is a happy man* on the basis of *any* belief whose content happens to mention *Claudio*. That means that Amanda's belief is formed in a silly way — and we do not want to say that she instantiated  $\alpha(B\phi) = B(\phi \vee \phi)$  by believing that *Claudio is a philosopher or Claudio is a happy man* on the basis of her belief that *Claudio is a philosopher*<sup>81</sup>. So our condition (I2) handles this case: *it is not because*  $\alpha(B\phi) = B(\phi \vee \phi)$  is available to Amanda that she forms the target belief on the basis of her reasons. That is: that  $\alpha(B\phi) = B(\phi \vee \phi)$  is available to her is not part of the explanation why she believed the target proposition on the basis of her reasons.

There is a crucial notion in (I) that was not explicated so far: the notion of *availability of an i-schema*. Roughly, given that *i-schemata* are supposed to represent ways of performing inferences, for an *i-schema* to be available to someone is for that person *to have a*

<sup>81</sup>Similar examples are also presented by Turri (2010, p. 317) and Goldman (2011, p. 133).

*way* of performing a certain piece of reasoning. There are two important features that, in general, we expect a person to exhibit when we assume that she has a way of performing a certain inference<sup>82</sup>.

First, we expect the following to be true: *if* the person were asked (by herself or by others) if the contents of the relevant pre-inferential beliefs give support to the content of the relevant inferential belief, she would make a positive judgment and answer ‘yes’ accordingly. For example, let us assume that Amanda has a way of performing an inference from her belief that *Claudio is a philosopher and 99% of the philosophers are teachers* to a belief that *Claudio is a teacher*. Under that assumption, we expect Amanda to judge that *Claudio is a philosopher and 99% of the philosophers are teachers* gives support to *Claudio is a teacher* if we prompt her to do so. Of course, we are not *always* entitled to expect such a thing when we assume that someone has a way of performing a certain inference. There are cases where one may be able to perform a certain inference without being able to recognize that the relevant support relation holds. We will see one example of this type in *Chapter 3*. For now, however, let us just note that if we want to include this feature in the explication of what it is for someone to have an *i-schema* available (or what is it for someone to have a way of performing an inference), we will need to make some idealization. This will be made clearer as we proceed.

So, when we attribute the possession of an *i-schema* to a person we expect her to judge, in some counterfactual situation, that some relevant support relation obtains. Someone might think that by acknowledging the legitimacy of this feature we are committing ourselves to the thesis that in order for a reasoner to perform an inference she must judge that the ‘premises’ (the contents of the pre-inferential beliefs) give support to the ‘conclusion’ (the content of the inferential belief). Boghossian (2012), for example, sustains that any account of inference must satisfy the following condition:

(*Taking Condition*) Inferring necessarily involves the thinker *taking* his premises to support his conclusion and drawing his conclusion *because* of that fact.

The feature we have been talking about, however, is quite different from Boghossian’s *Taking Condition*. Boghossian’s thesis is about actual inference, our about *inferring*, while our feature is about *having a way* of performing an inference. While someone cannot perform an inference without having a way of performing an inference, someone can have a way of performing an inference without performing an inference. We do not subscribe to Boghossian’s condition<sup>83</sup>, but the present point is that by acknowledging the legitimacy

<sup>82</sup>For the sake of simplicity, in what follows we will talk about inferences from beliefs to beliefs only.

<sup>83</sup>We are simply stating that we do not subscribe to the relevant thesis — not that it is false. So, we dispense ourselves of presenting any arguments here. Dealing with Boghossian’s *Taking Condition* is a task for another work.

of the feature described above we are not committed to the *Taking Condition*. We are committed, however, with something like a *Counterfactual Taking Condition*, which can be put as follows: if  $S$  has a way of performing an inference from  $R$  to  $B\phi$ , then  $S$  is disposed to judge that  $[R]$  gives support to  $\phi$ .

We just fleshed out one feature that, in general, we expect a person to exhibit when we assume that she has a way of performing a certain inference. The other one is that we also expect that person to perform the same type of inference across a relevant range of cases. This is some kind of *anti-luck* condition: as lucky success in a particular performance is not sufficient for having an ability to perform a certain action, forming a belief  $si_n(B\phi)$  on the basis of reasons  $si_n(R)$  in one particular situation is not sufficient for having an *i-schema*  $\alpha(R) = B\phi$  available.

Consider: why is it better (from an epistemological point of view) to believe something that gets support from the content of one's reasons by instantiating an *optimal* inferential schema than to do so by instantiating a *non-optimal* one? Or: why is it better to believe something on the basis of good reasons *in the right way* than to do so *in the wrong way*? In the example given above, Amanda believed a proposition of the form  $(\phi \vee \psi)$  on the basis of her belief in a proposition of the form  $\phi$  — but the way she reasoned from the latter to the former is not epistemically approvable.

We can answer the question above in the following way. Let us assume that  $S$  believes something on the basis of good reasons, that is, reasons whose contents give support to what  $S$  believes. It is better for  $S$  to do so by instantiating an *optimal i-schema* than to do so by instantiating a *non-optimal* one because the fact that an *optimal i-schema* was available to  $S$  explains the former but not the latter. That the availability of an *optimal inferential schema* explains the former instantiation means that  $S$  did not get it right (that is, formed a belief on the basis of good reasons) *out of luck* in that instantiation. When an *optimal i-schema* is available to someone, that person will get things right (in this case, believe things on the basis of good reasons) in a variety of situations, not just in a single, lucky one<sup>84</sup>.

Given these two features that we expect a person to have when we assume that she has a way of performing a certain inference (being disposed to believe in support relations, being successful across a variety of situations), let us propose the following:

- (A) An *i-schema*  $\alpha(B\psi_1, \dots, B\psi_m) = B\phi$  is available to  $S$  at  $t$  if and only if, for any  $si_n(B\psi_1, \dots, B\psi_m) = \{B\chi_1, \dots, B\chi_m\}$  and  $si_n(B\phi) = B\sigma$ , if at  $t$   $S$ 's available

---

<sup>84</sup>There is a similarity between our explanation above and the way virtue epistemologists deal with the relationship between knowledge and luck in Gettier-cases. See Sosa (2007), (2011), Greco (1993), (2007) and Zagzebski (1994). Luck is involved in Gettier-cases when one luckily believes what is true. Luck is involved in the cases relevant to our discussion when one luckily believes what gets support from the content of one's reasons.

---

reasons were  $\{B_{\chi_1}, \dots, B_{\chi_m}\}$  (and  $\{B_{\chi_1}, \dots, B_{\chi_m}\}$  alone), then:

- (1) If at  $t$   $S$  were properly prompted to determine if  $\{\chi_1, \dots, \chi_m\}$  (and  $\{\chi_1, \dots, \chi_m\}$  alone) gives support to  $\sigma$ , then  $S$  would believe that  $\{\chi_1, \dots, \chi_m\}$  gives support to  $\sigma$ ;
- (2) If at  $t$   $S$  were properly prompted to find out if  $\sigma$  is true then, provided  $S$  does not take any member of  $\{\chi_1, \dots, \chi_m, \sigma\}$  to be in conflict with anything else she happens to believe as a result of considering these propositions,  $S$  would believe that  $\sigma$  on the basis of  $\{\chi_1, \dots, \chi_m\}$ .

Condition (A1) takes into account the intuition that, when a certain inferential schema is available to  $S$  (when a person is able to reason in a certain way), she ( $S$ ) is disposed to judge that the content of the relevant pre-inferential beliefs gives support to the content of the relevant inferential belief. Of course, if  $S$  does not have the concept of *support relation* the appropriate prompting will provide the conditions for  $S$  to understand such a concept (same thing if  $S$  does not understand any of  $\{\chi_1, \dots, \chi_m, \sigma\}$ ). That is why our conditional begins with ‘If at  $t$   $S$  were properly prompted to determine if...’ — the ‘properly’ qualifier is supposed to exclude cases where  $S$  is prompted with a question she is not able to understand.

Condition (A2) takes into account the intuition that, in order for an *i-schema* to be available to someone, that person must be able to form a certain class of new beliefs on the basis of certain reasons. This condition assures us that the agent can be ‘put to work’, that is, to gather new beliefs in accordance with the *i-schema* available to her. The proviso ‘provided  $S$  does not take any member of  $\{\chi_1, \dots, \chi_m, \sigma\}$  to be in conflict with anything else she happens to believe as a result of considering these propositions’ is added because it may be that  $S$  forms new beliefs after considering the relevant propositions, and it may be that  $S$  comes to conclude that  $\sigma$  is false/probably false (or that one of  $\{\chi_1, \dots, \chi_m\}$  is false/probably false) or that it is not right for her to believe  $\sigma$  (or that it is not right for her to believe some member of  $\{\chi_1, \dots, \chi_m\}$ ). For example, after considering the members of  $\{\chi_1, \dots, \chi_m, \sigma\}$   $S$  may believe that these propositions form an inconsistent set, or that they are paradoxical, or simply that they disconfirm each other. However, we do not want to say that the inferential schema  $\alpha$  is not available to  $S$  just because some particular substitution instances of its input/output variables have such a property.

The subjunctive conditionals in (A1) and (A2) specify idealized scenarios, and we should not expect their truth to hold in some situations. For example, if  $S$  is tired, drunk or confused it may be that she will not believe that  $\{\chi_1, \dots, \chi_m\}$  gives support to  $\sigma$ , even after being properly prompted to judge if this is so. We could include these considerations by adding ‘and if  $S$  were in good shape’, or something like that, to the antecedent of (A1) and (A2). Similarly, we must suppose that  $S$  being stimulated to find out if  $\{\chi_1, \dots, \chi_m\}$

gives support to  $\sigma$  or to find out if  $\sigma$  is true would not cause  $S$  to have a heart attack, and so on. Although such conditions are not made explicit in (A), they should be part of the relevant idealized scenarios. The general idea, we take it, is not hindered by this lack of precision.

Now that we have an explication of what it is for an *i-schema* to be available for someone, let us explicate what it is for a reasoner to know how to reason.

## 2.5 *Knowing how to reason explicated*

In *Chapter 1* we saw that believing  $\phi$  is not rational or inferentially justified for  $S$  unless  $S$  knows how to infer that  $\phi$  from  $R$ . Being propositionally justified in believing  $\phi$  in virtue of reasons  $R$  — that is, satisfying the condition in (PJ) — is not sufficient for being inferentially justified in believing  $\phi$ . One must be able to perform the relevant inference. These conclusions are backed up both by the intuition that it is not rational for  $S$  to form those beliefs that are unreachable to  $S$  and by the fact that being inferentially justified in believing  $\phi$  in virtue of reasons  $R$  is, among other things, being in a position to infer that  $\phi$  from  $R$ .

We will try to fix (PJ) by including the relevant condition of procedural knowledge in the next chapter (*Chapter 3*). But the whole idea of inferential justification as involving propositional justification plus procedural knowledge about how to perform inferences is not fully understood until we make it clear what is required for one to know how to perform an inference. Now we have the appropriate tools to do this. Here is a first proposal:<sup>85</sup>

(KH\*)  $S$  knows how to perform an inference from  $R$  to  $B\sigma$  when and only when there is an *optimal i-schema*  $\alpha(T) = B\phi$  available to  $S$  such that  $R = si_n(T)$  and  $B\sigma = si_n(B\phi)$ , for some  $n$ .

That is,  $S$  knows how to perform a particular inference when and only when a certain inferential schema is available to  $S$ : one such that the pre-inferential belief of that particular inference is a substitution instance of its input variable, and such that the inferential belief of that particular inference is a substitution instance (uniform with the first one) of its output variable<sup>86</sup>. The relevant inferential schema must be *optimal* — otherwise there would be no *knowledge* of how to perform an inference.

---

<sup>85</sup>Both  $R$  and  $T$  are assumed to be sets of doxastic attitudes here — but while the contents of the doxastic attitudes in  $R$  are actual propositions (represented by *constants*), the ‘contents’ of the doxastic attitudes in  $T$  are actually just *content placeholders* (represented by *variables* and logical constants).

<sup>86</sup>(KH\*) looks just like a *conditional analysis* of an ability — see Maier (2010, *Section 3.1*).



---

Consider some simple examples applying (KH\*).  $S$  knows how to perform an inference from  $Bp$  to  $B(p \vee q)$  when and only when there is an *optimal  $i$ -schema* available to her,  $\alpha(B\phi) = B(\phi \vee \psi)$ , such that  $Bp$  is a substitution instance of its input variable and  $B(p \vee q)$  is a substitution instance (uniform with the first one) of its output variable. Likewise,  $S$  knows how to perform an inference from  $B(Pr(p) > 0.9)$  to  $Bp$  when and only when there is an *optimal  $i$ -schema* available to her,  $\alpha(B(Pr(\phi) > 0.9)) = B\phi$ , such that  $B(Pr(p) > 0.9)$  is a substitution instance of its input variable and  $Bp$  is a substitution instance (uniform with the first one) of its output variable.

(KH\*) looks just fine, but there is actually a problem with it. In order to see why, consider the following scenario<sup>87</sup>. Suppose  $S$  believes that  $(p \wedge q)$  and both inferential schemata are available to  $S$ :  $ce(B(\phi \wedge \psi)) = B\phi$  and  $di(B\phi) = B(\phi \vee \chi)$ <sup>88</sup>. No inferential schema that leads directly from  $B(p \wedge q)$  to  $B(p \vee r)$ , however, is available to  $S$ . Does  $S$  know how to perform an inference from  $B(p \wedge q)$  to  $B(p \vee r)$ ? Well, he knows how to perform an inference from  $B(p \wedge q)$  to  $B(p)$ , because  $ce$  is available to him, and he knows how to perform an inference from  $B(p)$  to  $B(p \vee r)$ , because  $di$  is available to him as well. So it would appear that  $S$  knows how to perform an inference from  $B(p \wedge q)$  to  $B(p \vee r)$ , even if he is not able to do so in a single step. It is not clear, though, that  $S$  knows how to perform such an inference according to (KH\*). Up to this point, we have said nothing about  *$i$ -schemata* that take the output of further  *$i$ -schemata* as input. Given our current theoretical resources, we cannot decide if  $S$  knows how to perform an inference from  $B(p \wedge q)$  to  $B(p \vee r)$ .

To fix this wrinkle, let us deal with a further issue about inferential schemata that we have been postponing — now is the appropriate time to do this. We have been working with examples of *first-order* inferential schemata and, although we did not make that explicit, all the inferential schemata presented in our examples up to this point are first-order inferential schemata. A *first-order* inferential schema is one that takes a set of doxastic attitudes that have already been formed as input. A *second-order* inferential schema is one that contains the output of a first-order inferential schema as input<sup>89</sup>. A *third-order* inferential schema is one that takes the output of a second-order inferential schema as input, and so on. In general, an inferential schema of order  $n$  is one that takes the output of an inferential schema of order  $(n - 1)$  as input.

---

<sup>87</sup>In the example that follows, we use the language of propositional logic ( $\mathcal{PL}$ ). In our language  $\mathcal{PL}$  the constants  $p, q, r, \dots$  are atomic formulas and the variables  $\phi, \psi, \chi, \dots$  are variables for atomic formulas.

<sup>88</sup>‘ $ce$ ’ stands for ‘conjunction elimination’ and ‘ $di$ ’ stands for ‘disjunction introduction’. We name these inferential schemata after known derivation rules from propositional logic, but we do so only for the purposes of making it easy to remember their structure.

<sup>89</sup>Maybe *not only* the output of a first-order inferential schema, though: its input can be a blend of doxastic attitudes that have already been formed with an output of a first-order inferential schema. The same applies to inferential schemata of higher orders.

---

So,  $S$  may know how to perform an inference from a set of reasons of the type  $R$  to an inferential belief of the type  $B\phi$  when a first order *i-schema*  $\alpha_1(R) = B\phi$  is available to  $S$ , when a second-order *i-schema*  $\alpha_2(\alpha_1(R)) = B\phi$  is available to  $S$ , when a third-order *i-schema*  $\alpha_3(\alpha_2(\alpha_1(R))) = B\phi$  is available to  $S$ , and so on. In general, an inferential schema of order  $n$ ,  $\alpha_n(\alpha_{n-1}(\dots(\alpha_1(R)))) = B\phi$ , may be available to  $S$ .

For the sake of simplicity, let us just talk about inferential schemata *of any order* and use our simpler notation ' $\alpha(R) = B\phi$ '. Let us keep in mind, however, that an inferential schema  $\alpha(R) = B\phi$  of *any order* may be a first-order *i-schema* that takes  $R$  as input, or a second-order *i-schema* whose input contains the output of a first-order *i-schema* that takes  $R$  as input, or a third-order *i-schema*... and so on. That would give us the following:

(KH)  $S$  knows how to perform an inference from  $R$  to  $B\sigma$  when and only when there is an *optimal i-schema*  $\alpha(T) = B\phi$  of *any order* available to  $S$  such that  $R = si_n(T)$  and  $B\sigma = si_n(B\phi)$ , for some  $n$ .

We can handle the case presented above using (KH). Although there is no first-order inferential schema  $\alpha_1(B(\phi \wedge \psi)) = B(\phi \vee \chi)$  available to  $S$ , there is a second-order inferential schema  $\alpha_2(\alpha_1(B(\phi \wedge \psi))) = B(\phi \vee \chi)$  available to  $S$ . The relevant inferential schema consists in the application of *di* to the output of *ce*:  $di(ce(B(\phi \wedge \psi))) = B(\phi \vee \chi)$ . When *ce* takes  $B(p \wedge q)$  as input it returns  $Bp$  as output, and when *di* takes  $Bp$  as input it returns  $B(p \vee r)$  as output. That means that  $S$  knows how to perform an inference from  $B(p \wedge q)$  to  $B(p \vee r)$  after all.

At this point one might think that (KH) makes it *too easy* for someone to know how to perform an inference — inferential paths that we would otherwise regard as too complex to dignify a reasoner as being able to perform are assumed to indicate that the relevant reasoner knows how to reason in such a complex way. But that is a mistake. It is not *that easy* for a higher-order inferential schema to be available to someone — and we are using the notion of *availability* to explicate knowledge of how to reason. We invite the reader to go back and check (A) for himself/herself again: in order for an inferential schema (of any order) to be available to someone, two counterfactual conditions, (A1) and (A2), must be satisfied across a certain range of situations. It turns out, then, that (KH) does not allow 'easy' attributions of knowledge of how to reason.

Now let us consider Nocond's case again (the one we presented in *Chapter 1*) and use our developments to diagnose it. We saw that at time  $t$  Nocond does not know how to infer that  $((q \rightarrow \neg p) \rightarrow \neg q)$  from his belief in  $p$ . According to (KH), that means that no inferential schema (of any order)  $\alpha(B\phi) = B(((\psi \rightarrow \neg\phi) \rightarrow \neg\psi))$  is available to him at  $t$ . If we were to ask him if  $p$  gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$  he would probably answer:

‘I have no idea’, or something like that (Nocond would not pass the test (A1)). Likewise, if Nocond were stimulated to find out if  $((q \rightarrow \neg p) \rightarrow \neg q)$  is true, given his available reason  $Bp$ , he would probably not manage to form a belief in such a proposition on the basis of  $Bp$  (Nocond would not pass the test (A2)). And even if Nocond were to answer ‘Yes’ in the first case and to form a belief in  $((q \rightarrow \neg p) \rightarrow \neg q)$  in the second one, that would not mean that the relevant inferential schema was available to him at  $t$ . The fact that one satisfies tests (A1) and (A2) in a *particular* situation does not entail that one would succeed in other situations, where further pre-inferential/inferential beliefs would constitute substitution instances of the input/output variables of the *i-schema* (the notion of *availability* present in (A) involves *universal quantification*). Similar points apply to Noind’s case. So (KH) seems to diagnose Nocond’s and Noind’s cases correctly.

What we fleshed out in (KH) is not supposed to be an *analysis* of what it is for an agent to know how to perform an inference. (KH) is just supposed to express a useful extensional equivalence (useful for our present purposes). It may appear that what we are offering here is an *anti-intellectualist*<sup>90</sup> account of the relevant procedural knowledge — one that aims to analyze procedural knowledge in terms of dispositions<sup>91</sup>. However, although we recognize that the notion of *availability of an i-schema* seems to denote a disposition, we are not committed to the claim that knowing how to perform an inference is just a matter of having certain dispositions.

To be sure, (KH) is compatible with *intellectualism* about procedural knowledge. Roughly, intellectualism says that *knowledge-how* is a type of *knowledge-that*. According to an influential intellectualist view, advanced by Stanley and Williamson (2001), for  $S$  to know how to  $A$  is for  $S$  to know, of some way  $w$ , that  $w$  is a way to  $A$ . Applied to the type of procedural knowledge we are discussing (i.e., knowledge-how to perform an inference), Stanley and Williamson’s theory would say that for  $S$  to know how to perform an inference  $I$  is for  $S$  to know, of some way  $w$ , that  $w$  is a way to perform  $I$ . Now, *it may be* that  $S$  knows, of some way  $w$ , that  $w$  is a way to perform  $I$  when and only when an inferential schema  $\alpha(R) = B\phi$  is available to  $S$  such that the pre-inferential beliefs in  $I$  are substitution instances of  $R$  and the inferential belief in  $I$  is a substitution instance (uniform with the first one) of  $B\phi$ . In this case, having an *i-schema* available would be equivalent to knowing a certain proposition to be true<sup>92</sup>. So we take it that (KH)

<sup>90</sup>See Fantl (2012) for a distinction between intellectualism, moderate anti-intellectualism and radical anti-intellectualism about knowledge-how in general.

<sup>91</sup>See Ryle (1945-1946) for the classical anti-intellectualist account of *knowledge-how* in terms of *dispositions to behave in a certain way*.

<sup>92</sup>We are not saying that *this is the case* — we are just saying that, as far as our theory goes, that possibility is left open. To be sure, one would need to argue in favor of the equivalence thesis presented above. We do not think that any such argument will succeed, however. But explaining why we think so is beyond the purposes of the present work. Some critics of Stanley and Williamson’s theory are Noë (2005),

---

is neutral on the intellectualism/anti-intellectualism debate about procedural knowledge. And although we are going to make an anti-intellectualist move (a '*Rylean move*') against one particular strategy for dealing with Nozick's and Nozick's types of cases in the next chapter, that will not imply, again, that (KH) is an anti-intellectualist thesis.

# Chapter 3

## *Ex ante rationality*

In this chapter we will explicate the notion of *ex ante* rationality (or *ex ante* justification). We aim to show that the strategy we use to deal with the cases that are problematic for (PJ) is better than a further strategy: the ‘*adding beliefs*’ strategy. By doing so we make a ‘Rylean move’, arguing in a way that is similar to the way Ryle argued against an intellectualist theory about knowledge-how. Finally, we show how our theory differs from similar theories already advanced in the literature (by Goldman and Turri) and we use our developments from *Chapter 2* to explicate *ex post* rationality as well.

### 3.1 *Ex ante* rationality explicated (a Rylean move)

We saw in *Chapter 1* that (PJ) — the thesis that *ex ante* rationality is solely a function of available reasons — seems to be falsified by cases like Nocond’s and Noind’s:

#### **Nocond’s case:**

Nocond rationally believes that  $p$  at  $t$ . As one can check through basic propositional logic,  $p$  entails  $((q \rightarrow \neg p) \rightarrow \neg q)$ . Unfortunately, Nocond is not able to infer that  $((q \rightarrow \neg p) \rightarrow \neg q)$  from his belief that  $p$  — he does not know how to perform this kind of inference. Further, Nocond has no reasons to disbelieve or doubt that  $((q \rightarrow \neg p) \rightarrow \neg q)$ , and he has no other reasons for believing that proposition (such as the testimony from someone else).

#### **Noind’s case:**

Noind rationally believes that *only 0.01% of the Fs that are Hs are also Gs* and that *a is an F and an H*. The set of propositions  $\{\textit{only 0.01\% of the Fs that are Hs are also Gs}, \textit{a is an F and an H}\}$  gives inductive support to *a is not G*. However, Noind is not able to infer that *a is not G* from his beliefs that *only 0.01% of the Fs that are Hs are also Gs* and that *a is an F and an H*. He does not know how to perform this kind of inference. Further, Noind has no

---

reasons to disbelieve or doubt that *a is not G*, and he has no further reasons for believing that proposition.

We have a pretty strong intuition that forming what we have called ‘unreachable beliefs’ is not rational for a doxastic agent. The belief that  $((q \rightarrow \neg p) \rightarrow \neg q)$  is unreachable to Nocond at  $t$ , and the belief that *a is not G* is unreachable to Noind at  $t$ . We also saw that being inferentially justified in believing  $\phi$  in virtue of reasons  $R$  is a matter of *being in a position* to infer that  $\phi$  from  $R$  in the right way, and just having reasons  $R$  whose contents give support to  $\phi$  do not put one in such a position. As we put it in *Chapter 1*, in order for a belief in  $\phi$  to be rational or inferentially justified for  $S$  it is not sufficient that  $S$  be *propositionally justified* in believing  $\phi$ .

But it is not our purpose just to show that the ‘*ex ante* rationality as propositional justification’ account is wrong: we have to find a suitable (informative) substitute for it. Now we can try to fix (PJ). The new set of conditions is supposed to entail that believing  $((q \rightarrow \neg p) \rightarrow \neg q)$  is not rational or inferentially justified for Nocond, and that believing *a is not G* is not rational or inferentially justified for Noind. So here is our (unsurprising) proposal:

- (IJ) Believing  $\phi$  is rational or inferentially justified for  $S$  at  $t$  if and only if (1) there is a set of undefeated reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  available to  $S$  at  $t$  such that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  and (2)  $S$  knows how to perform an inference from  $R$  to  $B\phi$  at  $t$ .

Our thesis successfully deal with both, Nocond’s and Noind’s case: having those beliefs is not rational for them, because they do not know how to form those beliefs on the basis of their reasons. Given our explication of the notion of *knowledge of how to perform an inference*, that would be equivalent to say that having those beliefs is not rational for Nocond and Noind because there is no *optimal* inferential schema available to them such that it returns the relevant beliefs as output given their reasons as input.

Also, (IJ) has a further virtue. It is often stressed that principles of human rationality should not overlook cognitive or computational limitations of humans’ cognitive capacities<sup>93</sup>. (IJ) offers a direct way of addressing this point — we should not regard the beliefs that  $S$  is not able to form as rational or justified for her. Of course, this type of consideration is usually made in a context where epistemic *obligations* are being considered<sup>94</sup>, but it applies to epistemic *permissions* as well. If one is not able to form a belief in a certain way (on the basis of the right reasons in the right way) then one has no epistemic

---

<sup>93</sup>See Harman (1986, p. 9-10), Neta (2008, p. 114), Pollock and Cruz (1999, p. 165) for different ways of making this point.

<sup>94</sup>About the *ought implies can* principle in the theory of epistemic normativity, see Wedgwood (2013).

permission to form that belief, for it is not epistemically permitted to form a belief in a non-competent way (by way of guessing, committing fallacies, etc).

So, (IJ) maintains the condition of support relation already fleshed out in (PJ) and adds a condition of procedural knowledge. In order for one to be justified in believing  $\phi$  it is not sufficient that one has reasons whose contents give support to  $\phi$  — one needs to know how to use those reasons to form a belief in  $\phi$ . The reasons that one does not know how to reason with do not ‘epistemize’ beliefs for one. To borrow a term used by Goldman<sup>95</sup> again, (IJ) is a ‘two-component’ account of rationality or inferential justification, only it is not about *ex post* rationality (or *doxastic* justification) — it is rather about *ex ante* rationality<sup>96</sup>. When it comes to *ex post* rationality, Goldman describes a two-component theory as one that takes into account both a certain *fit relation* that is supposed to hold between pre-inferential and inferential doxastic attitudes (condition present in *evidentialist* theories of justification) and the reliability of a certain type of cognitive process (condition present in *reliabilist* theories of justification)<sup>97</sup>. When it comes to *ex ante* rationality, so we suggest, a two-component theory is one that takes into account both a certain *fit relation* that is supposed to hold between inferential and pre-inferential doxastic attitudes<sup>98</sup> and the possession of a certain *cognitive ability* (in the case of (IJ), knowledge of how to reason in a certain way). So, a two-component theory of *ex ante* rationality contrasts with any theory for which *ex ante* rationality is solely a function of available reasons.

Now let us compare some evidentialist theories of *ex ante* justification with a two-component theory like ours. We saw before that (PJ) looks like an evidentialist theory of inferential justification, of the type defended by Conee and Feldman (1985, p. 83):

- (E) Doxastic attitude  $D$  toward proposition  $p$  is epistemically justified for  $S$  at  $t$  if and only if having  $D$  toward  $p$  fits the evidence  $S$  has at  $t$ <sup>99</sup>.

If we interpret (PJ) as an instantiation of an evidentialist theory of this type, it turns out that its particularity is taking the fit relation mentioned in Conee and Feldman’s thesis to be analyzed in terms of a support relation among propositions: where  $R$  is a set

<sup>95</sup>See Goldman (2011, p. 132). We are using the pagination of the reprinted version.

<sup>96</sup>We will consider a process-reliabilist theory of *ex ante* rationality suggested by Goldman in the next section (3.2). We refrain from assuming that our theory is a process-reliabilist theory of *ex ante* rationality because, as we saw in *Chapter 2*, there is more than one interpretation about the nature of inferential schemata and a reliabilist interpretation is just one among them.

<sup>97</sup>See Goldman (2011: 132-133).

<sup>98</sup>In the case of (IJ) the fit relation could be analyzed in terms of a support relation, or lack thereof, between the contents of the pre-inferential doxastic attitudes and the contents of the inferential ones — see below.

<sup>99</sup>We will occupy ourselves, again, only with *inferential* justification. Accordingly, the reading of (E) we are interested in is one where the evidence is a set of doxastic attitudes available to  $S$ . Whenever we mention an ‘evidentialist account of *ex ante* rationality’ we mean a theory of this type.

of reasons, believing  $\phi$  fits  $R$  when and only when  $[R]$  gives support to  $\phi$ , and suspending judgment about  $\phi$  fits  $R$  when and only when  $[R]$  gives support neither to  $\phi$  nor to  $\neg\phi$ . We have been arguing that (PJ) is false — although it states a necessary condition for *ex ante* rationality it does not state a sufficient condition for *ex ante* rationality. But by defending a two-component theory of *ex ante* rationality we are implying not only that (PJ) is false: other evidentialist accounts of type (E) are false as well. Some evidentialist account of that type, however, may also be presented as a solution to the problem of *unreachable beliefs* that we faced. So, we have to explain why a two-component theory such as ours does better than other evidentialist theories that also purport to fix the problem with (PJ).

Consider different ways an evidentialist could (purportedly) deal with Nocond's and Noind's types of cases. First, consider again a strategy that consists in analyzing the *fit relation* in terms of a *support relation* among propositions. This time, however, the support relation is interpreted in the following way:  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  only when at least one  $\psi_i \in \{\psi_1, \dots, \psi_n\}$  says that  $\{\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_n\}$  gives support to  $\phi$ . For example, the set of propositions  $\{(p \vee q), \neg p\}$  does not give support to  $q$ , but the set of propositions  $\{(p \vee q), \neg p, \{(p \vee q), \neg p\} \text{ gives support to } q\}$  does.

A diagnosis of Nocond's case based on this strategy would consist in stating: believing  $((q \rightarrow \neg p) \rightarrow \neg q)$  is not justified for Nocond at  $t$  because the content of Nocond's reasons at  $t$ ,  $p$ , does not give support to  $((q \rightarrow \neg p) \rightarrow \neg q)$  (contrary to what we assumed). If Nocond had a belief that  $p$  gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$ , however, it would be rational for him to believe  $((q \rightarrow \neg p) \rightarrow \neg q)$ , because that would mean that the content of his reasons gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$ . Similarly, a diagnosis of Noind's case based on the above strategy would be: believing that  $a$  is not  $G$  is not justified for Noind at  $t$  because the content of Noind's reasons at  $t$ ,  $\{\text{only } 0.01\% \text{ of the } Fs \text{ that are } Hs \text{ are also } Gs, a \text{ is an } F \text{ and an } H\}$ , does not give support to  $a$  is not  $G$ . If Noind had a belief to the effect that those propositions give support to  $a$  is not  $G$ , however, it would be rational for him to believe that  $a$  is not  $G$ , because that would mean that the content of his reasons gives support to that proposition.

But notice that, in whatever plausible interpretation of the support relation,  $p$  gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$  without the need of the additional information *stating that very fact*. The inclusion of a proposition about a support relation among the supporting propositions is not necessary for the relevant support relation to hold. If for  $\{\psi_1, \dots, \psi_n\}$  to give support to  $\phi$  it is necessary that at least one  $\psi_i \in \{\psi_1, \dots, \psi_n\}$  states that  $\{\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_n\}$  gives support to  $\phi$ , then only an *infinite* set of propositions can give support to a further proposition: we only need to reiterate the same requirement again and again. It is a fact, however, that *finite* sets of propositions give sup-



port to further propositions. So this strategy does not succeed because of its implausible implications about the support relation.

So maybe the way to go for the evidentialist here is not to offer an alternative interpretation of the *support relation* of that kind. Maybe the evidentialist would do better by letting the support relation be interpreted as we did and just work his way out by offering an alternative interpretation of the *fit relation*.

Following that direction, a second evidentialist strategy would be to hold that believing  $\phi$  fits reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  when there is a  $B\psi_i \in \{B\psi_1, \dots, B\psi_n\}$  such that  $\psi_i$  states that  $\{\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_n\}$  gives support to  $\phi$  (or something similar)<sup>100</sup> and, further, it is true that  $\{\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_n\}$  gives support to  $\phi$  (that is,  $B\psi_i$  is a belief in a true proposition). This is a clearly different strategy from the previous one: this proposal is not committed to an extraneous and implausible thesis about the support relation (in fact, it is compatible with any plausible account of support relation). In order to clarify the present evidentialist strategy, let us consider some examples:

**Amanda's case:**

Amanda (rationally) believes that *Claudio is a philosopher* and that *99% of philosophers are teachers*.

**Rachel's case:**

Rachel (rationally) believes that *Claudio is a philosopher*, that *99% of philosophers are teachers*, and that these facts *give support to* (or *are reliable indicators of the truth of*) the proposition *Claudio is a teacher*.

The present evidentialist thesis would say that, while believing that *Claudio is a teacher* does not fit Amanda's reasons, believing that proposition *does fit* Rachel's reasons. That is because Rachel believes that the set of propositions  $\{\textit{Claudio is a teacher}, \textit{99\% of philosophers are teachers}\}$  gives support to *Claudio is a teacher* (or something along these lines).

Finally, a third evidentialist strategy would consist in advancing the idea that *to have reasons* necessarily involves having beliefs in support relations (or in relations of reliable indication, etc). Notice the difference between the previous evidentialist strategy and the present one. According to the former one, Amanda has reasons to believe that *Claudio is a teacher*, but believing that proposition does not fit her reasons. According to the latter one, Amanda *does not* have reasons to believe that *Claudio is a teacher*. These

---

<sup>100</sup>Other propositions such as  $\{\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_n\}$  makes it true/probably true that  $\phi$  and  $\{\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_n\}$  reliably indicates the truth of  $\phi$  would do as well. We will use the term 'support relation' to describe the content of the relevant belief, but it should be understood as an umbrella term for cognate terms such as 'makes true/probably true', 'reliably indicates the truth of', etc.

strategies agree, however, that believing the proposition *Claudio is a teacher* is rational for Rachel but not for Amanda. While the first strategy explains this fact (if it is a fact) by pointing out that believing the proposition *Claudio is a teacher fits* Rachel's reasons but not Amanda's reasons, the second one explains it by pointing out that Rachel *has reasons* to believe that Claudio is a philosopher, while Amanda does not.

The details need not bother us now. There is a *shared implication* by these evidentialist views and the objection we are going to make will be about this very implication. Both views use this shared implication to explain why the epistemic status of unreachable beliefs is negative. If our objection succeeds, it will follow that neither evidentialist strategy solves the problem of unreachable beliefs.

In order to do this, let us first see how both evidentialist strategies explain the failure of justification in Nocond's and Noind's cases. Both accounts purport to show that what Nocond lacks is a belief in something like: *p gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$* , and that what Noind lacks is a belief in something like: *{only 0.01% of the Fs that are Hs are also Gs, a is an F and an H} gives support to a is not G*. So both views will ultimately explain the negative epistemic status of Nocond's and Noind's unreachable beliefs by pointing out that these subjects 'see no connection' between the contents of their reasons and the content of those unreachable beliefs — where 'seeing the connection' is something like having beliefs in the relevant support relations<sup>101</sup>. These evidentialist strategies are committed to the claim that by *adding beliefs* (beliefs about support relations, presumably cases of rational belief or knowledge) to these subjects' mental states we make the target beliefs justified for them and, therefore, that the absence of such beliefs explains why the target beliefs are not justified for those subjects in the original cases. Let us call this the '*adding beliefs*' solution to our problem. The idea can be put as follows:

(*AddBel*) Believing  $\phi$  is rational or inferentially justified for  $S$  at  $t$  if and only if there is a set of undefeated reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  available to  $S$  at  $t$  such that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  and  $S$  believes/rationally believes/knows that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ <sup>102</sup>.

Given that Nocond does not believe that *p gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$* , (*AddBel*) correctly entails that believing  $((q \rightarrow \neg p) \rightarrow \neg q)$  is not rational for him (same for Noind).

<sup>101</sup>When it comes to the objection we are about to make, it will not help to sophisticate the evidentialist strategy a little bit by saying that *to see* the relevant connection is *to have an intuition/to be acquainted with* the fact that the support relation holds.

<sup>102</sup>We use 'believes/rationally believes/knows' to leave open all of these choices to the proponent of this thesis. The point we are about to make is independent of the epistemic status that one requires for the belief in the relevant support relation. Also, notice that (*AddBel*) leaves open the possibility that  $S$ 's belief about the support relation is part of the set  $\{B\psi_1, \dots, B\psi_n\}$ .

So (*AddBel*) seems to solve the problem: it purportedly explains why it is not rational for one to form those beliefs that are unreachable to one. In what follows, we want to show that (*AddBel*) is not a solution to the problem of unreachable beliefs. Before doing so, however, let us explain what we *do not* aim to do.

First, we do not aim to show that (E) is false. (E) is a *general schema*, and one could think of our proposal as being one of interpreting the fit relation also in terms of the possession of cognitive abilities: believing  $\phi$  fits reasons  $R = \{B\psi_1, \dots, B\psi_n\}$  for  $S$  only when  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  and  $S$  knows how to reason from  $R$  to  $B\phi$ . In this case, the ‘two-component’ aspect of our theory would be totally embedded in the analysis or explication of the *fit relation* and we could be called ‘evidentialists’<sup>103</sup>. Of course, although in this case our theory could be regarded as an evidentialist theory it would still be in conflict with Conee and Feldman’s theory. That is because Conee and Feldman (2008, p. 83) subscribe to the following ‘strengthening’ of the general evidentialist thesis (we adapt Conee and Feldman’s thesis to the special case of inferential justification):

(SE) Necessarily, if  $S_1$  is inferentially justified in believing  $\phi$ , and  $R$  is the set of reasons available to  $S_1$  then (1) on balance the content of  $R$  gives support to  $\phi$ , and (2) if  $R$  is the set of reasons available to  $S_2$ , then  $S_2$  is justified in believing  $\phi$ .

Our thesis (IJ) entails the falsity of (SE): it is possible for two subjects to have exactly the same set of reasons  $R$  and, still, believing  $\phi$  is justified for one but not for the other. According to our theory this is so because it is possible that one of these subjects knows how to infer that  $\phi$  from  $R$  while the other does not.

Second, we do not want to suggest that Conee and Feldman subscribe to the evidentialist strategies we have been talking about. It may appear that Conee and Feldman (2001, 2008) are committed to (*AddBel*). We are not sure about this, however<sup>104</sup>. It might be useful to consider an example that Conee and Feldman themselves deal with (2008, p. 85). Suppose an investigator knows that  $p$ : *The fingerprints at the scene of the crime have characteristics X, Y, Z*. Further, the truth of  $p$  is strong objective evidence for the truth of  $q$ : *Lefty was at the scene of the crime*<sup>105</sup>. Given that much, Conee and

<sup>103</sup>A similar point applies if one advances the idea that *having reasons* requires *knowing how to use them* in at least some inferences: in this case the ‘two-component’ aspect of our theory would be embedded in a theory about the nature of reasons.

<sup>104</sup>Conee and Feldman (2001) explicitly recognize that both options are open to internalists: to deny that it is always necessary for  $S$  to be justified in believing  $\phi$  that  $S$  has beliefs about the relevant support relations, and to affirm this. The first option would be open because internalists could hold that we do not need to have beliefs about certain elementary logical connections (such as the connection between a conjunction and one of its conjuncts) in order for the corresponding inferential beliefs to be justified. The idea would be that propositions related in this way would have a ‘primitive or basic epistemic connection’ (2001, p. 252).

<sup>105</sup>Where we use the term ‘objective evidence’ Conee and Feldman use the term ‘scientific evidence’:  $p$  is scientific evidence for  $q$  when the fact that  $p$  is publicly available and  $p$  reliably indicates the truth of

Feldman claim that the investigator *does not* have justifying evidence to believe that  $q$  if he is unaware of the connection between  $p$  and  $q$ . In order to have justifying evidence the investigator needs to be aware of the truth–connection between the characteristics of the fingerprints in the scene of the crime and Lefty’s presence. Apparently, then, they seem to subscribe to something similar to (*AddBel*), at least if we assume that being aware of (the truth of)  $\phi$  entails believing  $\phi$ . Of course, the belief about the truth–connection does not need not be an ‘explicit’ one. As Conee and Feldman themselves emphasize, in order for the investigator to be justified in believing  $q$  he does not need to ‘formulate the thought’ (2008, p. 85) about the truth–connection between  $p$  and  $q$ . So the relevant belief could in some sense be an ‘implicit’ one.

It is not clear, however, that the fact that Conee and Feldman deal with these cases in the way they do commits them to (*AddBel*). Notice that (*AddBel*) makes two requirements for a belief in  $\phi$  to be inferentially justified for  $S$  in virtue of  $S$ ’s reasons  $R = \{B\psi_1, \dots, B\psi_n\}$ : that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$  and that  $S$  believes that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ . But it is controversial that the proposition *The fingerprints at the scene of the crime have characteristics X, Y, Z* actually gives support to the proposition *Lefty was at the scene of the crime*. It is not controversial, on the other hand, that the proposition *The fingerprints at the scene of the crime have characteristics X, Y, Z* together with some further proposition like *Lefty’s fingerprints are identified by characteristics X, Y, Z* gives support to the proposition *Lefty was at the scene of the crime*. The investigator’s case may just be a case of *insufficient evidence*: the investigator lacks more information, in the sense that the evidence available to him does not by itself confirm that *Lefty was at the scene of the crime*. So, it is not clear that cases such as this one motivate (*AddBel*) after all.

Third, we do not aim to raise an objection to (*AddBel*) that is similar to an objection that Goldman raises against Conee and Feldman’s internalist view about justification. Goldman (2009, pp. 103-104) objects to the idea of ‘internalizing’ the support relation — that is, the idea that relations of support (or other similar relations) make an epistemic difference only when agents get to believe them to hold — by pointing out that people do not need to have beliefs about support relations in order to have justified beliefs that ‘rely’ on these support relations. That would be too strong a requirement. We do not subscribe to Goldman’s criticism because the evidentialist could hold that not only do explicit and actual beliefs make a contribution to the justification of further beliefs, but non–explicit or dispositional beliefs also do. Consider (*AddBel*) again. As it is stated, it is

---

$q$ . This sense of ‘evidence’ is present in ordinary language when we say, for example, that the presence of Koplik spots in Amanda’s body (or the fact that Amanda has Koplik spots) is evidence that Amanda has measles. See Kelly (2006) for discussion.

open the possibility that the mentioned beliefs about support relations are non-explicit or dispositional beliefs. Further, maybe non-explicit or dispositional beliefs about support relations also count as part of one's set of available reasons, in such a way that it is not hard or extraordinary for doxastic agents to satisfy this further constraint. So, we find Goldman's objection inconclusive.

Now that we have made clear what we *do not* aim to do, let us make our point against (*AddBel*). Here we will make a 'Rylean move' — one that has some similarity to the way Ryle argued against a certain intellectualist view about *knowledge-how*. Ryle (1945, p. 6) argues against the thesis that knowing how to reason 'is analysable into the knowledge or supposal of some propositions': adding information to one's cognition does not make one smarter, more skilled or more competent when it comes to knowing how to perform inferences. He asks us to consider the following case (Ryle 1945, p. 6):

"A pupil fails to follow an argument. He understands the premises and he understands the conclusion. But he fails to see that the conclusion follows from the premises. The teacher thinks him rather dull but tries to help. So he tells him that there is an ulterior proposition which he has not considered, namely, that *if these premises are true, the conclusion is true*. The pupil understands this and dutifully recites it alongside the premises, and still fails to see that the conclusion follows from the premises even when accompanied by the assertion that these premisses entail this conclusion. So a second hypothetical proposition is added to his store, namely, that *the conclusion is true if the premises are true* as well as the first hypothetical proposition that *if the premisses are true the conclusion is true*. And still the pupil fails to see. And so on forever. He considers reasons, but he fails to reason."

The point here is to show that the intellectualist thesis that *knowledge-how* is analysable into *knowledge-that* is false. One could give all the information in the world to a bad reasoner and, still, he could fail to perform the relevant inferences<sup>106</sup>. It is not our purpose to argue against intellectualism about *knowledge-how* and neither to show that this concept is 'logically prior' (Ryle 1945, p. 5) to the concept of *knowledge-that*. To be sure, there is more than one way the intellectualist could tell us her story about *knowledge-how* being analyzed as *knowledge-that*<sup>107</sup>. An intellectualist may not require the presence of *that* kind of *knowledge-that* (knowledge that something follows from something else, or knowledge that something gains support from something else) in order for *knowledge-how* to take place, but the presence of some other kind of *knowledge-that*.

<sup>106</sup>Ryle identifies this problem with Lewis Carroll's (1895) puzzle in 'What the Tortoise Said to Achilles', where the tortoise pretends to be such a dull reasoner.

<sup>107</sup>See Stanley (2011).

---

We will use, however, cases that are structurally similar to the ones presented by Ryle to argue that (*AddBel*) does not solve the problem of unreachable beliefs and, therefore, that those alternative evidentialist ‘solutions’ to the problem of unreachable beliefs fail. Consider the following:

**Nocond’s new case:**

At  $t$  Nocond rationally believes that  $p$  and that  $p$  entails  $((q \rightarrow \neg p) \rightarrow \neg q)$ . Unfortunately, however, Nocond is not able to infer that  $((q \rightarrow \neg p) \rightarrow \neg q)$  from his new set of reasons  $\{p, p \text{ entails } ((q \rightarrow \neg p) \rightarrow \neg q)\}$  — he does not know how to perform this kind of inference either. Further, Nocond has no reasons to disbelieve or doubt that  $((q \rightarrow \neg p) \rightarrow \neg q)$ , and he has no other reasons for believing that proposition.

**Noind’s new case:**

At  $t$  Noind rationally believes that only *only 0.01% of the  $F$ s that are  $H$ s are also  $G$ s*, that  $a$  is an  $F$  and an  $H$  and that  $\{\text{only 0.01\% of the } F\text{s that are } H\text{s are also } G\text{s, } a \text{ is an } F \text{ and an } H\}$  gives support to  $a$  is not  $G$ . However, Noind is not able to infer that  $a$  is not  $G$  from his new set of reasons. He does not know how to perform this kind of inference either. Further, Noind has no reasons to disbelieve or suspend judgment about  $a$  is not  $G$ , and he has no further reasons for believing that proposition.

In the original case, given that Nocond did not believe that  $p$  gives support to  $((q \rightarrow \neg p) \rightarrow \neg q)$ , (*AddBel*) correctly entailed that believing  $((q \rightarrow \neg p) \rightarrow \neg q)$  was not rational for him. In *Nocond’s new case*, however, Nocond *does believe* the relevant support proposition (he believes an *entailment* relation to hold) — but he still does not know how to infer that  $((q \rightarrow \neg p) \rightarrow \neg q)$  is true from his reasons. It turns out that, in this new case, Nocond has more reasons than before to believe  $((q \rightarrow \neg p) \rightarrow \neg q)$  but, again, he is not able to form a belief in such a proposition on the basis of his evidence. So, (*AddBel*) only postpones the solution to our problem. We could also have *Nocond’s new new case*, *Nocond’s new new new case*, and so on. However many *strata* of beliefs about support relations we add to Nocond’s set of available reasons, the same problem regarding his lack of procedural knowledge recurs, putting into question the claim that it is rational for Nocond to form the belief that is unreachable to him. Similar points apply to Noind’s case.

In general, *having more reasons* does not guarantee *being able to reason*. The point can be made with full generality as follows. Let ‘ $s$ ’ be a meta-variable for propositions of the type:  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ . Now let us assume that, at  $t$ ,  $S$  has reasons

$\{B\psi_1, \dots, B\psi_n\}$  such that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ . Further,  $S$  does not know how to reason from  $\{B\psi_1, \dots, B\psi_n\}$  to  $B\phi$  at  $t$ . A diagnosis based on (*AddBel*) says that believing  $\phi$  is not rational for  $S$  at  $t$  because she does not believe that  $\phi$  gets support from her evidence. So let us add to  $S$ 's reasons a belief in a proposition  $s$ , saying that  $\{\psi_1, \dots, \psi_n\}$  gives support to  $\phi$ . In this new version,  $S$ 's reason at  $t$  are  $\{B\psi_1, \dots, B\psi_n, Bs\}$ . However, it may as well be that  $S$  does not know how to reason from  $\{B\psi_1, \dots, B\psi_n, Bs\}$  to  $B\phi$  at  $t$ . Again, it would appear that we are not entitled to say that believing  $\phi$  is rational or inferentially justified for  $S$  at  $t$  —  $S$  is not able to believe  $\phi$  on the basis of his reasons in this new version as well. A further diagnosis based on (*AddBel*), then, can be proposed: believing  $\phi$  is not rational for  $S$  at  $t$  because she does not believe that  $\phi$  gets support from  $\{\psi_1, \dots, \psi_n, s\}$ . So let us add to  $S$ 's reasons a belief in a proposition  $s^*$ , saying that  $\{\psi_1, \dots, \psi_n, s\}$  gives support to  $\phi$ . Again, it may as well be that  $S$  does not know how to reason from  $\{B\psi_1, \dots, B\psi_n, Bs, Bs^*\}$  to  $B\phi$  at  $t$ . And so on, *ad infinitum*.

Therefore, adding beliefs about support relations to the set of available reasons does not solve the problem we began with: to find a further condition (besides the one fleshed out in (PJ)) for a belief to be inferentially justified for someone — a condition that explains why forming unreachable beliefs is not rational for one. Such a condition needs to take into account the fact that, in order for a belief to be inferentially justified for  $S$  in virtue of  $S$ 's reasons  $R$ , she ( $S$ ) needs to be in a position to infer that  $\phi$  from  $R$  in the right way. Adding beliefs about support relations does not do the job: we have an agent with more reasons, but she still may not know how to reason. From here we conclude that those evidentialist strategies do not work. The relevant evidentialist theories are committed to (*AddBel*), and they purport to solve the problem of unreachable beliefs by using that thesis. But (*AddBel*) does not solve the problem of unreachable beliefs. A two-component theory is our best choice to deal with the types of cases we have been considering. Now let us consider other two-component theories and compare them to our own.

### 3.2 Similar theories

Our theory is certainly not the first one to bring the issue of procedural knowledge (*knowledge of how to reason*) into the theory of epistemic justification. For example, in more than one place (1995, 1999) John Pollock makes the point that epistemic norms of justification are descriptive of our knowledge of how to cognize. Among these epistemic norms there are the ones that describe right ways of reasoning: norms for inferential justification. Pollock (1995) calls this the ‘procedural concept of epistemic justification’.

More importantly, there are further theories that are also supposed to deal with the

problem of unreachable beliefs — theories that are quite similar to the one we are endorsing here (that is, there are further ‘two-component’ theories of *ex ante* rationality). Nevertheless, we aim to present a counterexample to these theories that is not a counterexample to our (IJ), at least when the relevant procedural knowledge mentioned in (IJ) is interpreted as we did in (KH) (see *Chapter 2, Section 2.5*).

In order to distinguish our own theory from the ones we have in mind, notice that our thesis (IJ) plus our explication of *knowledge of how to reason* differs from (and does not entail) the following thesis:

(*Would-J*) If believing  $\phi$  is inferentially justified for  $S$  at  $t$  (in virtue of reasons  $R$  and *optimal* inferential schema  $\alpha$ ) then, if at  $t$   $S$  were to believe that  $\phi$  on the basis of  $R$  as a result of an instantiation of  $\alpha$ ,  $S$ 's belief that  $\phi$  would thereby be *doxastically justified*.

So, suppose that there is a set of undefeated reasons  $R$  available to  $S$  at  $t$  such that they are good reasons for  $S$  to believe  $\phi$  and, further, there is an *optimal* inferential schema  $\alpha(T) = B\psi$  available to  $S$  at  $t$  such that  $R = si_n(T)$  and  $B\phi = si_n(B\psi)$  (that is,  $R$  is a substitution instance of the input-variable of  $\alpha$  and  $B\phi$  is a substitution instance of the output-variable of  $\alpha$ ). According to (IJ), believing that  $\phi$  is inferentially justified for  $S$  at  $t$ . Moreover, (*Would-J*) entails that if  $S$  were to form a belief in  $\phi$  on the basis of  $R$  as a result of an instantiation of  $\alpha$  (if  $S$  were to form a belief in  $\phi$  on the basis of  $R$  in the right way),  $S$ 's belief that  $\phi$  would thereby be *doxastically justified*.

We saw before (*Chapter 1*) that, in order for  $S$  to be justified in believing  $\phi$  in virtue of reasons  $R$ ,  $S$  needs to be in a position to infer that  $\phi$  from  $R$  in the right way (and for one to be in a position to infer that  $\phi$  on the basis of  $R$  in the right way one needs to know how to infer that  $\phi$  from one's reasons  $R$ ). But being in a position to infer that  $\phi$  from  $R$  in the right way does not entail *maintaining a rational belief* in  $\phi$  right after performing the relevant inference: it may be that after inferring  $\phi$  one notices that  $\phi$  is unacceptable (example given below). Accordingly, our (IJ) does not entail (*Would-J*).

We explicated the concept of *knowledge of how to reason* using the notion of *availability of inferential schemata*:

- (A) An *i*-schema  $\alpha(B\psi_1, \dots, B\psi_m) = B\phi$  is available to  $S$  at  $t$  if and only if, for any  $si_n(B\psi_1, \dots, B\psi_m) = \{B\chi_1, \dots, B\chi_m\}$  and  $si_n(B\phi) = B\sigma$ , if at  $t$   $S$ 's available reasons were  $\{B\chi_1, \dots, B\chi_m\}$  (and  $\{B\chi_1, \dots, B\chi_m\}$  alone), then:
- (1) If at  $t$   $S$  were properly prompted to determine if  $\{\chi_1, \dots, \chi_m\}$  (and  $\{\chi_1, \dots, \chi_m\}$  alone) gives support to  $\sigma$ , then  $S$  would believe that  $\{\chi_1, \dots, \chi_m\}$  gives support to  $\sigma$ ;



---

(2) If at  $t$   $S$  were properly prompted to find out if  $\sigma$  is true then, provided  $S$  does not take any member of  $\{\chi_1, \dots, \chi_m, \sigma\}$  to be in conflict with anything else she happens to believe as a result of considering these propositions,  $S$  would believe that  $\sigma$  on the basis of  $\{\chi_1, \dots, \chi_m\}$ .

*Availability of inferential schemata* is analyzed here by means of two counterfactuals restricted to worlds where one's reasons  $\{B\chi_1, \dots, B\chi_m\}$  are those doxastic attitudes that constitute a particular input to the relevant  $i$ -*schema* (that does not mean, of course, that inferential schemata are available only in those worlds). *But* it may still be the case that  $S$  forms new beliefs, conflicting ones, by considering the relevant propositions  $\chi_1, \dots, \chi_m, \sigma$  involved in the relevant inferential process. That is why we include that *proviso* at clause (A2). That means that there are some situations where one has reasons  $R$  and an inferential schema  $\alpha$  that outputs  $B\phi$  given  $R$  as input and, yet, one can reason from  $R$  (and even  $R$  alone) by instantiating  $\alpha$  but fail to *rationaly* maintain  $B\phi$ . So, it can be rational for  $S$  to believe  $\phi$  in virtue of  $S$ 's reasons  $R$  and in virtue of an *optimal* inferential schema  $\alpha$  available to  $S$  even though it is not the case that, if  $S$  were to believe  $\phi$  on the basis of  $R$  as a result of an instantiation of  $\alpha$ ,  $S$ 's belief in  $\phi$  would thereby be doxastically justified. We will consider such a case in a moment. For now, suffice it to say that by advancing (IJ) together with our theory about knowledge of how to reason we are not committed to (*Would-J*).

Something similar to (*Would-J*) was advanced by Turri (2010). Turri is in the business of arguing against what he calls 'the orthodox view of the relationship between propositional and doxastic justification' (2010, p. 312). Roughly, the orthodox view says that *if* believing  $\phi$  is propositionally justified for  $S$  in virtue of  $S$ 's reasons  $R$  and  $S$  forms a belief in  $\phi$  on the basis of  $R$ , then  $S$ 's belief that  $\phi$  is doxastically justified. Turri correctly points out that such a view is false: one can base one's beliefs in good reasons but do so in a *wrong way*<sup>108</sup>. Not only is the orthodox view false but, according to Turri, it misses a crucial point about the relationship between propositional and doxastic justification: 'The way in which the subject performs, the manner in which she makes use of her reasons, fundamentally determines whether her belief is doxastically justified' (2010, p. 318).  $S$  can have the best reasons in the world to believe that  $\phi$  and base her belief on those reasons but, still, if  $S$ 's inferential performance is not epistemically approvable her belief in  $\phi$  will not thereby be doxastically justified.

Motivated by these points<sup>109</sup>, Turri (2010, p. 320) advances something like the following thesis about the relationship between *ex ante* rationality and *ex post* rationality (or, in

---

<sup>108</sup>We presented one such example in *Section 2.4* of the previous chapter.

<sup>109</sup>We agree with these points. We disagree with the moral that Turri draws from them, as we will show in a moment.

the language used by Turri, between *propositional* justification and *doxastic* justification):

(TT) Necessarily, if believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$ , then believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$  *because*  $S$  possesses (at  $t$ ) at least one way of coming to form a doxastically justified belief in  $\phi$  on the basis of  $R$ <sup>110</sup>.

According to (TT), one's being justified in believing  $\phi$  is always explained by the fact that one has a way of coming to form a doxastically justified belief in  $\phi$ . Turri's thesis entails that, when it is justified for  $S$  to believe  $\phi$  in virtue of  $S$ 's reasons  $R$ ,  $S$  possesses at least one way of coming to form a doxastically justified belief that  $\phi$  on the basis of  $R$ . But we want to show that there are cases where believing  $\phi$  is inferentially justified for  $S$  in virtue of reasons  $R$  even though  $S$  possesses no way of coming to form a doxastically justified belief in  $\phi$  on the basis of  $R$ <sup>111</sup>. There are beliefs such that by forming them  $S$  will lose justification for holding them. Turri himself touches on the problem, but because he deals with a poor counterexample he says he is 'unpersuaded by such examples' (2010, p. 321). Here is a relevant case for us to consider<sup>112</sup>:

#### **Gottlob's case:**

Gottlob has been studying logic and making a lot of exercises. He masters all the rules of derivation that he has learnt in the logic class. Looking back, however, Gottlob notices that all long proofs he did in the past were incorrect. Gottlob does not happen to notice this, but the long proofs he did in the past were wrong just because he was distracted by the television while doing them (Gottlob does not remember that he made those exercises while watching television). At  $t$  Gottlob knows axioms  $A1-A3$  and has no reason to doubt them to be true. From  $A1-A3$  a certain theorem  $T$  follows, but in order to derive it one has to go through several steps. Gottlob knows how to infer that the theorem is true on the basis of his knowledge of axioms  $A1-A3$  (by performing a derivation from  $A1-A3$  to  $T$ ), but he does not perform the relevant inference

<sup>110</sup>'TT' stands for 'Turri's Thesis'. We have adapted Turri's thesis to the specific case of inferential justification. It should be emphasized here, again, that saying that something holds *in virtue of* something else does not imply that something holds *only in virtue of* something else. So, when the antecedent of (TT) is true it does not follow that believing  $\phi$  is justified for  $S$  at  $t$  *only* in virtue of  $S$ 's reasons  $R$ . Further, Turri's original formulation differs a little bit from ours: 'Necessarily, for all  $S$ ,  $p$ , and  $t$ , if  $p$  is propositionally justified for  $S$  at  $t$ , then  $p$  is propositionally justified for  $S$  at  $t$  because  $S$  currently possesses at least one means of coming to believe  $p$  such that, were  $S$  to believe  $p$  in one of those ways,  $S$ 's belief would thereby be doxastically justified' (Turri 2010, p. 320).

<sup>111</sup>This is not to say, however, that there are cases where believing  $\phi$  is inferentially justified for  $S$  in virtue of reasons  $R$  even though  $S$  does not know how to infer that  $\phi$  on the basis of  $R$ !

<sup>112</sup>Our example differs from Turri's in that it is not supposed to be one where the agent's evidence is 'destroyed' after forming the target belief.

---

at time  $t$ . Nevertheless, by deriving  $T$  from  $A1-A3$  (in the only way available to him) Gottlob would notice that the proof from  $A1-A3$  to  $T$  is a long one<sup>113</sup>.

Here is our diagnosis of Gottlob's case: At time  $t$  Gottlob *is justified* in believing  $T$ . To be sure, at  $t$  he has good, undefeated reasons for believing  $T$ , and he knows how to infer this theorem from his knowledge of the axioms  $A1-A3$ . However, by using his cognitive abilities to derive the theorem he would not form a doxastically justified belief in  $T$ . For, if Gottlob were to perform the relevant derivation he would gain reasons to suspend judgment about  $T$ , because he would notice that the derivation from  $A1-A3$  is a long one, and he knows that in the past he made a lot of mistakes in similar situations. In this case, it would not be rational or justified for him to believe  $T$ . Therefore, if he were to form a belief in  $T$ , it would not be a *doxastically justified* belief. Gottlob's case seems to be a counterexample to both, (*Would-J*) and Turri's thesis.

We anticipate at least five ways one could object to our counterexample. In what follows, we will address each of these objections (*Objections 1-5*) individually<sup>114</sup>.

*Objection 1:*

There is still *a way* by means of which Gottlob could justifiably believe  $T$  on the basis of his beliefs in  $A1-A3$  at time  $t$ . After all, Gottlob could slowly *recheck* the proof after doing it (or ask professional logicians if the proof is sound) and come to discover that the errors he made in the past (when doing long derivations) were merely circumstantial, etc.

*Reply to objection 1:*

The concept of a *way* must not be *that broad* such as to include gathering any further evidence in the future. Of course, *some* new evidence can be generated as a *byproduct* of actualizing a way of coming to form a certain belief (the most obvious type of byproduct evidence of this sort is evidence gathered via introspection or conscious monitoring of what is going on in one's mind). In our example, if Gottlob were to perform a derivation from  $A1-A3$  to  $T$ <sup>115</sup>, he would notice that such a derivation is a long one. This type of new evidence (*byproduct* evidence) can be included in the things that we need to take into account when we are judging if a certain belief would be justified if one were to form it by using a *way* available to one. But other types of evidence (pieces of evidence that are not *byproduct* evidence) should not be included among these things. In Gottlob's case,

---

<sup>113</sup>This example is the result of several private conversational epicycles we had with John Turri. We are very thankful to him for all his objections and suggestions.

<sup>114</sup>We thank Tristan Haze, Tiegüie Vieira and Murali for useful feedback about this counterexample at our website, URL = <http://fsopho.wordpress.com/2014/01/03/a-counterexample-to-turris-thesis-about-justification/>.

<sup>115</sup>It does not matter for the present purposes if Gottlob is to do the derivation 'mentally' or to write it down on a piece of paper.

rechecking the proof (or asking professional logicians if the proof is sound) and coming to discover that his errors in the past (when doing long derivations) were merely circumstantial would give Gottlob the latter type of evidence — not byproduct evidence.

*Objection 2:*

There is still a way by means of which Gottlob could justifiably believe  $T$  on the basis of his beliefs in  $A1-A3$  at time  $t$ . After all, we are assuming that Gottlob *knows how* to infer that  $T$  is true from his beliefs in  $A1-A3$  at  $t$ . In Gottlob's situation, performing an inference from his beliefs in  $A1-A3$  to a belief in  $T$  (exercising one's *knowledge of how to reason*) still counts as a way of coming to form a *doxastically justified* belief in  $T$ .

*Reply to objection 2:*

That  $S$  knows how to reason from  $R$  to  $B\phi$  *does not guarantee* that, by exercising this procedural knowledge,  $S$  *will not* gain reasons to believe that  $S$ 's belief-forming process from  $R$  to  $B\phi$  is unreliable. We are assuming that Gottlob *knows* (remembers) that all long proofs he did in the past were incorrect and that by performing a derivation from  $A1-A3$  to  $T$  he would notice that *that* is a long derivation. By going through the relevant belief-forming process Gottlob would acquire reasons to form the higher-order belief that *that* belief-forming process is unreliable. So, *Objection 2* implies that one can form a doxastically justified belief by means of a certain process-type  $P$  (in this case, doing a logical proof from  $A1-A3$  to  $T$  in one's head, or something like that) and *recognize* that such a belief was generated by a process of type  $P$  while one rationally believes that process-type  $P$  has been 100% unreliable up to the time one forms the relevant belief. But if one rationally believes that process-type  $P$  has been 100% unreliable and one believes that one's belief in  $\phi$  was generated by a process of type  $P$  then one's belief in  $\phi$  is not justified (absent any reason to think that *this particular* instantiation of process-type  $P$  is somehow different from the others when it comes to its accuracy).

*Objection 3:*

Contrary to what we have assumed, believing  $T$  *is not justified* for Gottlob at  $t$ . If at time  $t$  a subject  $S$  possesses no way of coming to form a doxastically justified belief in  $\phi$  on the basis of  $R$  (just like Gottlob when it comes to theorem  $T$ ) then it cannot be the case that believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$ <sup>116</sup>.

<sup>116</sup>Someone might take this to be the very thesis we have been defending, but that is a mistake. The truth-conditions for the conditional: (If at  $t$   $S$  possesses no way of coming to form a doxastically justified belief in  $\phi$  on the basis of  $R$  then it cannot be the case that believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$ ) clearly differ from the truth-conditions for the conditional: (If at  $t$   $S$  is not able to form a belief in  $\phi$  on the basis of  $R$  then it cannot be the case that believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$ ). The latter does not require that one be able to form any *doxastically justified* belief in order for that belief to be justified for one — just that one has the ability to form it in a certain way.

---

*Reply to objection 3:*

This way of arguing against our counterexample is circular. The conditional (If at  $t$   $S$  possesses no way of coming to form a doxastically justified belief in  $\phi$  on the basis of  $R$  then it cannot be the case that believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$ ) is just the contrapositive of the claim that believing  $\phi$  is justified for  $S$  in virtue of  $S$ 's reasons  $R$  at  $t$  just in case  $S$  possesses (at  $t$ ) at least one way of coming to form a doxastically justified belief in  $\phi$  on the basis of  $R$ . Gottlob's example is supposed to be a counterexample to that very thesis. The defender of such a thesis cannot non-circularly object to our case by pointing out that believing the target proposition is not justified for Gottlob *because* he possesses no way of coming to justifiably believe it.

*Objection 4:*

Contrary to what we have assumed, believing  $T$  is *not justified* for Gottlob at  $t$ . That is because the epistemic status of the (non-actual) belief that  $T$  is *defeated* for Gottlob at time  $t$ .

*Reply to objection 4:*

It may appear that, since we are assuming that at  $t$  Gottlob knows that all long derivations he did in the past were incorrect, the epistemic status of the (non-actual) belief that  $T$  is defeated for him at  $t$ . But notice that Gottlob's knowledge of his poor track-record does not *by itself* defeat the epistemic status of that belief at time  $t$ . It is only when Gottlob's knowledge is put together with the information that *the proof from A1-A3 to T is a long one* that Gottlob gains a defeater for the belief in  $T$ . But at  $t$  Gottlob *does not have* that further information, because he did not make the proof at a time earlier than or equal to  $t$ . Therefore, the epistemic status of the (non-actual) belief that  $T$  is undefeated for Gottlob at time  $t$ .

*Objection 5:*

Contrary to what we have assumed, Gottlob *does not know how* to infer that  $T$  from his beliefs in  $A1-A3$ . His poor-track record on doing long derivations gives support to this negative judgment.

*Reply to objection 5:*

The errors Gottlob made in the past were just *performance* errors: he has the ability and competence to do long proofs of the relevant kind and, therefore, the ability and competence to form a belief in  $T$  on the basis of his beliefs in  $A1-A3$ . Unfortunately, all the long proofs he did in the past were made in unfavorable situations: he was distracted by the television, although he does not know that he was distracted (therefore, he does not know that his performance was jaundiced by this environmental factor — so he cannot

use that information to dismiss the evidence he has to believe that he is bad at doing long proofs). That one knows how to perform an inference is perfectly consistent with the fact that one never succeeded in performing it in the past. As pointed out by John Greco (2007, p. 61): ‘Actual track records can be the result of good luck rather than ability. Likewise, actual track records can be the result of *bad* luck rather than *lack* of ability’.

In addition to the five objections we dealt with above, one could make an ‘internal coherence’ question to us: Is it possible for Gottlob to know how to infer  $T$  from his beliefs in axioms  $A1-A3$  according to our own theory? (Our theory consists in the combination of an explication of *ex ante* rationality (IJ), an explication of *knowledge of how to reason* (KH), and an explication of what it is for an inferential schema to be available to someone (A)). Notice that, according to our theory, neither the fact that Gottlob went wrong in the past nor the fact that Gottlob would (rationally) refrain from forming a belief in  $T$  if he were to derive it from the axioms at  $t$  tells against the idea that Gottlob knows how to infer  $T$  from his beliefs in  $A1-A3$ . The theory we defend, (IJ) + (KH) + (A), allows us to attribute *knowledge of how to reason* to Gottlob because what matters for deciding if Gottlob has the relevant inferential ability are the counterfactual tests that we need to make on the following assumptions: that the only reasons available to Gottlob are his beliefs in  $A1-A3$  and that Gottlob is in ‘good shape’ (not distracted, awake, sober, etc). The former assumption excludes cases where Gottlob has knowledge of his bad track record in performing long proofs from the range of our counterfactuals. The second one excludes cases where Gottlob is distracted and (as a result) goes awry at making the relevant derivations from the range of our counterfactuals. As far as our theory goes, it is consistent with the assumption that Gottlob knows how to infer that  $T$  on the basis of his knowledge of the axioms  $A1-A3$ . So, we take it that Gottlob’s case is a successful counterexample to both (*Would-J*) and Turri’s thesis, but not to our own theory.

A similar but stronger thesis than Turri’s was proposed earlier by Goldman (1979, p. 21)<sup>117</sup>. Goldman’s proposal is that we can analyze propositional justification or ‘*ex ante* justifiedness’ in terms of doxastic justification or ‘*ex post* justifiedness’. The idea here is that believing  $\phi$  is justified for  $S$  at  $t$  if and only if there is a reliable belief-forming process  $P$  available to  $S$  at  $t$  such that the application of  $P$  to  $S$ ’s cognitive state at  $t$  would result in  $S$  having a doxastically justified belief that  $\phi$ <sup>118</sup>.

Call this the ‘reliabilist analysis of *ex ante* rationality’. Gottlob’s case constitute a counterexample to such analysis as well. There is no reliable inferential process available to

<sup>117</sup>Turri himself (2010, footnote 18) explicitly notes this.

<sup>118</sup>The present formulation of Goldman’s view is due to a suggestion given by Goldman himself in private conversation. We thank him for this.

Gottlob at  $t$  such that, when applied to his cognitive state at that time, it would generate a *doxastically justified* belief in  $T$ . That is because, by applying the relevant reliable process available to him (performing a derivation from  $A1-A3$  to  $T$ ) to his cognitive state at  $t$  and reaching a new cognitive state, Gottlob would notice that he performed a long proof and would thereby gain a defeater for his belief in  $T$  (assuming, again, that he has knowledge of his poor track-record when doing long derivations and that he does not know that he was distracted in the past).

Both Turri's and Goldman's theses would equally give a solution to the problem of unreachable beliefs, and both entail that being justified is not just a matter of having reasons. Consider Nocond's case again. There is no way or reliable process available to Nocond such that it would lead him to justifiably believe that  $((q \rightarrow \neg p) \rightarrow \neg q)$  on the basis of his reasons at  $t$ . So both theses say that believing  $((q \rightarrow \neg p) \rightarrow \neg q)$  is not rational for Nocond. However, Gottlob's case seems to be a counterexample to both Turri's and Goldman's theses and, therefore, we have a reason to reject these theses. On the other hand, (IJ) gives a solution to the problem of unreachable beliefs *and* is immune to such a counterexample.

### 3.3 A very brief section on *ex post* rationality

Having argued that our theory of *ex ante* rationality is superior to similar theories, we want to briefly present an explication of *ex post* rationality as well. Despite the considerations we made above about Turri's thesis we still suggest, as Turri does (2010), that there is a strong relationship between the state of having justification and the state of justifiably believing. Here is a plausible way of establishing such a strong relationship:

(DJ)  $S$  justifiably believes (or rationally believes) that  $\phi$  on the basis of  $R$  if and only if (i) believing  $\phi$  is inferentially justified for  $S$  in virtue of both, reasons  $R$  and an *optimal* inferential schema  $\alpha$ , and (ii)  $S$  believes  $\phi$  on the basis of  $R$  as a result of instantiating  $\alpha$ .

Notice that (DJ) takes into account Turri's observation that 'the way a subject makes use of his reasons' (Turri 2010, p. 315) matters to whether his belief is doxastically justified or not: if  $S$  forms her belief in  $\phi$  on the basis of  $R$  by instantiating a *non-optimal* inferential schema, then her belief is not doxastically justified. At the same time, it does not commit us to an explanatory or analytical divorce between the notion of *justifiably believing* (*ex post* rationality) and the notion of *having justification to believe* (*ex ante* rationality). There are a number of details that need to be filled out here. We already saw

---

what it is for a subject to instantiate an inferential schema in *Chapter 2, Section 2.4*. But we have given no precise account of the *basing relation* yet. This is a task for future work, though. By presenting (DJ) we just expect to show that a plausible explication of *ex post* rationality can be fleshed out on the basis of our previous developments and, therefore, that our investigation ‘has something to say’ about doxastic justification and, presumably, about knowledge as well (assuming that knowledge requires doxastic justification).



## *PART 2*

### *Modelling Epistemic Rationality*

# Chapter 4

## *Setting the stage again*

In this chapter we start thinking about how to formalize attributions of epistemic rationality. Further, we flesh out what we need to take into account when trying to offer a semantics for attributions of epistemic rationality. Certain *criteria of adequacy* for such a semantics are established.

### 4.1 Language

Let us start fleshing out a language by means of which we can formalize attributions of epistemic rationality. In order to have a model-theoretic semantics for attributions of epistemic rationality we need first to establish which types of formulas will be made true by our models. Let us call our language for attributions of epistemic rationality ‘ $\mathcal{E}$ ’. Our language  $\mathcal{E}$  will contain formulas of the following type: formulas with *doxastic operators* (a *belief* operator and a *doubt* operator) and formulas with *rationality operators*. Given that we attribute rationality to beliefs and doubts for certain subjects, the scope of the rationality operators will consist in formulas with the doxastic operators.

Since we are in this type of intensional setting, we will also need a language  $\mathcal{L}$  whose formulas will be in the scope of the doxastic operators. Here,  $\mathcal{L}$  is supposed to be the language that we use to describe the *contents* of doxastic attitudes. And just as we use a certain language to describe the contents of doxastic attitudes, we will need elements of a model for formulas in  $\mathcal{L}$  inside our models (see the next section and *Chapter 5*). So what we have to do is to index the language  $\mathcal{L}$  to the language  $\mathcal{E}$ :  $\mathcal{E}_{\mathcal{L}}$ . Our language  $\mathcal{E}_{\mathcal{L}}$  is always an extension of a further language  $\mathcal{L}$ . As the index  $\mathcal{L}$  is used here as a *variable*, we can have as many languages for attributions of rationality as we want. What remains the same throughout all these languages is the fact that they contain the doxastic operators, the rationality operators and they obey certain rules, to be precisely formulated in the next chapter.

Let  $\Omega$  be a set of well-formed formulas (*wffs*) from a certain language  $\mathcal{L}$  (the language that we will use to describe the contents of doxastic attitudes) and let  $\phi, \psi \in \Omega$ . We have already been using the doxastic operators  $B$  and  $D$ . For example, when representing the reasons  $\Gamma$  available to a certain agent  $S$  we used  $\Gamma = \{B\psi_1, \dots, B\psi_n\}$ <sup>119</sup>. As we emphasized earlier,  $B\phi$ ,  $D\psi$ , etc., are supposed to denote particular doxastic attitudes. As such,  $B\phi$ ,  $D\psi$ , etc., by themselves do not have truth-values. So we have to include the reference to subjects in our formalization, because we want to attribute doxastic attitudes to *those subjects* in our language  $\mathcal{E}_{\mathcal{L}}$ .

We will do this by means of a subject-index: ' $B_s\phi$ ' expresses the fact that  $S$  believes  $\phi$  (or that the doxastic attitude  $B\phi$  is possessed by  $S$ ) and ' $D_s\phi$ ' expresses the fact that  $S$  doubts  $\phi$  (or that the doxastic attitude  $D\phi$  is possessed by  $S$ ). Likewise, if we want to say that  $S$  has reasons  $\Gamma = \{B\psi_1, \dots, B\psi_n\}$  we can use the conjunction ' $B_s\psi_1 \wedge \dots \wedge B_s\psi_n$ '.

So whenever  $\phi, \psi \in \Omega$  and  $S$  is a subject-index, not only  $B_s\phi$ ,  $D_s\phi$ ,  $B_s\psi$  and  $D_s\psi$  are *wffs* of our language  $\mathcal{E}_{\mathcal{L}}$ , but also any conjunction of these formulas is a *wff* of  $\mathcal{E}_{\mathcal{L}}$ . Further, by adding the negation ' $\neg$ ' and the disjunction ' $\vee$ ' to our vocabulary we can follow the usual formation rules for formulas with these connectives.

So far so good for attributions of doxastic attitudes to subjects. But we also need a further operator here — a *rationality operator*  $R$ . Now, when  $\phi, \psi$  are *wffs* of  $\mathcal{L}$  it does not follow that  $R\phi$  and  $R\psi$  are *wffs* of  $\mathcal{E}_{\mathcal{L}}$ . The rationality operator, remember, is supposed to have formulas attributing doxastic attitudes to agents in their scope and, therefore, only formulas whose main operator is a doxastic operator can be preceded by  $R$ . For example, let us use the language of *propositional logic*,  $\mathcal{P}\mathcal{L}$ , as an index to our language  $\mathcal{E}$ , in such a way as to have  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ . Now suppose  $\phi$  is a well-formed formula of  $\mathcal{P}\mathcal{L}$ . Given what we said so far,  $B_s\phi$  and  $D_s\phi$  are *wffs* of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  — but  $R\phi$  is not. It would make no sense to say that a proposition or sentence with no doxastic operators 'is rational'. Rather, the well-formed formulas that we will have here are as follows:  $R(B_s\phi)$  and  $R(D_s\phi)$ . The first one expresses the fact that *believing  $\phi$  is rational for  $S$*  — or that *the (actual or otherwise) attitude of belief towards  $\phi$  is rational for  $S$*  — and the second one expresses the fact that *doubting  $\phi$  (or suspending judgment about  $\phi$ ) is rational for  $S$*  — or that *the (actual or otherwise) attitude of doubt towards  $\phi$  is rational for  $S$* .

These syntactical forms may look strange at first. As we postulated above, ' $B_s\phi$ ' expresses a fact: the fact that  $S$  believes  $\phi$ . So it would appear that ' $R(B_s\phi)$ ' says that *the fact that  $S$  believes  $\phi$  is rational*, which is as weird as it is to say that (the truth of) a proposition is rational. But this is a case where syntactical forms hide subtleties that can only be taken into account by an appropriate semantics.

<sup>119</sup>Actually we have been using ' $R$ ' for that purpose, but since we will use that letter for another purpose from now on (see below), let us use the Greek capital letters ' $\Gamma$ ' and ' $\Sigma$ ' to denote sets of doxastic attitudes.

Consider an attribution of *ex ante* rationality: ‘Forming doxastic attitude  $\lambda$  towards proposition  $\phi$  is rational for  $S$ ’. Such an attribution clearly involves talking about *possible worlds*, *possible situations* or *possible states*. By saying that a certain doxastic attitude is *ex ante* rational for  $S$ , one is not thereby saying that  $S$  *already has* that doxastic attitude. Of course, the claim that  $S$  already believes  $\phi$  is *consistent* with the claim that believing  $\phi$  is rational for  $S$  — but the truth of the former is not *necessary* for the truth of the latter. When we attribute *ex ante* rationality to a doxastic attitude for  $S$  we are saying that, in the present situation,  $S$  *has reasons to form* that doxastic attitude and  $S$  *knows how* to form such a doxastic attitude on the basis of those reasons (if our explication of *ex ante* rationality is correct). The relevant modalization throughout other situations (or worlds or states-of-affairs) than the actual one lies in the *knowledge-how* condition: that  $S$  knows how to reason from  $\Gamma$  to  $B_s\phi$  while in state  $w$  means that, in  $w$ ,  $S$  is able to perform an inference from  $\Gamma$  to  $B_s\phi$  and, consequently, to reach a new state  $w'$  where  $S$  believes  $\phi$  on the basis of reasons  $\Gamma$ <sup>120</sup>. So part of what is expressed by ‘ $R(B_s\phi)$ ’ is the information that there is at least one situation or possible world where  $S$  competently believes  $\phi$  on the basis of certain reasons.

The truth of  $R(B_s\phi)$  in a state  $w$  should not, then, imply that  $B_s\phi$  is true in  $w$ . One might ask, then: ‘where’ (in which *state* or *possible world*) is the fact expressed by ‘ $B_s\phi$ ’ supposed to hold when the proposition expressed by our formula ‘ $R(B_s\phi)$ ’ is true? Consider: the truth of an attribution of *ex ante* rationality to a belief in  $\phi$  for  $S$  entails that  $S$  knows how to form that belief on the basis of certain reasons  $\Gamma = \{B\psi_1, \dots, B\psi_n\}$ ; that  $S$  knows how to form such a belief in  $\phi$  on the basis of those reasons entails that there is a series of ordered-pairs of possible states  $\langle w, w' \rangle$  where  $(B\psi_1 \wedge \dots \wedge B\psi_n)$  is true in  $w$  and  $B_s\phi$  is true in  $w'$ <sup>121</sup>.

So the short answer to that question is this: when the proposition expressed by our formula ‘ $R(B_s\phi)$ ’ is true, the fact expressed by ‘ $B_s\phi$ ’ is supposed to hold in all states  $w'$  pertaining to the second position of the mentioned ordered-pairs. The long answer would involve explaining why the truth of an attribution of the relevant *knowledge-how* entails such a series of ordered-pairs of possible worlds or states  $\langle w, w' \rangle$  (the agent that is said to know how to perform a certain inference is taken to be able to reason from her doxastic attitudes in  $w$  and to form new doxastic attitudes (or to abandon previously held ones), achieving a further state  $w'$ ). We will not go through such a long answer. Suffice it to say that the use of the notions of *procedural knowledge* and *ability* involves making reference to possible situations where certain ‘deeds’ are accomplished by the relevant agent, and it

<sup>120</sup>It does not follow from here that  $S$  will *rationally* believe  $\phi$  in whatever state that  $S$  can reach from  $w$  by performing the relevant inference.

<sup>121</sup>This is entailed by our explication of *knowledge of how to reason* presented in *Chapter 2* and, presumably, it is also entailed by other accounts of this type of procedural knowledge.

is in these possible situations that formulas of the type  $B_s\phi$  are supposed to be true when a formula of the type  $R(B_s\phi)$  is true<sup>122</sup>.

We just fleshed out one more operator that will be used in our formal language  $\mathcal{E}_{\mathcal{L}}$ : the one present in  $R(B_s\phi)$ , where  $\phi$  is a formula of  $\mathcal{L}$ . For reasons that will be presented in the next few paragraphs, we will call the operator  $R$  as occurring in ‘ $R(B_s\phi)$ ’ the ‘absolute–rationality operator’. Accordingly, a more appropriate way of reading ‘ $R(B_s\phi)$ ’ is: ‘Believing  $\phi$  is *absolutely* rational for  $S$ ’, or: ‘Believing  $\phi$  is rational for  $S$  *all things considered*’. Appropriate truth–conditions for formulas involving the absolute–rationality operator will be given in the next chapter (*Chapter 5*).

Besides an *absolute–rationality* operator, we will also need a *relative–rationality* operator. The reason why we need a relative–rationality operator is that we need to take into account the fact that the reasons in virtue of which believing  $\phi$  is (absolutely) rational for  $S$  are undefeated. Suppose  $\Sigma$  is the *total set* of reasons available to  $S$  at  $t$ . Further, assume that there is a proper subset  $\Gamma \subset \Sigma$  such that  $\Gamma$  rationalizes believing  $\phi$  for  $S$  (the content of  $\Gamma$  gives support to  $\phi$  and  $S$  knows how to perform an inference from  $\Gamma$  to  $B\phi$ ). However, there is another subset  $\Gamma' \subset \Sigma$  such that  $\Gamma'$  rationalizes believing  $\neg\phi$  for  $S$  (or such that  $\Gamma \cup \Gamma'$  rationalizes doubting  $\phi$  for  $S$ ), and there is no other subset of  $\Sigma$  that does the job of ‘restoring’ justification for  $S$  to believe  $\phi$ . Assuming that the members of  $\Gamma$  are as justified as the members of  $\Gamma'$ , we can say that although believing  $\phi$  is rational for  $S$  relative to the set of reasons  $\Gamma$ , believing  $\phi$  is not rational for  $S$  relative to the set of reasons  $\Gamma \cup \Gamma'$  and, therefore, believing  $\phi$  is not rational for  $S$  relative to the set of reasons  $\Sigma$ . In other words, the epistemic status of  $B_s\phi$  is defeated.

It would follow, then, that believing  $\phi$  is not absolutely rational for  $S$ ,  $\neg R(B_s\phi)$ , but believing  $\phi$  is rational for  $S$  relative to the set of reasons  $\Gamma$ , or  $R(B_s\phi \mid \Gamma)$ . We can interpret an attribution of relative–rationality like  $R(B_s\phi \mid \Gamma)$  in the following way: believing  $\phi$  is rational for  $S$  relative to  $S$ ’s reasons  $\Gamma$ , and  $\Gamma$  only. If  $S$ ’s available reasons were  $\Gamma$  and  $\Gamma$  only, believing  $\phi$  would be absolutely rational for  $S$ . So the truth of  $R(B_s\phi \mid \Gamma)$  does not guarantee the truth of  $R(B_s\phi)$ , but the truth of  $R(B_s\phi)$  requires the truth of  $R(B_s\phi \mid \Gamma)$ , for some  $\Gamma$ .

Together, the operators  $B$ ,  $D$ ,  $R( )$ ,  $R( \mid )$  plus the usual connectives ( $\neg, \wedge, \vee$ ) and some set–theoretic notation will give us what we need to formalize attributions of epistemic rationality. Precise rules for well–formed formulas in our language  $\mathcal{E}_{\mathcal{L}}$  will be given in the next chapter.

---

<sup>122</sup>See Fantl (2012) for discussion.

## 4.2 Models – what we need to take into account

We want to develop a model–theoretic semantics to validate/invalidate formulas involving the rationality operators mentioned above. The validation of a formula occurs when we have a certain class or family of models such that the relevant formula is true in every model pertaining to that class or family. What we expect to validate in this way is a number of general principles of epistemic rationality. It is clear that some formulas are so obviously true that we can use them as a guiding idea to build our model-theoretic framework. As an example, consider the following formula-schema in  $\mathcal{E}_{\mathcal{L}}$  (where  $\phi, \psi$  are formulas of  $\mathcal{L}$  and  $S$  is an index to any subject, behaving as a free–variable):

$$(E) \quad R(B_s(\phi \wedge \psi)) \rightarrow R(B_s\phi)$$

This formula says that *if* believing  $(\phi \wedge \psi)$  is rational for  $S$ , then believing  $\phi$  is rational for  $S$ <sup>123</sup>. It is far from clear that one could object to **(E)** and, therefore, we can take this formula for granted and assume that no model–family that we expect to use as a validation structure for a logic with our rationality operators can contain a model that counter–exemplifies it. A similarly obvious principle is a version of **(E)** with the relative–rationality operator:

$$(E/) \quad R(B_s(\phi \wedge \psi) \mid \Gamma) \rightarrow R(B_s\phi \mid \Gamma)$$

that is: if believing  $(\phi \wedge \psi)$  is rational for  $S$  relative to  $S$ 's reasons  $\Gamma$ , then believing  $\phi$  is rational for  $S$  relative to  $S$ 's reasons  $\Gamma$ .

But while some of these formulas can be taken for granted when we are building our semantics, others generate controversy among epistemologists. For example, consider the following formula–schemata of  $\mathcal{E}_{\mathcal{L}}$ :

$$(G) \quad (R(B_s\phi) \wedge R(B_s\psi)) \rightarrow R(B_s(\phi \wedge \psi))$$

$$(C) \quad (R(B_s\phi) \wedge (\phi \rightarrow \psi)) \rightarrow R(B_s\psi)$$

schema **(G)** says that *if* believing  $\phi$  is rational for  $S$  *and* believing  $\psi$  is rational for  $S$ , then believing  $(\phi \wedge \psi)$  is rational for  $S$ . We will call schema **(G)** the ‘principle of agglomeration of rationality’, or ‘principle of agglomeration’ for short. In view of well–known paradoxes of rationality – most notably the *lottery paradox* and the *preface paradox*<sup>124</sup> — some philosophers have argued that we should reject principles of agglomeration like **(G)**<sup>125</sup>

<sup>123</sup>A more informal version of this principle can be found in Klein (1981, p. 45) under the name ‘Conjunction Principle’.

<sup>124</sup>See Sorensen (2006).

<sup>125</sup>See Kyburg (1961). Kyburg’s rejection of agglomeration commits him to the thesis that one can have a set of beliefs in jointly inconsistent propositions all of which are rational.

while others argue that we should not. Other versions of agglomeration, this time using the relative-rationality operator, are:

$$(\mathbf{G}/) \quad (R(B_s\phi \mid \Gamma) \wedge R(B_s\phi \mid \Gamma)) \rightarrow R(B_s(\phi \wedge \psi) \mid \Gamma)$$

$$(\mathbf{G}||) \quad (R(B_s\phi \mid \Gamma) \wedge R(B_s\phi \mid \Gamma')) \rightarrow R(B_s(\phi \wedge \psi) \mid \Gamma \cup \Gamma')$$

schema **(C)** says that if believing  $\phi$  is rational for  $S$  and  $\phi$  implies  $\psi$ , then believing  $\psi$  is rational for  $S$ . We will call schema **(C)** ‘principle of closure of rationality’, or ‘principle of closure’ for short. While it may appear to many that **(C)** is false, some epistemologists subscribe to modified versions of it<sup>126</sup>. Here are some relevant modified versions:

$$(\mathbf{C}^\wedge) \quad R(B_s(\phi \wedge (\phi \rightarrow \psi))) \rightarrow R(B_s\psi)$$

$$(\mathbf{C}/) \quad (R(B_s\phi \mid \Gamma) \wedge R(B_s(\phi \rightarrow \psi) \mid \Gamma)) \rightarrow R(B_s\psi \mid \Gamma)$$

$$(\mathbf{C}||) \quad (R(B_s\phi \mid \Gamma) \wedge R(B_s(\phi \rightarrow \psi) \mid \Gamma')) \rightarrow R(B_s\psi \mid \Gamma \cup \Gamma')$$

It is not our purpose to decide if schemata **(G)**, **(C)** or their derived versions are true now. The present point is that the fact that these schemata are controversial (as general principles of rationality) is a sufficient reason for us not to use them as guiding-ideas to build our models. Some schemas in  $\mathcal{E}_{\mathcal{L}}$  such as the ones presented above will be validated by certain family of models, and some will not. Maybe some of them will be validated *only if* others are validated, and maybe we can find a co-extensionality between two or more of these general schemas (whenever a family of models makes one of these schemas valid, it makes another schema valid as well, and vice-versa). But in order for us to have validity of formulas we need to have *families of models*. And in order for us to have families of models we need first to develop the structure of the particular models. So this must be where we begin.

As we said in our *Introduction*, our models will be abstract or formal representations of situations that subjects are in. As the word ‘situation’ is vague here, we need to make it clearer what we mean by it before we decide what our models need to take into account. In order for us to explain what a situation is supposed to be, let us compare two agents  $S_1$  and  $S_2$ . There are at least two scenarios in which  $S_1$  and  $S_2$  may be in *different situations*. In the first one,  $S_1$  and  $S_2$  simply have a different set of reasons. The fact that the reasons available to  $S_1$  differ from the reasons available to  $S_2$  means that  $S_1$  and  $S_2$  are in different

<sup>126</sup>See Luper (2002, *Section 6*) for discussion. Peter Klein (1981) shows that closure of justification is not refuted by counterexamples to the principle of *transmissibility of evidence* (If  $E$  is evidence for  $\phi$  and  $\phi$  entails  $\psi$  then  $E$  is evidence for  $\psi$ ). When a belief in  $\phi$  is justified for  $S$  in virtue of  $S$ ’s reasons  $R$  and  $\phi$  entails  $\psi$ , the relevant principle of closure does not entail that believing  $\psi$  is justified for  $S$  *in virtue of reasons*  $R$  — but it entails that believing  $\psi$  is justified for  $S$  in virtue of *some* reasons  $R'$  ( $B_s\phi$  would do).

situations. In the second one, the same set of reasons is available to both  $S_1$  and  $S_2$  — they believe and doubt exactly the same things — but they differ in their inferential abilities. What  $S_1$  can do with the reasons available to her is not the same thing as what  $S_2$  can do with the reasons available to her:  $S_1$  is able to perform a certain inferential state–transition that  $S_2$  *is not* able to perform, or vice–versa.

This can be explained as follows. Let us say that both  $S_1$  and  $S_2$  are in a state  $w$ , consisting in a certain set of beliefs  $\Gamma$ . There is an inferential schema available to  $S_1$ , but not to  $S_2$ , such that its application to the set of reasons  $\Gamma$  would lead to a further state  $w'$  containing a belief in  $\phi$ . In this case,  $w'$  is *reachable* (or *accessible*) to  $S_1$  from  $w$  — but it is not *reachable* (or *accessible*) to  $S_2$  from  $w$ . Here, again, the situations that  $S_1$  and  $S_2$  are in are different.

So the situation of a subject  $S$  is a function of (i) the reasons available to  $S$  and (ii)  $S$ 's inferential abilities, which determines the state–transitions that  $S$  is able to perform. The first element establishes the *state* (the *doxastic state*) that a subject is in, and the second one establishes the *states that a subject is able to go to* from a certain state she is in. So we could sum up what we have so far by saying that a situation is a combination of a particular state with the set of states that are ‘accessible’ from it<sup>127</sup>.

It would appear, then, that if our models take into account (i) and (ii) we have all we need to validate the relevant formulas (the ones attributing epistemic rationality). *But* if this is all that our models will take into account, we will hardly manage to derive rich logical systems from our semantic framework. How could we, for example, use axiom–schemata and theorem–schemata from basic *propositional logic* and substitute their sentence variables to derive some relevant formulas? These axiom/theorem–schemata from propositional logic would not be validated by our models unless our models also contain what it takes to make true/false formulas from propositional logic: a plain truth–assignment to atomic formulas with recursive clauses for complex (non–atomic) formulas. Only in this way could our models make true/false not only epistemic and doxastic formulas (that is, formulas whose main operators are the *rationality* and *doxastic* operators), but also formulas from propositional logic.

So our models for formulas of  $\mathcal{E}_{\mathcal{L}}$  will contain the elements of models for formulas of another language  $\mathcal{L}$  (for example, the language for propositional logic  $\mathcal{PL}$ ). In other words, our models for  $\mathcal{E}_{\mathcal{L}}$  will *embed* models for  $\mathcal{L}$  (call the latter ‘embedded models’). We have a *logical* motivation to use embedded models in our models — we need formulas of a certain language  $\mathcal{L}$  to be validated in order for us to use them in our axiomatic systems.

---

<sup>127</sup>All of this starts suggesting that we may use a *possible–worlds semantics* for attributions of rationality. But, as we will see, our states differ from what is commonly understood as a ‘world’ in possible–worlds semantics. Accordingly, we will call our semantics a ‘*possible–states semantics*’ (see *Chapter 5*).



Our models will be structured, then, as follows. First, they will contain a set  $W$  of (possible or actual) states of a modelled subject  $S$ . Second, they will contain an assignment of doxastic attitudes  $dox_s$  to each state  $w \in W$ . Such an assignment will determine which reasons are available to  $S$  in each (possible or actual) state. Third, they will contain a function  $i_s$  that maps inferential schemata (applied to particular doxastic attitudes) onto pairs of states  $\langle w, w' \rangle$ . The function  $i_s$  will determine which reasoned state–transitions the subject is able to perform in each state. Finally, our models will embed the elements of models for formulas in a language  $\mathcal{L}$ . By means of models structured in this way we will be able to make true/false formulas like  $B_s\phi$ ,  $D_s\phi$ ,  $R(B_s\phi \mid \Gamma)$ ,  $R(B_s\phi)$ , etc., and we will be able to validate/invalidate schemata such as **(E)**, **(G)** and **(C)**.

So in general our models will be structures:

$$M = \langle W, dox_s, i_s, [\text{elements of a model for } \mathcal{L}] \rangle,$$

where [elements of a model for  $\mathcal{L}$ ] can be substituted by a truth–assignment to atomic formulas (when  $\mathcal{L}$  is the language of *propositional logic*), or by a domain  $D$  of objects and a function that maps predicates and proper–names onto subsets and members of  $D$  respectively (when  $\mathcal{L}$  is the language of *first–order logic*) and so on. In principle, we can try to embed any model for any other language  $\mathcal{L}$  in our models. (Notice that the relevant language  $\mathcal{L}$  is the same that will be used to describe the contents of doxastic attitudes in our syntax).

The structure of our models will be explained in more detail in the next chapter. Now we need to set forth some *criteria of adequacy* for a semantics of rationality attributions. We want our models to ‘make’ attributions of epistemic rationality true or false *in an accurate way*<sup>128</sup>. That is, given a certain model representing a situation of a certain subject  $S$ , we want it to make true the formula  $R(B_s\phi)$  in a certain state  $w$  when and only when believing  $\phi$  is rational for  $S$  in state  $w$ . Further, we want the ‘judgments’ of our model–theoretic semantics to agree with the judgments of competent speakers of natural languages. If our semantics makes formulas of the type  $R(B_s\phi)$  true in several situations where competent speakers do not judge that it is rational for  $S$  to believe  $\phi$ , then we have a problem. Maybe our models *failed to include* a crucial element, one that we need to take into account when we are judging if a certain belief is rational or not, or maybe our models did not fail to include all crucial elements in this sense, *but the way we defined* the relationship between our models and the relevant formulas is problematic.

<sup>128</sup>This may sound strange — but notice that the truth–making relation that holds between a formal model and a set of formulas is, presumably, different from the truth–making relation that holds between facts in the world and sentences in natural language.

So, roughly, a semantics for attributions of rationality will be *adequate* when it is accurate and when it is in a certain level of agreement with the judgments of competent speakers of natural languages. Let us formulate our criteria of adequacy in a more precise way now. Begin with accuracy for a model  $M$ . A model  $M$  will be said to be accurate when the following holds:

A formula  $R(B_s\phi)$  is true in a state  $w$  of  $M$  when and only when it is rational for  $S$  to believe  $\phi$  in state  $w$ .

Here, we have to count on what we think is the best theory of epistemic rationality. The theory we have been defending says that it is rational for  $S$  to believe  $\phi$  when and only when there is a set of undefeated reasons  $\Gamma$  available to  $S$  such that the content of  $\Gamma$  gives support to  $\phi$  and  $S$  knows how to form an inferential belief in  $\phi$  on the basis of  $\Gamma$ . That means that our notion of accuracy becomes:

A formula  $R(B_s\phi)$  is true in a state  $w$  of  $M$  when and only when  $M$  represents the situation of an agent  $S$  such that, in state  $w$ ,  $S$  has a set of undefeated reasons  $\Gamma$ , the content of which gives support to  $\phi$ , and there is an inferential schema  $\alpha(\Gamma) = B_s\phi$  available to  $S$  in  $M$  such that there is a further state  $w'$  that is reachable from  $w$  via an application of  $\alpha$  to  $\Gamma$ .

Roughly, this gives us the following general criterion of adequacy for models for rationality attributions:

- (*Cr1*) The truth or falsity of formulas that attribute/deny rationality to doxastic attitudes for a certain subject  $S$  in a certain state should be a function of two things: the reasons available to  $S$  in that state and the inferential abilities possessed by  $S$  in that state.

If our theory of epistemic rationality is right, a model for rationality attributions will be adequate only if it satisfies (*Cr1*). This criterion describes one aspect of what it means to say that a model of the relevant type is accurate.

Of course, this is not all about accuracy. We also want our semantic framework to be *consistent*. In particular, we will have reasons to doubt the accuracy of our semantics if it says that believing  $\phi$  is rational and believing  $\phi$  is not rational for a certain subject  $S$  in a certain state  $w$ . That would give us one more criterion of adequacy:

- (*Cr2*) A semantics for rationality attributions should not make true both, a formula that attributes rationality to a doxastic attitude for a certain subject in a certain state *and* its negation.

This is a ‘non–dialetheist’ condition of adequacy for models of epistemic formulas<sup>129</sup>. It says that an attribution of rationality and its negation cannot both be true in the same model and the same state (if we interpret negation classically, of course). Why is this a criterion of adequacy for these models? Why could not we allow for ‘dialetheias’ of the type  $R(B_s\phi) \wedge \neg R(B_s\phi)$ ? Suppose that believing  $\phi$  is rational for  $S$  in state  $w$ . That means that, in  $w$ ,  $S$  has a set of undefeated reasons  $\Gamma$  whose contents give support to  $\phi$  and  $S$  knows how to reason from  $\Gamma$  to  $B_s\phi$ . What could also make it the case that believing  $\phi$  is *not* rational for  $S$  in such a situation? Either  $S$  does not know how to reason from  $\Gamma$  to  $B_s\phi$ , or  $S$  has equally good reasons for believing  $\neg\phi$ /suspending judgment about  $\phi$  in  $w$ , or  $S$  has no reasons at all for believing  $\phi$ .

In the first case, we have a contradiction with our initial assumption (that believing  $\phi$  is rational for  $S$  in  $w$ ) and, therefore, it follows that either our assumption is false or it is false that  $S$  does not know how to reason from  $\Gamma$  to  $B_s\phi$ . In the second one, we also have a contradiction: if it is true that  $S$  has equally good reasons for believing  $\neg\phi$ /suspending judgment about  $\phi$  in  $w$ , then the reasons  $\Gamma$  that  $S$  has to believe  $\phi$  in  $w$  are *not undefeated*. It follows that either our initial assumption is false or it is false that  $S$  has equally good reasons for believing  $\neg\phi$ /suspending judgment about  $\phi$  in  $w$ . In the third case we have again a contradiction with our initial assumption, for the fact that it is rational for  $S$  to believe  $\phi$  entails that  $S$  has good reasons to believe  $\phi$ . So, if our theory of epistemic rationality is right, a model for rationality attributions will be adequate only if it satisfies (*Cr2*).

There are still other questions about accuracy that we would have to deal with here. For example, one might worry that the representations of doxastic attitudes and abilities to reason in the relevant models are not empirically adequate, or not coherent with current cognitive psychology. If they are not, the accuracy of our models might be threatened. Accordingly, only empirically adequate models would count as adequate overall. This is a legitimate worry, but taking it into account now would require covering the literature on doxastic attitudes and reasoning abilities in cognitive psychology, and we have no space to do that here — that would require a separate investigation. Still, we think that the representations of doxastic attitudes and reasoning abilities (in terms of state–transitions) in our models will be sufficiently neutral as not to suffer from any danger of empirical disconfirmation<sup>130</sup>. Notice that there are two types of accuracy here: the accuracy of a model *as a representation of cognitive states and abilities* (the model accurately represents the cognitive states and abilities of subjects) and the accuracy of a model as a truth–maker

<sup>129</sup>About dialetheism and the possibility of there being true contradictions, see Priest and Smiley (1993).

<sup>130</sup>All we do is to represent doxastic attitudes as having propositional contents (represented by assertoric sentences) and to represent abilities to reason as abilities to reach certain cognitive states from the cognitive state one is in.

of attributions of rationality (the model accurately makes formulas of the form  $R(B_s\phi)$  true, in the sense described above). It is the latter type of accuracy that we aim to achieve here.

So much for accuracy. We also saw before that a semantics for attributions of rationality will be *adequate* only when it is in a certain level of agreement with the judgments of competent speakers of natural languages. It is quite vague what a ‘certain level of agreement’ means here, so let us try to precisify this requirement a bit. There are some cases where almost every competent speaker of English agrees that a certain belief is rational for someone. For example, if we were to describe a situation such as the following to a competent speaker of English:

Rachel rationally believes that *several reliable historians report that John Locke died in 1704*, she knows how to infer that *John Locke died in 1704* from that belief and she has no reason to doubt or disbelieve that *John Locke died in 1704*,

and if we were to ask the competent speaker if it is rational for Rachel to believe that *John Locke died in 1704*, he would most likely answer ‘yes’. In fact, if the relevant speaker were to answer our question negatively we would begin to doubt that he is using the term ‘rational’ in a competent way (maybe the speaker thinks that believing  $\phi$  is rational for  $S$  only if  $S$  is *absolutely certain* that  $\phi$  is true, or something like that).

But not all attributions of epistemic rationality are *that* unanimous and uncontroversial. There are some problematic cases where opinions may diverge greatly. For example, when a subject rationally believes each of  $\{\phi_1, \dots, \phi_n\}$  some people may judge that it is rational for that subject to believe  $(\phi_1 \wedge \dots \wedge \phi_n)$ , depending on how great  $n$  is, while some people may judge that it is not. Or take the case of disagreement among ‘epistemic peers’<sup>131</sup>, that is, equally responsible and competent subjects with access to roughly the same evidence. Suppose  $S_1$  rationally believes that  $\phi$ . Then  $S_1$  learns that an epistemic peer of her,  $S_2$ , believes that  $\neg\phi$ .  $S_1$  knows that  $S_2$  is as responsible and competent as  $S_1$  is, and that they both have access to roughly the same evidence. What should  $S_1$  do (regarding her opinion about  $\phi$ ) when she learns that her peer  $S_2$  believes  $\neg\phi$ ? Should  $S_1$  ‘stick to her guns’ and maintain her belief in  $\phi$ , or should she change her initial opinion about  $\phi$ ? Some people may think that it is rational for  $S_1$  to maintain her previous position, some people may think it is not.

So when it comes to attributions of rationality there is a class of cases that can be classified as *controversial*. The closer the divide between the judgments of competent

---

<sup>131</sup>For an up-to-date discussion on the topic of disagreement among epistemic peers, see the special volume by Christensen and Lackey (2013).

speakers is to 50/50, the more controversial is the case. Accordingly, we do not want to use controversial cases to test the adequacy of our semantics for attributions of rationality. In order to do so, we would have to decide first which one of the ‘camps’ in a given controversy is the right one — and maybe the theory of rationality we have chosen to ground our semantics, as it stands, does not allow us to decide that. It turns out, then, that the agreement we expect our semantics to have with the judgments of competent speakers of a natural language is restricted to *non-problematic* cases: simple and uncontroversial, *quasi* unanimous, cases. The bigger the overlap of the ‘judgments’ about non-problematic cases outputted by our semantics with the judgments of competent speakers, the more adequate our semantics is. This gives us the following criterion of adequacy:

- (Cr3) A semantics for attributions of rationality should make true/false formulas that attribute/deny rationality to doxastic attitudes in such a way as to agree with the judgments of competent speakers of English about *non-problematic* cases.

Several issues need to be worked out here. We mention three. First, in order for us to have a class of judgments about non-problematic cases to be tested against our semantics do we need to survey people’s opinions about particular cases and ask them what is rational for fictitious characters to believe? Or does our acquaintance with English-speakers and our own linguistic competence suffice for us to judge if the relevant cases are problematic or not? It is clear that getting statistically robust data about the judgments of speakers, if not a necessary step towards a decision about the adequacy of our semantics, is a desirable move. That would give empirical confirmation/disconfirmation for the claim that a certain class of cases is a class of non-problematic cases. Still, some such cases are so simple and clearly uncontroversial that maybe it would be a waste of time to survey people’s opinions about them.

Second, assuming that we have a class of judgments about non-problematic cases, how big must be the overlap between these judgments and the ‘judgments’ of our semantics? Must they have complete agreement (an overlap of 100%)? Or will a lower threshold suffice (say, 95%)?

Third, and this is related to the first issue, there is no guarantee that the concept of *rationality* used by different speakers, or by the same speaker in different cases, has the same semantic properties (same implications, same truth-conditions, etc). To be sure, there are many concepts of *rationality*: rationality as *intelligence*, rationality as *unblameworthiness*, rationality as *maximization of goals*, rationality as *emotional control*, etc. Maybe the task of finding a (theoretically relevant) common factor between all of these notions is unfeasible. But it is clear that speakers are stimulated to use different concepts of *rationality* depending on the type of case they are judging, on how the case

---

is described, on what questions are made to them exactly, on what are the most relevant features of the case, etc. So if we are to use the judgments of speakers to determine a class of non-problematic cases we need to use cases with a certain structure that is suited to our task. Only in this way we can avoid stimulating judgments about concepts of *rationality* that are distant from the one we are interested in.

The details about these three issues need not bother us at this point. The important thing to notice now is that in order for us to determine if a semantics for attributions of rationality satisfies (*Cr3*) we may be required to use experimental methods (unlike (*Cr2*), the satisfaction of which can be verified purely *a priori*).

So, we fleshed out three *criteria* of adequacy for a semantics for attributions of epistemic rationality. (*Cr1*) and (*Cr2*) are criteria of accuracy, and (*Cr3*) is a criterion of coherence with natural language. Of course, there may be more *criteria* of adequacy to be fleshed out and, therefore, there is more meta-theoretical work to be done here. But the *criteria* we exposed above can already be used to test and to assess our semantics (to be presented in the next chapter).

# Chapter 5

## *Syntax and semantics*

In this chapter we develop a language  $\mathcal{E}_{\mathcal{L}}$  that will be used to formalize attributions of epistemic rationality. A model-theoretic semantics for formulas in a version of  $\mathcal{E}_{\mathcal{L}}$  ( $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ ) will be sketched. We will present *simple models*, which constitute a family of models in our ‘possible-states’ semantics. Finally, we discuss the properties of our semantics in the light of the *criteria* of adequacy advanced in the previous chapter.

### 5.1 A formal language

As we saw in *Chapter 4* our language  $\mathcal{E}_{\mathcal{L}}$  — the language that we will use to formalize attributions of epistemic rationality — is always an extension of a further language  $\mathcal{L}$ . Now we will establish precise recursive rules of formulæ-formation in  $\mathcal{E}_{\mathcal{L}}$ . ‘ $\mathcal{L}$ ’ will be use here as a variable for any other language, and we should assume that  $\mathcal{L}$  has precise rules of formulæ-formation as well.

Let  $\Omega$  be the set of well-formed formulas (*wffs*) of a certain language  $\mathcal{L}$ . Further, let  $Sub = \{S_1, \dots, S_n\}$  be a set of  $n$  subject-names (a variable ‘ $S$ ’ for the members of  $Sub$  will be used as an index to formulas with doxastic operators). Then we can give the following recursive rules for well-formed formulas in  $\mathcal{E}_{\mathcal{L}}$ :

For every formulas  $\Phi, \Psi$  and set of formulas  $\Gamma$ :

1. If  $\Phi \in \Omega$  then  $\Phi$  is a *wff*;
2. If  $\Phi$  is a *wff*, then  $\neg\Phi$  is a *wff*;
3. If  $\Phi, \Psi$  are *wffs*, then  $(\Phi \wedge \Psi)$  is a *wff*;
4. If  $\Phi \in \Omega$  and  $S \in Sub$ , then  $B_s\Phi, D_s\Phi$  are *wffs*;
5. If  $B_s\Phi, D_s\Phi$  are *wffs*, then  $R(B_s\Phi), R(D_s\Phi)$  are *wffs*;

- 
6. If  $B_s\Phi$ ,  $D_s\Phi$  are *wffs* and  $\Gamma$  is a set of *wffs* of  $\mathcal{E}_{\mathcal{L}}$  whose main operators are doxastic operators, then  $R(B_s\Phi \mid \Gamma)$ ,  $R(D_s\Phi \mid \Gamma)$  are *wffs*;
7. Nothing else is a *wff*.

We also assume the standard identities by definition among formulas with boolean operators when  $\wedge$  and  $\neg$  are taken to be primitive. So:

$$(\Phi \vee \Psi) =_{def.} \neg(\neg\Phi \wedge \neg\Psi),$$

$$(\Phi \rightarrow \Psi) =_{def.} \neg(\Phi \wedge \neg\Psi),$$

$$(\Phi \leftrightarrow \Psi) =_{def.} \neg(\Phi \wedge \neg\Psi) \wedge \neg(\neg\Phi \wedge \Psi).$$

for any formulas  $\Phi$ ,  $\Psi$  of  $\mathcal{E}_{\mathcal{L}}$ .

It is noteworthy that the rules presented above quantify not only over formulas — but also sets of formulas. More specifically, rule 6 makes use of the notion of a set of *wffs* of a certain kind (*wffs* whose main operators are doxastic operators). Examples of sets allowed in the scope of the relative-rationality operator are:  $\Gamma = \{B_s\Phi, D_s\Psi\}$ ,  $\Gamma' = \{B_s(\Phi \rightarrow \Psi)\}$ ,  $\Sigma = \{B_s\Phi_1, \dots, B_s\Phi_n\}$  (we use  $\Gamma$  and  $\Sigma$  for sets of doxastic attitudes, sometimes with the usual ‘prime’ symbol).

The notion of *set* mentioned in rule 6 must be assumed to involve the syntactic properties of the standard notation used in set-theory. Assume that  $\Gamma$  and  $\Sigma$  are sets of *wffs* of  $\mathcal{E}_{\mathcal{L}}$  whose main operators are doxastic operators. Given that much, if  $R(B_s\Phi \mid \Gamma)$ ,  $R(D_s\Phi \mid \Gamma)$  are well-formed formulas of  $\mathcal{E}_{\mathcal{L}}$ , then  $R(B_s\Phi \mid \Gamma \cup \Sigma)$ ,  $R(D_s\Phi \mid \Gamma \cup \Sigma)$ , where ‘ $\cup$ ’ represents the usual *union* operation, are well-formed formulas of  $\mathcal{E}_{\mathcal{L}}$  as well<sup>132</sup>.

Summing up,  $\mathcal{E}_{\mathcal{L}}$  is an extension of a previous language  $\mathcal{L}$ , it contains indexed doxastic operators and it contains the absolute and relative-rationality operators, where some set-theoretic notation is used in the scope of the latter. In principle,  $\mathcal{E}_{\mathcal{L}}$  can be established for any formalized language  $\mathcal{L}$  whose formulas represent assertive sentences. We will occupy ourselves with the simplest case where  $\mathcal{L}$  is the language of *propositional logic*,  $\mathcal{P}\mathcal{L}$ . That will give us  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ . In  $\mathcal{P}\mathcal{L}$  we will use letters  $p$ ,  $q$ ,  $r$ ,  $s$  as atomic formulas (sometimes with natural numbers as indexes) and also the boolean operators ( $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$ ,  $\leftrightarrow$ ) in the usual way. Metalinguistically, lowercase Greek letters  $\phi$ ,  $\psi$ ,  $\chi$  will be used as variables for formulas of  $\mathcal{P}\mathcal{L}$  only, and capital Greek letters  $\Phi$ ,  $\Psi$ , will be used as variables for any formula of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  (so  $\Phi$  can be used both, as a variable for formulas of  $\mathcal{P}\mathcal{L}$  and formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ , while  $\phi$  can be used as a variable for formulas of  $\mathcal{P}\mathcal{L}$  but not for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ ).

---

<sup>132</sup>‘ $\Gamma \cup \Sigma$ ’ reads as ‘the union of  $\Gamma$  and  $\Sigma$ ’, and it is defined as  $\Gamma \cup \Sigma = \{\Phi : \Phi \in \Gamma \text{ or } \Phi \in \Sigma\}$ .



The rules for well-formed formulas in  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  are established by taking  $\Omega$  to be the set of well-formed formulas of  $\mathcal{P}\mathcal{L}$  in the rules presented above.

Whenever we use a language  $\mathcal{E}_{\mathcal{L}}$  to represent attributions of epistemic rationality, our semantics will be constituted by models *over formulas of  $\mathcal{L}$* . The language  $\mathcal{L}$  is both, the one that will be extended by  $\mathcal{E}_{\mathcal{L}}$  and the one that will be used to represent the contents of the doxastic attitudes held by the modelled subjects. Furthermore, our models for formulas of  $\mathcal{E}_{\mathcal{L}}$  will not only be models *over formulas of  $\mathcal{L}$* , but also models *over a set of inferential schemata* whose parameter-language is  $\mathcal{L}$ .

A language  $\mathcal{L}$  is the parameter-language of a certain inferential schema  $\alpha$  when only variables and logical constants of  $\mathcal{L}$  are used to represent the ‘content placeholders’ in  $\alpha$ . So, suppose we are using a language  $\mathcal{E}_{\mathcal{F}\mathcal{O}\mathcal{L}}$ , where  $\mathcal{F}\mathcal{O}\mathcal{L}$  is the language of *first-order logic* that includes the identity sign ‘=’. In the language  $\mathcal{F}\mathcal{O}\mathcal{L}$ ,  $x, y, z$  are variables for proper-names and  $a, b, c$  are proper-names. A model for  $\mathcal{E}_{\mathcal{F}\mathcal{O}\mathcal{L}}$  will be a model over formulas of  $\mathcal{F}\mathcal{O}\mathcal{L}$  and inferential schemata whose parameter-language is  $\mathcal{F}\mathcal{O}\mathcal{L}$ . An inferential schema of this type is  $tid(B(x = y \wedge y = z)) = B(x = z)$  (here, the ‘content placeholders’ are  $(x = y \wedge y = z)$  and  $(x = z)$ )<sup>133</sup>. In this case, a model for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  would neither be a model over formulas of  $\mathcal{F}\mathcal{O}\mathcal{L}$  nor a model over inferential schemata such as  $tid$ . A model for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  must be a model over formulas of  $\mathcal{P}\mathcal{L}$  and inferential schemata whose parameter-language is  $\mathcal{P}\mathcal{L}$ .

It is also noteworthy that we can capture different properties of a type of inference depending on which parameter-language we are using. For example, if the parameter-language of  $tid$  is  $\mathcal{P}\mathcal{L}$  we will have  $tid(B(\phi \wedge \psi)) = B\chi$ <sup>134</sup>. It is clear that the latter is not an *optimal* inferential schema. But  $tid(B(x = y \wedge y = z)) = B(x = z)$  is an *optimal* inferential schema. So  $tid$  parameterized with  $\mathcal{P}\mathcal{L}$  has different properties than  $tid$  parameterized with  $\mathcal{F}\mathcal{O}\mathcal{L}$  (these are not identical inferential schemata at all). Since there is a relation of dependence between our language  $\mathcal{E}_{\mathcal{L}}$  and the inferential schemata over which our models apply, we have to choose  $\mathcal{L}$  wisely in each situation. It may be that  $\mathcal{P}\mathcal{L}$  is too ‘coarse grained’ for us to make true some attributions of rationality. If, for example,  $S$  believes  $(a = b \wedge b = c)$  and  $S$  knows how to infer that  $(a = c)$  from that belief, the model used to represent  $S$ ’s situation will not do a good job unless it is a model over formulas of  $\mathcal{F}\mathcal{O}\mathcal{L}$  and inferential schemata whose parameter-language is  $\mathcal{F}\mathcal{O}\mathcal{L}$ . If the model representing  $S$ ’s situation is a model over formulas of  $\mathcal{P}\mathcal{L}$  and inferential schemata

<sup>133</sup>‘ $tid$ ’ abbreviates ‘transmissibility of identity’. The inferential schema  $tid$ , however, should not be confused with the derivation rule of *first-order logic* that could be called this way. Again, we use names of inferential schemata after names of derivation rules merely as a mnemonic resource.

<sup>134</sup>When the doxastic operators occur as input-variables/output-variables in the formal representation of an inferential schema, we avoid using an index to a subject  $S$ . Nevertheless, when we want to represent *particular* applications of inferential schemata to the doxastic attitudes of some particular agent, we will use subject-indexes attached to the relevant doxastic operators.

whose parameter–language is  $\mathcal{PL}$ , it will make false the claim that believing  $(a = c)$  (in this case an atomic formula  $r$ ) is rational for  $S$  in virtue of  $S$ 's belief in  $(a = b \wedge b = c)$  (in this case a conjunction of atomic formulas  $(p \wedge q)$ ), because the inferential schema  $\text{tid}(B(\phi \wedge \psi)) = B\chi$  is not an *optimal* one.

So, the syntactic structure of the sentences representing the contents of doxastic attitudes is crucial for us to make true/false formulas that attribute rationality to beliefs, because syntactic structure is crucial for us to determine which inferential schemata are available to reasoners. By using  $\mathcal{E}_{\mathcal{PL}}$ , then, we will not manage to capture a lot of epistemic properties we want to capture. Not only subjects reason in other, deeper levels than the propositional one — they also reason with modalities, probabilities, quantifiers, etc. The richer the syntactic structure of  $\mathcal{L}$  (a deeper level of representation, more logical constants and operators), the more reasoning abilities we can take into account and, therefore, the more accurate will be our attributions of epistemic rationality.

At this point, however, we are looking for *simplicity*. Our aim here is to start working with the simplest language, the simplest models, and to check how they work together. This is just supposed to be beginning of an adequate semantics for attributions of rationality — we expect to develop richer structures in future work. In what follows, we develop a semantics for attributions of epistemic rationality to reasoners that are able to reason at the propositional level (and the propositional level only).

## 5.2 A *possible–states* semantics

In this section we will develop what might be properly called a ‘*possible–states* semantics’ (or ‘ps–semantics’ for short), to be contrasted with a *possible–worlds* semantics<sup>135</sup>. Our models will not have the same structure as the one present in the well known *Kripke–models* of standard possible–worlds semantics<sup>136</sup>.

Roughly, Kripke–models are tuples  $\langle W, R, V \rangle$  over a set of formulas  $\Omega$ , where  $W$  is supposed to be a set of possible worlds,  $R$  is a binary relation on  $W$  (it says which worlds are ‘accessible’ from each world) and  $V$  is a valuation function that maps formulas in  $\Omega$  onto sets of possible worlds (it says which formulas are true in each world). These models validate both, formulas with a ‘universal’ modal operator (formulas whose truth–conditions consist in a quantification over *all* possible worlds) and formulas with an ‘existential’ modal operator that is the dual of the former one (formulas whose truth–conditions consist in a quantification over *some* possible worlds). So, for example, Kripke–models can be used to validate formulas that are supposed to formalize talk of *possibility* and *necessity*, where

<sup>135</sup>About possible–worlds semantics, see Menzel (2013).

<sup>136</sup>For a nice presentation of Kripke–models, see Blackburn, de Rijke and Venema (2001, p. 16).

both a necessity operator  $\Box$  and its dual  $\Diamond =_{def.} \neg\Box\neg$  are used. They can also be used to validate formulas with knowledge operators, doxastic operators, deontological modalities, tense operators, etc, and they give rise to a variety of intensional logics<sup>137</sup>.

Our models, with the structure  $M = \langle W, dox_s, i_s, [\text{elements of a model for } \mathcal{L}] \rangle$ , will differ from Kripke-models, in that we will have not only an alethic valuation function  $V$  but also a doxastic valuation function  $dox$  that says what is believed, disbelieved or doubted by a certain agent in each *state* of a set  $W$ . That means that there are two main differences between our models and Kripke-models.

First, the members of  $W$  are supposed to be actual or possible *states* of a certain subject, not *possible worlds* in the canonical sense. Possible worlds in the canonical sense are *complete* — for each world  $w$  either  $\Phi$  or  $\neg\Phi$  is true in  $w$ <sup>138</sup> — but our states are not like that (that is the main reason why we call our semantics a ‘possible-states’ semantics instead of a ‘possible-worlds’ semantics). For some members of our set  $W$ , neither  $B_s\phi$  will be true nor  $\neg B_s\phi$  will be true. Only three types of formulas are supposed to be made true in our states (the members of  $W$ ): formulas about what is believed/doubted by a certain subject, formulas attributing rationality to doxastic attitudes, and formulas of a certain language  $\mathcal{L}$  (initially,  $\mathcal{PL}$ ). The truth or falsity of the the first two types of formulas is not determined by  $V$  — only the truth or falsity of the latter type of formula is determined by  $V$  (that is, the truth of formulas of the type  $B_s\phi$  and formulas of the type  $R(B_s\phi)$  is not determined by  $V$ ). The truth of formulas of the former type is determined by  $dox$ , while the truth of formulas of the latter type is determined by  $dox$  plus our ‘accessibility’ function, to be presented below. Of course, that does not mean that the truth of formulas involving doxastic and rationality operators, but whose *main connective* is one of  $(\neg, \wedge, \vee, \rightarrow, \leftrightarrow)$ , cannot be also determined by  $V$ . This will be made clearer as we proceed.

Further, the truth-conditions for attributions of *doxastic attitudes* assumed here are not the usual ones in possible-worlds semantics — we chose a ‘syntactic’ approach to model those attributions. It is our overall purpose to offer a semantics for attributions of epistemic rationality that has something to say about *real*, instead of *ideal*, agents. Usually, Kripke-models for modal epistemic/doxastic logics tend to overlook cognitive limitations of agents, in the sense that they take agents to know or to believe things that are not in fact known or believed by those agents<sup>139</sup>. For example, modal epistemic logics minimally assume that knowledge/belief is ‘closed’ under entailment (agents are taken to know everything that is entailed by what they already know) and that all axioms/theorems

<sup>137</sup>See van Benthem (1988) for intensional logics in general and Blackburn, de Rijke and Venema (2001) for modal logics in particular.

<sup>138</sup>See Beal and Restall (2006, p. 50).

<sup>139</sup>See Duc (1997), Fagin, Halpern, Moses and Vardi (1995, *Chapter 9*).

of a certain logical system are known/believed by the modelled subjects. If we build our semantics with standard Kripke-models, we will have a semantics that is not as realistic as we expect it to be.

The models that we will develop in this section will be called ‘*simple models*’. These are models over the set  $\Omega$  of *wffs* of the language  $\mathcal{PL}^{140}$  and the set  $A$  of *optimal* inferential schemata whose parameter-language is  $\mathcal{PL}$ . Our models will look like models for *dynamic logics*<sup>141</sup> in that they are not only models over formulas but also over other structures, in this case, *inferential schemata* (in the case of some models for dynamic logics, *programs*).

Simple models are models for formulas of the language  $\mathcal{E}_{\mathcal{PL}}$  — an extension of  $\mathcal{PL}$ , with the doxastic and rationality operators, obeying the rules presented in the previous section. They have the following structure:

$$M = \langle W, dox_s, i_s, V \rangle,$$

which is an instantiation of the general structure:

$$M = \langle W, dox_s, i_s, [\text{elements of a model for } \mathcal{L}] \rangle$$

of our possible-states models. Since our language  $\mathcal{L}$  is  $\mathcal{PL}$ , the [elements of a model for  $\mathcal{L}$ ] consists simply in a valuation function  $V$  (a model for  $\mathcal{PL}$  is just such a valuation function): a plain truth-assignment to atomic formulas with recursive clauses for complex formulas.

Let us describe each element of our simple models in more detail now.  $W$  is a set  $\{w_1, \dots, w_n\}$  of (actual or otherwise) states of a subject  $S$  (the one whose situation is being modelled). In each of these states,  $S$  is supposed to have a certain set of reasons (this will be determined by  $dox_s$ ).  $S$  is also supposed to be able to perform transitions among the members of  $W$  (this will be determined by  $i_s$ ).

Notice that in our informal talk (*Chapters 1-3*) we talked about subjects having a certain set of reasons *at a certain time t* and being able to reason in a certain way *at a certain time t*. That was our way of individuating the states that reasoners are in (or could be in). Now we explicitly talk about a certain subject having a certain set of reasons (or being able to reason in a certain way) *in a certain state w* (instead of *at a certain time t*)<sup>142</sup>.

<sup>140</sup>From now on we always assume  $\Omega$  to have this interpretation.

<sup>141</sup>See Pacuit (2013).

<sup>142</sup>Some complication can arise here. According to the way we individuate states, if the same set of doxastic attitudes hold in a certain state  $w$  and in a certain state  $w'$ , and the same inferential abilities are present in  $w$  and  $w'$ , then  $w = w'$ . But it could be perfectly true that  $w$  and  $w'$  are actualized in different

The second element of our simple models is  $dox_s$  — a partial function that maps the set of formulas  $\Omega$  onto  $\{0, 1\}$  in each  $w \in W$ . So  $dox_s(w): \Omega \mapsto \{0, 1\}$ . When  $\phi \in \Omega$  and  $dox_s(w)(\phi) = 1$ ,  $S$  is said to *believe that*  $\phi$  in state  $w$ . On the other hand, if  $dox_s(w)(\phi) = 0$ ,  $S$  is said to *doubt that* (or to *suspend judgment about*)  $\phi$  in state  $w$  ( $S$  will be said to *disbelieve*  $\phi$  in  $w$  when  $dox_s(w)(\neg\phi) = 1$ ). This function determines the reasons that are available to  $S$  in each state  $w$ . As  $dox_s$  is a *partial* function, it does not assign a value to every member of  $\Omega$ . That is, there are some members of  $\Omega$  that  $dox_s$  does not map them to  $\{0, 1\}$  — intuitively, those sentences that are not even considered by  $S$ . The rules constraining  $dox_s$  will be presented in *sub-Section 5.2.2* below.

The third element of simple models is the function  $i_s$ . Let  $A$  be the set of *optimal* inferential schemata whose parameter-language is  $\mathcal{PL}$ . Then we can say that  $i_s$  is a function that maps the set of applications of members of  $A$  to particular doxastic attitudes onto the set of ordered pairs  $\langle w, w' \rangle$ , where  $w, w' \in W$ . Such a function determines the availability of inferential schemata for  $S$  in each state  $w$ . When  $\langle w, w' \rangle \in i_s(\alpha(\Gamma))$ , that means that state  $w'$  is *reachable* from  $w$  via an application of inferential schema  $\alpha$  to the set of doxastic attitudes  $\Gamma$ , written  $w \xrightarrow{\alpha(\Gamma)} w'$ . That  $w \xrightarrow{\alpha(\Gamma)} w'$  means that, in  $w$ ,  $S$  is able to reach  $w'$  by instantiating  $\alpha$  applied to  $\Gamma$ .

In the previous chapter we have been working with an *intensional* representation of inferential schemata, of the type  $\alpha(\Sigma) = B\phi$ , where  $\Sigma$  is a set of doxastic attitudes. Items of this type are the members of  $A$  over which our models range. We also said before (*Chapter 2*) that, extensionally, each inferential schema corresponds to a set of ordered pairs of states  $\langle w, w' \rangle$  where certain doxastic attitudes towards propositions represented in a certain language  $\mathcal{L}$  hold. It is important to ask, then: How are these representations, the intensional and the extensional, related?

Let us assume that  $w \xrightarrow{\alpha(\Gamma)} w'$  or, what is the same, that  $\langle w, w' \rangle \in i_s(\alpha(\Gamma))$ , for some simple model  $M$  and pair of states  $w, w' \in W$ . The intensional representation of an inferential schema is related with its extension of ordered pairs of states in the following way. *If* the intensional representation of  $\alpha$  is  $\alpha(\Sigma) = B\phi$  then<sup>143</sup>:

- $\Gamma = si_n(\Sigma)$  for some  $n$  and  $M, w \models \bigwedge \Gamma$ ;
- For the same  $n$ , there is a formula  $B_s\psi = si_n(B\phi)$  such that  $M, w' \models B_s\psi$ .<sup>144, 145</sup>

times, so they would not really be identical. We sidestep this complication for now, however, since the inclusion of time-points in our models would add a considerable degree of complexity.

<sup>143</sup> $\bigwedge \Gamma$  is, remember, the conjunction of all formulas in  $\Gamma$ . So if, for example,  $\Gamma = \{B_s\phi, B_s\psi, D_s\chi\}$ , then the formula  $\bigwedge \Gamma$  is equivalent with the formula  $(B_s\phi \wedge B_s\psi \wedge D_s\chi)$ .

<sup>144</sup>The symbol ' $\models$ ' is the usual 'models' or 'makes-true' symbol.

<sup>145</sup>Here, of course, the substitution instances are parameterized with  $\mathcal{PL}$ , that is, the relevant function

That is, if the intensional representation of  $\alpha$  is  $\alpha(\Sigma) = B\phi$  and  $w \xrightarrow[\alpha(\Gamma)]{} w'$ , then the set of doxastic attitudes  $\Gamma$  is a substitution instance of the input-variable  $\Sigma$  of  $\alpha$  and all doxastic attitudes in  $\Gamma$  are held by  $S$  in  $w$ . Further, there is a substitution instance (*uniform* with the first one) of the output-variable  $B\phi$  of  $\alpha$ ,  $B_s\psi$ , such that it is held by  $S$  in  $w'$  (in this case, we assume that  $si_n(B\phi) = B_s\psi$  holds in the (possible or actual) state  $w'$  because  $\alpha$  was applied to  $\Gamma$  in state  $w$ ).

Let us consider an example. Consider the ‘doxastic *modus ponens*’ inferential schema  $dmp(B\phi, B(\phi \rightarrow \psi)) = B\psi$ . Now suppose we have a simple model  $M = \langle W, dox_s, i_s, V \rangle$  where:

$$\begin{aligned} dox_s(w)(p) &= 1, \\ dox_s(w)(p \rightarrow q) &= 1, \end{aligned}$$

which means that  $M, w \models B_s p$  and  $M, w \models B_s(p \rightarrow q)$  (as we will see in the general truth-conditions for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  below, it follows from here that  $M, w \models B_s p \wedge B_s(p \rightarrow q)$ ). Further, let us assume that:

$$\langle w, w' \rangle \in i_s(dmp(B_s p, B_s(p \rightarrow q)))$$

and let us use  $\Gamma$  to represent the set of doxastic attitudes  $\{B_s p, B_s(p \rightarrow q)\}$ . That means that  $w \xrightarrow[dmp(\Gamma)]{} w'$ , or that  $w'$  is reachable from  $w$  via an application of  $dmp$  to  $\Gamma$ .

Here we can see what exactly is the relationship between the intensional representation of  $dmp$ ,  $dmp(B\phi, B(\phi \rightarrow \psi)) = B\psi$ , and its extensional aspect. The set of formulas  $\Gamma = \{B_s p, B_s(p \rightarrow q)\}$  is a substitution instance of  $dmp$ ’s input-variable  $\{B\phi, B(\phi \rightarrow \psi)\}$ , and  $M, w \models \bigwedge \Gamma$ . Further,  $B_s q$  is a substitution instance (uniform with the former one) of  $dmp$ ’s output-variable  $B\Psi$ , and  $M, w' \models B_s q$ . That means that, assuming that the inferential schema  $dmp$  is available to  $S$ , when  $S$  is in  $w$  she can perform a state-transition to  $w'$  by instantiating  $dmp$ , and  $S$  will also be able to instantiate  $dmp$  in other states where doxastic attitudes that constitute substitution instances of the input-variable of  $dmp$  hold. So, the availability of  $dmp$  gives rise to a set of ordered pairs of states, where the second member of these pairs is reachable from the first member via an application of  $dmp$  to

---

is  $si^{\mathcal{P}\mathcal{L}}$ . We omit the index to  $\mathcal{P}\mathcal{L}$  for easier readability. Further, the intended substitution instance of  $\Sigma$  is one where the index to the subject  $S$  is added to the doxastic operators. In order to be fully precise, we would need to flesh out a further substitution-instance function that does just that, but let us step aside that complication for now and let us just assume that  $si$  not only substitutes the  $\mathcal{P}\mathcal{L}$ -variables by  $\mathcal{P}\mathcal{L}$ -constants in the contents of doxastic attitudes, but also adds a subject-index  $S$  to the relevant doxastic operators.

doxastic attitudes that constitute substitution instances of the input-variable of  $dmp$ .

Some states are reachable from a state  $w$  only if other states are reachable from  $w$ . Suppose that in state  $w$  subject  $S$  believes that  $(p \wedge q)$  and that state  $u$  is reachable from  $w$  via an application of the inferential schema  $ce(B(\phi \wedge \psi)) = B\phi$ . Given that in state  $u$  subject  $S$  believes that  $p$ , let us assume that a further state  $v$  is reachable from  $u$  via an application of the inferential schema  $di(B\phi) = B(\phi \vee \chi)$ . Is  $S$  able to reach state  $v$ , where  $B_s(p \vee r)$  holds, from state  $w$ , where  $B_s(p \wedge q)$  holds? Intuitively, yes. For we are assuming that  $S$  knows how to infer a conjunct on the basis of a belief in a conjunction, and that  $S$  knows how to infer a disjunction on the basis of a belief in one of its disjuncts. It is a simple ‘two-steps’ chain of reasoning. So we want to say that  $v$  is *reachable* from  $w$ . Yet, as far as our assumptions go, the pair  $\langle w, v \rangle$  does not belong to the scope of a *first-order* inferential schema (see *Chapter 2, Section 2.5*). But maybe that pair belongs to the scope of a *second-order* inferential schema. How can we take this into account?

We will assume that state-transitions are *transitive* under compositionality of inferential schemata: where  $\alpha(\Gamma) = \lambda$  is a particular application<sup>146</sup> of  $\alpha$  and  $\lambda$  is an attribution of a doxastic attitude, if  $w \xrightarrow{\alpha(\Gamma)} u$  and  $u \xrightarrow{\beta(\lambda)} v$  then  $w \xrightarrow{\beta(\alpha(\Gamma))} v$ . Here,  $\beta$  is a second-order inferential schema, that is, an inferential schema that takes the output of a first-order inferential schema as input. That  $w \xrightarrow{\beta(\alpha(\Gamma))} v$  means that  $v$  is reachable from  $w$  via an application of  $\beta$  to an output of an application of  $\alpha$  to  $\Gamma$  in  $w$ .

When it comes to the intensional representation of second-order inferential schemata, if  $\alpha(\Sigma) = B\phi$  and  $\beta(B\phi) = B\psi$ , then  $\beta(\alpha(\Gamma)) = B\psi$ . Alternatively, we can talk about the transitivity of state-transitions under compositionality of inferential schemata by saying that, where again  $\alpha(\Gamma) = \lambda$  is a particular application of  $\alpha$  and  $\lambda$  is an attribution of a doxastic attitude, if  $\langle w, u \rangle \in i_s(\alpha(\Gamma))$  and  $\langle u, v \rangle \in i_s(\beta(\lambda))$ , then  $\langle w, v \rangle \in i_s(\beta(\alpha(\Gamma)))$ .

Finally, the fourth element of simple models is  $V$ : a function with the same domain as  $dox_s$ , only it maps onto the set  $\{\mathbf{true}, \mathbf{false}\}$ . This is a truth-assignment function that models formulas of  $\mathcal{PL}$ , obeying the usual recursive rules for the truth-values of formulas with the boolean operators. So  $V(w): \Omega \mapsto \{\mathbf{true}, \mathbf{false}\}$ . The function  $V$  establishes which formulas from  $\Omega$  are true and which are false in each member of  $W$ . For our purposes, it really does not matter which atomic formulas from  $\Omega$  are assigned values *true* or *false* by  $V$  in each state: our interest is merely on ‘necessary truths’ of  $\mathcal{PL}$  here. Be  $V$  as it may (if it makes  $p$ ,  $q$ , etc. *true* or *false* is not important), what

<sup>146</sup>We should not confuse *particular applications* of inferential schemata with their *intensional representations*. When we say that  $\alpha(\Gamma) = \lambda$  is a particular application of  $\alpha$  we are saying that the members of  $\Gamma$  are formulas whose main operators are doxastic operators with a subject-index, and that  $\lambda$  is a formula whose main operator is a doxastic operator with a subject-index as well. In intensional representations, however, the input and output variables have no subject-index (and, obviously, no truth-values), and the contents of the doxastic attitudes are represented by means of variables of a language  $\mathcal{L}$ . (We will sometimes use  $\lambda$  as a meta-variable for formulas with doxastic operators, as we did above).

matters here is that it will validate formulas ( $\mathcal{P}\mathcal{L}$ -tautologies) such as  $p \rightarrow p$ ,  $\neg(p \wedge \neg p)$ ,  $(p \wedge q) \leftrightarrow \neg(p \vee \neg q)$ , etc. As we said in the previous chapter, this valuation function has a purely logical function: we need the axioms and theorems from propositional logic to use them in the axiomatics corresponding to our semantics (we will not develop the axiomatics here, but this is clearly needed for such a task — to be pursued in future work).

### 5.2.1 Truth-conditions, consequence and validity

Now let us establish truth-conditions for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  using our simple models. First, we have the following truth-conditions for formulas of  $\mathcal{P}\mathcal{L}$  (that is, formulas that belong to  $\Omega$ ). Where  $M = \langle W, dox_s, i_s, V \rangle$  is a *simple model* and  $\phi \in \Omega$ :

$M, w \models \phi$  if and only if  $V(w)(\phi) = \mathbf{true}$ ;  
 $M, w \not\models \phi$  if and only if  $V(w)(\phi) = \mathbf{false}$ .

Second, we need truth-conditions for formulas attributing doxastic attitudes to the subjects whose situations are represented by simple models. These are pretty straightforward. Where  $M = \langle W, dox_s, i_s, V \rangle$  is a *simple model* and  $\phi \in \Omega$ :

$M, w \models B_s \phi$  if and only if  $dox_s(w)(\phi) = 1$ ;  
 $M, w \models D_s \phi$  if and only if  $dox_s(w)(\phi) = 0$ .

Third, we must establish truth-conditions for formulas with the relative-rationality operator (the truth of formulas with the relative-rationality operator will be used to establish the truth-conditions for formulas with the absolute-rationality operator). Assuming that  $M$  is a simple model over  $\Omega$  (the set of *wffs* of  $\mathcal{P}\mathcal{L}$ ) and  $A$  (the set of inferential schemata whose parameter-language is  $\mathcal{P}\mathcal{L}$ )<sup>147</sup>, that  $\Gamma$  is a set of formulas whose main operators are doxastic operators, and that  $\phi \in \Omega$ :

$M, w \models R(B_s \phi \mid \Gamma)$  if and only if there is an inferential schema  $\alpha$  of any order in  $A$ , where  $\alpha(\Gamma) = B_s \phi$ , and there is a  $w' \in W$  such that  $w \xrightarrow[\alpha(\Gamma)]{} w'$ .<sup>148</sup>

<sup>147</sup>The set  $A$  of inferential schemata over which simple models range, remember, is a set of *optimal* inferential schemata. So no *non-optimal* inferential schema will be part of  $A$ . We will talk about this idealization below.

<sup>148</sup>That  $\alpha$  is an inferential schema of *any order*, remember, means that it is an inferential schema  $\alpha_n(\dots(\alpha_1(\Gamma))) = \lambda$  of order  $n$  (where  $\Gamma$  is a set of doxastic attitudes and  $\lambda$  is a doxastic attitude). In the simplest case, where  $n = 1$ ,  $\alpha$  is a first-order inferential schema whose input-variable is just  $\Gamma$ .



Let us explain these truth-conditions in more detail. We saw above that from the fact that  $w \xrightarrow{\alpha(\Gamma)} w'$  it follows that  $M, w \models \bigwedge \Gamma$ . That means that believing  $\phi$  is rational for  $S$  relative to reasons  $\Gamma$  in state  $w$  only if the doxastic attitudes ascribed to  $S$  by means of the members of  $\Gamma$  are actually held by  $S$  in  $w$ . It cannot be rational for someone to believe something on the basis of a certain set of reasons if one *does not have* those reasons. So, if  $M, w \models R(B_s\phi \mid \Gamma)$  and  $\Gamma = \{B_s\psi_1, \dots, B_s\psi_n\}$  then  $M, w \models (B_s\psi_1 \wedge \dots \wedge B_s\psi_n)$ . On the basis of that consequence relation we can easily prove that, for any  $\phi$  and  $\Gamma$ ,  $M, w \models R(B_s\phi \mid \Gamma) \rightarrow \bigwedge \Gamma$ .

Further, the truth of  $R(B_s\phi \mid \Gamma)$  in  $w$  implies that, from  $w$ ,  $S$  is able to reach a state  $w'$  where  $S$  believes  $\phi$  on the basis of  $\Gamma$  as a result of an instantiation of a certain inferential schema  $\alpha$ , where  $\alpha(\Gamma) = B_s\phi$  is a *particular application* of  $\alpha$  to  $\Gamma$ . That means that we will have  $M, w' \models B_s\phi$  whenever  $M, w \models R(B_s\phi \mid \Gamma)$  and  $w \xrightarrow{\alpha(\Gamma)} w'$  (assuming, as we are, that  $\alpha(\Gamma) = B_s\phi$ ).

So, the presence of an inferential schema of any order such that it takes  $\Gamma$  as input in  $w$  and it returns  $B_s\phi$  as output in  $w'$  is required for the truth of  $M, w \models R(B_s\phi \mid \Gamma)$ . That assures us that a relevant state  $w'$  is *reachable* from state  $w$  for  $S$  via an application of a relevant inferential schema  $\alpha$  that is available to  $S$  in  $w$ . In the simpler and most common case where  $M, w \models R(B_s\phi \mid \Gamma)$ , there will be a *first-order* inferential schema  $\alpha$  such that  $\alpha(\Gamma) = B_s\phi$  and a  $w' \in W$  such that  $w \xrightarrow{\alpha(\Gamma)} w'$ . We could choose to restrict attributions of relative-rationality only to cases where a first-order inferential schema is available, but that would make our semantics too restrictive: it would only attribute rationality in  $w$  to beliefs that can be reached in a *single step* from  $w$ . It is useful for us to consider a case that we already presented above once again.

Let us build a simple model  $M = \langle W, dox_s, i_s, V \rangle$  over  $\Omega$  and  $A$ , representing the situation of a certain subject  $S$ . Assuming that both, the first-order inferential schemata:

$$\begin{aligned} ce(B(\phi \wedge \psi)) &= B\phi, \\ di(B\phi) &= B(\phi \vee \chi) \end{aligned}$$

and the second-order inferential schema:

$$di(ce(B(\phi \wedge \psi))) = B(\phi \vee \chi)$$

are members of  $A$ , and also that  $w, u, v$  are members of  $W$ , here are the specifications of our model:

$$dox_s(w)(p \wedge q) = 1$$

$$\begin{aligned}
& \text{dox}_s(u)(p) = 1 \\
& \langle w, u \rangle \in i_s(\text{ce}(B_s(p \wedge q))) \\
& \langle u, v \rangle \in i_s(\text{di}(B_s p)) \\
& \langle w, v \rangle \in i_s(\text{di}(\text{ce}(B_s(p \wedge q))))
\end{aligned}$$

This is all we need to make our point. Of course, there may be other ordered-pairs in the set  $i_s(\text{ce}(B_s(p \wedge q)))$  than  $\langle w, u \rangle$ , and there may be other ordered-pairs in the set  $i_s(\text{di}(B_s p))$  than  $\langle u, v \rangle$ , although that is not necessarily the case. Also, other formulas may be assigned value 1 in each of these states. But this is not important now. Using our truth conditions for attributions of belief, we can derive:

$$\begin{aligned}
M, w & \models B_s(p \wedge q) \\
M, u & \models B_s p
\end{aligned}$$

Now let us use ‘ $\lambda 1$ ’ to represent  $B_s(p \wedge q)$  and ‘ $\lambda 2$ ’ to represent  $B_s p$ . Given that there is an inferential schema,  $\text{ce}$ , such that  $w \xrightarrow[\text{ce}(\lambda 1)]{} u$  and given that there is an inferential schema,  $\text{di}$ , such that  $u \xrightarrow[\text{di}(\lambda 2)]{} v$ , our truth-conditions for attributions of relative rationality give us:

$$M, w \models R(B_s p \mid B_s(p \wedge q))$$

*Proof.* There is an inferential schema in  $A$ ,  $\text{ce}$ , where  $\text{ce}(B_s(p \wedge q)) = B_s p$ , and a state  $u \in W$  such that  $w \xrightarrow[\text{ce}(\lambda 1)]{} u$ , or  $\langle w, u \rangle \in i_s(\text{ce}(B_s(p \wedge q)))$ . By the truth-conditions for attributions of relative rationality, it follows that, in  $w$ ,  $R(B_s p \mid B_s(p \wedge q))$ .  $\square$

$$M, u \models R(B_s(p \vee r) \mid B_s p)$$

*Proof.* There is an inferential schema in  $A$ ,  $\text{di}$ , where  $\text{di}(B_s p) = B_s(p \vee r)$ , and a state  $v \in W$  such that  $u \xrightarrow[\text{di}(\lambda 2)]{} v$ , or  $\langle u, v \rangle \in i_s(\text{di}(B_s p))$ . By the truth-conditions for attributions of relative rationality, it follows that, in  $u$ ,  $R(B_s(p \vee r) \mid B_s p)$ .  $\square$

Now, we also want to use the fact that  $\text{di}(\text{ce}(B(\phi \wedge \psi))) = B(\phi \vee \chi)$  is such that  $\langle w, v \rangle \in i_s(\text{di}(\text{ce}(B_s(p \wedge q))))$  to show that, in  $w$ , it is rational for  $S$  to believe  $(p \vee r)$  relative to  $S$ ’s reason  $B_s(p \wedge q)$ . That is, we want the following to hold:

$$M, w \models R(B_s(p \vee r) \mid B_s(p \wedge q))$$

But if only ‘one-step’ state transitions are allowed in the truth-conditions for attributions of relative-rationality, then we cannot make the formula  $R(B_s(p \vee r) \mid B_s(p \wedge q))$  true in

$w$  using the model we presented above. That is because neither *ce* nor *di* alone give us what we want. And although  $S$  is not able to reach a state where  $B_s(p \vee r)$  holds in a *single* step from  $w$ , she is able to do so by performing a ‘two-steps’ state transition: when *di* takes the output of *ce* as input,  $S$  goes from a state where  $B_s(p \wedge q)$  holds ( $w$ ) to a state where  $B_s(p \vee r)$  holds ( $v$ ). That means that  $S$  is able to perform an inference from  $B_s(p \wedge q)$  to  $B_s(p \vee r)$  in  $w$ , and that it is (relatively) rational for  $S$  to believe  $(p \vee r)$  on the basis of  $S$ ’s belief that  $(p \wedge q)$  in  $w$ . Therefore, our truth-conditions for attributions of relative-rationality should not be restricted to cases where only first-order inferential schemata are available.

Now that we established the truth-conditions for formulas with the relative-rationality operator, we can establish the truth-conditions for formulas with the absolute-rationality operator. Where  $M = \langle W, dox_s, i_s, V \rangle$  is a simple model over  $\Omega$  and  $A$ :

$M, w \models R(B_s\phi)$  if and only if (i) there is a set of formulas  $\Gamma$  such that  $M, w \models R(B_s\phi \mid \Gamma)$  and (ii) there is no further set  $\Sigma$  such that  $M, w \models R(B_s\neg\phi \mid \Sigma)$  or  $M, w \models R(D_s\phi \mid \Gamma \cup \Sigma)$ .

Here, the first condition is just the requirement that believing  $\phi$  is rational for  $S$  only when there is a set of reasons had by  $S$  such that it is rational for  $S$  to believe  $\phi$  on the basis of those reasons alone. The second one is a ‘no-defeater’ condition.  $S$  will be absolutely justified in believing  $\phi$  only when  $S$  has no reasons to believe  $\neg\phi$ , or reasons that taken together with the reasons  $S$  has to believe  $\phi$  rationalize doubting  $\phi$ . One might wonder why our second condition is not:

... there is no further set  $\Sigma$  such that  $M, w \models R(B_s\neg\phi \mid \Sigma)$  or  $M, w \models R(D_s\phi \mid \Sigma)$ .

Here is why: requiring such a thing would almost always make  $R(B_s\phi)$  false. It is pretty easy for  $S$  to have reasons that rationalize belief in  $\phi$  while  $S$  also has further reasons that taken alone rationalize doubt in  $\phi$  — simply because the content of these reasons ‘has nothing to do’ with  $\phi$ . For example, suppose that  $S$  believes  $(p \wedge q)$  in  $w$  and that  $M, w \models R(B_s p \mid B_s(p \wedge q))$  (say, because there is a  $w'$  such that  $w'$  is reachable from  $w$  via an application of *ce* to  $B_s(p \wedge q)$ ). Suppose, further, that  $S$  also believes  $r$  in  $w$  and that  $M, w \models R(D_s p \mid B_s r)$ . However, it is still not the case that  $M, w \models R(D_s p \mid \{B_s(p \wedge q), B_s r\})$ . Clearly, in this case the justification that  $S$  has to believe  $p$  should not be ‘taken away’ by the irrelevant belief in  $r$  (that is, irrelevant to the belief that  $p$ ).

Finally, we have the following *general* truth-conditions. Where  $M$  is a simple model, for every formulas  $\Phi, \Psi$  of  $\mathcal{E}_{\mathcal{PL}}$ :

$M, w \models \neg\Phi$       *if and only if*  $M, w \not\models \Phi$ ;  
 $M, w \models (\Phi \wedge \Psi)$     *if and only if*  $M, w \models \Phi$  and  $M, w \models \Psi$ ;  
 $M, w \models (\Phi \vee \Psi)$     *if and only if*  $M, w \models \Phi$  or  $M, w \models \Psi$ ;  
 $M, w \models (\Phi \rightarrow \Psi)$  *if and only if*  $M, w \models \neg\Phi$  or  $M, w \models \Psi$ .

These apply to any formula of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  whatsoever — be it a member of  $\Omega$  (a *wff* of  $\mathcal{P}\mathcal{L}$ ) or a formula with the doxastic or the rationality operators. So these are the truth-conditions for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  in simple models. Now we need to work the notions of *simple model consequence* (or *SM-consequence*) and *simple model validity* (or *SM-validity*) out.

Where both  $\Phi$  and  $\Psi$  are formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ ,  $\Psi$  is a SM-consequence of  $\Phi$ , written  $\Phi \models_{SM} \Psi$ , when, for every model  $M = \langle W, dox_s, i_s, V \rangle$ ,  $\Psi$  is true in every state of  $M$  where  $\Phi$  is true. Formally (and equivalently):

$\Phi \models_{SM} \Psi$  *if and only if* there is no simple model  $M$  and a state  $w \in W$  such that  $M, w \models \Phi$  but  $M, w \not\models \Psi$ .

As elsewhere, a formula is valid if it is ‘always’ true. Here, that means that a formula  $\Phi$  of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  is SM-valid, written  $\models_{SM} \Phi$ , if  $\Phi$  is true in every state in every simple model  $M$ :

$\models_{SM} \Phi$  *if and only if* for every simple model  $M$  and every state  $w \in W$ ,  $M, w \models \Phi$ .

In order to prove some of the SM-valid formulas, we will need to use a version of the *deduction theorem*<sup>149</sup>:

**(DT)** *If*  $\Phi \models_{SM} \Psi$  *then*  $\models_{SM} \Phi \rightarrow \Psi$ .

Now let us consider some SM-valid formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ .

### 5.2.2 SM-valid formulas

Before presenting some of the most important SM-valid formulas, let us consider some important features of simple models that we did not consider yet.

First, the function *dox* is constrained by certain rules:

(d1) For no  $w \in W$  and  $\phi \in \Omega$ ,  $dox_s(w)(\phi) = 1$  and  $dox_s(w)(\phi) = 0$ ;

(d2) For no  $w \in W$  and  $\phi \in \Omega$ ,  $dox_s(w)(\phi) = 1$  and  $dox_s(w)(\neg\phi) = 1$ ;

<sup>149</sup>To be more precise, this is actually a *meta-theorem*. See Kleene (1967, p. 39).

---

(d3) For all  $w \in W$  and all  $\phi \in \Omega$ ,  $dox_s(w)(\phi) = 0$  if and only if  $dox_s(w)(\neg\phi) = 0$ ;

The rules constraining *dox* will validate formulas involving the doxastic operators — general principles of belief and doubt. The first rule, (d1), says that an agent will not believe *and* doubt a proposition in the same state (it may well be that, when  $S$  believes  $\phi$  in state  $w$ ,  $S$  will doubt  $\phi$  in a ‘next’ state, though). The idea here is that to say that  $S$  does not doubt  $\phi$  in a certain state is part of what it means to say that  $S$  believes  $\phi$  in that state. The same type of consideration applies to the relation between beliefs and disbeliefs, which is why the rule (d2) is also postulated. This gives rise to the following SM-valid formula:

$$(1) \quad \models_{SM} B_s\phi \rightarrow (\neg D_s\phi \wedge \neg B_s\neg\phi)$$

*Proof.* Let  $M = \langle W, dox_s, i_s, V \rangle$  be an arbitrary simple model. Suppose that  $M, w \models B_s\phi$  for an arbitrary  $w \in W$  and an arbitrary  $\phi \in \Omega$ . By the truth-conditions for formulas with the doxastic operator  $B$ , we have  $dox_s(w)(\phi) = 1$ . By (d1) it follows that it is not the case that  $dox_s(w)(\phi) = 0$ . It also follows that  $M, w \not\models D_s\phi$  (by the truth-conditions for formulas with  $D$ ) and, therefore,  $M, w \models \neg D_s\phi$  (by the general truth-condition for negations). So  $M, w \models B_s\phi \rightarrow \neg D_s\phi$ . Further, given  $dox_s(w)(\phi) = 1$ , it follows by (d2) that it is not the case that  $dox_s(w)(\neg\phi) = 1$ . Given the truth-conditions for attributions of belief again, it follows that  $M, w \not\models B_s\neg\phi$  and, therefore,  $M, w \models \neg B_s\neg\phi$ . So  $M, w \models B_s\phi \rightarrow \neg B_s\neg\phi$ . By simple propositional logic, it follows that  $M, w \models B_s\phi \rightarrow (\neg D_s\phi \wedge \neg B_s\neg\phi)$ . Since that applies for any simple model  $M$ , for every  $w \in W$  and for every  $\phi \in \Omega$ , it follows (by our definition of SM-validity), that  $\models_{SM} B_s\phi \rightarrow (\neg D_s\phi \wedge \neg B_s\neg\phi)$ .  $\square$

Similar considerations about (d1) apply to attitudes of doubt: to say that  $S$  does not believe  $\phi$  in  $w$  is part of what it means to say that  $S$  doubts (or suspends judgment about)  $\phi$  in  $w$ .

The constraint advanced in (d2) rules out the possibility of beliefs in contradictory propositions holding in the *same* state. Rule (d3), in its turn, says that if one doubts  $\phi$  in a state  $w$  one also doubts  $\neg\phi$  in state  $w$ , and vice-versa. The idea is that one does not suspend judgment about a proposition unless one also suspends judgment about the negation of that proposition. This gives rise to the following SM-valid formulas:

$$(2) \quad \models_{SM} D_s\phi \rightarrow (\neg B_s\phi \wedge \neg B_s\neg\phi)$$

$$(3) \quad \models_{SM} D_s\phi \leftrightarrow D_s\neg\phi$$

We will not prove these formulas — the proof for **(1)** is presented mainly for expository purposes. But we invite the reader to prove **(2)**, **(3)** and **(4)** (hint: in order to get **(2)** you need to use both, *(d1)* and *(d3)*). A further SM-valid formula, one that can be proved using only *(d2)*, is:

$$(4) \quad \models_{SM} \neg(B_s\phi \wedge B_s\neg\phi)$$

That is: it is never the case that  $S$  believes  $\phi$  and  $S$  believes  $\neg\phi$ . Nothing here implies, however, that *dox* never assigns 1 to inconsistent formulas. We will get back to this below.

Finally, we have the following rule constraining *dox* — a rule that is directly related to the function  $i_s$  and to the ‘reachability’ relations among states in  $W$ :

$$(d4) \quad \text{For all states } w, w' \in W, \text{ if } w \xRightarrow{\alpha}_{\alpha(\Gamma)} w', \text{ where the intensional representation of } \alpha \text{ is } \alpha(\Sigma) = B\phi, \text{ and } \Gamma = \{B_s\chi_1, \dots, B_s\chi_n\} = si_n(\Sigma) \text{ and } B_s\psi = si_n(B\phi), \text{ then } dox_s(w)(\chi_1) = 1, \dots, dox_s(w)(\chi_n) = 1 \text{ and } dox_s(w')(\psi) = 1.$$

Roughly, *(d4)* says that  $dox_s$  must be coherent with  $i_s$ . When  $i_s$  establishes that  $w'$  is reachable from  $w$  via an application of  $\alpha$  to  $\Gamma = \{B_s\chi_1, \dots, B_s\chi_n\}$ ,  $dox_s$  must be such that it assigns 1 to all members of  $\{\chi_1, \dots, \chi_n\}$  in  $w$ , because  $dox_s$  must make it the case that all doxastic attitudes in  $\Gamma$  are in fact held by  $S$  in  $w$  (the same applies, of course, when  $\Gamma$  contains attitudes of doubt, case in which  $dox_s$  assigns value 0 to the contents of these attitudes). Further, assuming that  $w \xRightarrow{\alpha}_{\alpha(\Gamma)} w'$  and  $\alpha(\Sigma) = B\phi$ , where  $\Gamma = si_n(\Sigma)$  and  $B_s\psi = si_n(B\phi)$ ,  $dox_s$  must assign 1 to  $\psi$  in  $w'$ . For, the fact that  $w \xRightarrow{\alpha}_{\alpha(\Gamma)} w'$  means that  $w'$  is reachable from  $w$  via an application of  $\alpha$  to  $\Gamma$ , and the output of  $\alpha(\Gamma)$  is  $B_s\psi$ .

We just saw one feature of simple models that gives rise to SM-valid formulas **(1–4)**: the rules constraining the function *dox*. Now let us consider two particularly relevant properties that *dox* *does not* have. The first one is that the rules constraining *dox* validate neither the converse of **(1)** nor the converse of **(2)**, that is:

$$\begin{aligned} &\not\models (\neg D_s\phi \wedge \neg B_s\neg\phi) \rightarrow B_s\phi \\ &\not\models (\neg B_s\phi \wedge \neg B_s\neg\phi) \rightarrow D_s\phi \end{aligned}$$

That is because *dox* is not *complete*: it does not need to map *all* members of  $\Omega$  (the set of well-formed formulas of  $\mathcal{PL}$ ) either to 1 or 0. That is, it is not the case that for every  $\phi \in \Omega$ ,  $w \in W$  either  $dox_s(w)(\phi) = 1$  or  $dox_s(w)(\phi) = 0$ . Further, if it is not the case that  $dox_s(w)(\phi) = 1$  and it is not the case that  $dox_s(w)(\phi) = 0$ , it does not follow that  $dox_s(w)(\neg\phi) = 1$ . In other words, there are propositions that the agents modelled by simple models neither believe, nor doubt, nor disbelieve. Presumably, this is a desirable

feature, since we do not want to model agents that are fully opinionated over the formulas in  $\Omega$ .

The second one is that *dox*-valuations are not *logically omniscient*. That means, first, that these valuations are not *transparent*: when  $\phi$  and  $\psi$  are logically equivalent formulas, it does not necessarily follow that if  $dox_s(w)(\phi) = 1$  then  $dox_s(w)(\psi) = 1$ , and it also does not follow that if  $dox_s(w)(\phi) = 0$  then  $dox_s(w)(\psi) = 0$ . When  $dox_s(w)(\phi) = 1$  (or  $dox_s(w)(\phi) = 0$ ) and  $\phi$  is logically equivalent to  $\psi$ , it may be that  $dox_s$  does not decide  $\psi$  (that is, it may be that it does map  $\psi$  either to 1 or to 0). Second, that means that *dox*-valuations are not *closed under logical entailment*. When  $\phi$  logically entails  $\psi$ , it does not necessarily follow that if  $dox_s(w)(\phi) = 1$  then  $dox_s(w)(\psi) = 1$  (same for valuation to 0). Finally, that means that *dox* does not necessarily assign value 1 to tautologies of  $\mathcal{PL}$ . For example, it is not necessarily the case that, for an arbitrary  $w \in W$ ,  $dox_s(w)((\phi \wedge \psi) \rightarrow \neg(\neg\phi \vee \neg\psi)) = 1$ , where  $(\phi \wedge \psi) \rightarrow \neg(\neg\phi \vee \neg\psi)$  is a tautology of  $\mathcal{PL}$ .

Let us consider now another feature of our simple models that will give rise to relevant SM-valid formulas, this time involving our rationality operators. Consider the following inferential schemata that belong to  $A^{150}$  (the set of inferential schemata whose parameter-language is  $\mathcal{PL}$ ):

$$\begin{aligned}
 id(B\phi) &= B\phi \\
 dmp(B\phi, B(\phi \rightarrow \psi)) &= B\psi \\
 ce(B(\phi \wedge \psi)) &= \{B\phi, B\psi\} \\
 ci(B\phi, B\psi) &= B(\phi \wedge \psi) \\
 di(B\phi) &= B(\phi \vee \psi) \\
 dn(B\phi) &= B\neg\neg\phi \\
 nd(B\neg\phi) &= B\neg(\phi \wedge \psi) \\
 dc(D\phi, D\psi) &= D(\phi \wedge \psi) \\
 dd(D\phi, D\psi) &= D(\phi \vee \psi) \\
 d(D\phi) &= D\neg\phi
 \end{aligned}$$

In simple models, whenever a substitution instance of the input-variable of any of these ten inferential schemata holds in a state, there is a further state that is reachable from it: one where a substitution instance of the output-variable of the relevant inferential schema holds. We will call these inferential schemata ‘*always available* inferential schemata’.

In simple models, all the inferential schemata from the list presented above are always

---

<sup>150</sup>Notice that *ce* and *dd* are supposed to output more than one doxastic attitude — up to this point, we have been representing inferential schemata that output a single doxastic attitude.

available. Whenever a state  $w$  makes true attributions of doxastic attitudes that match<sup>151</sup> the input–variable of any of these inferential schemata, there is a state  $w'$  that is reachable from  $w$  via an application of the relevant inferential schema to the relevant doxastic attitudes. There is no *guarantee*, however, that *if* there are two always available inferential schemata such that there is a state  $w$  where some doxastic attitudes hold that match their input–variables, *then* there is a state  $w'$  (one and the same) where some doxastic attitudes hold that match the output–variable of both schemata. On the basis of this feature we will prove that the absolute–rationality operator *does not agglomerate* (see below).

Before presenting some of the formulas that are SM–valid in virtue of this feature (i.e., the ‘always availability’ feature), let us anticipate an idealization of our simple models.

The idealization is the following: each doxastic attitude held by  $S$  in a state  $w$  is *reasonably held by  $S$*  in  $w$ . So  $M, w \models B_s\phi$  means here not only that  $S$  believes  $\phi$  in  $w$ , but also that  $S$  does so reasonably. A belief in  $\phi$  is reasonably held by  $S$  in a state  $w$  when, in  $w$ , there is no set of reasons  $\Gamma$  such that  $R(B_s\neg\phi \mid \Gamma)$ . An attitude of doubt is reasonably held by  $S$  in a state  $w$  when, in  $w$ , there is no set of reasons  $\Gamma$  such that  $R(B_s\phi \mid \Gamma)$  or  $R(B_s\neg\phi \mid \Gamma)$ .

Several questions arise here. For example, suppose that  $M, w \models B_s\phi$  and  $M, w \models B_s\psi$ , but  $\phi$  and  $\psi$  are *inconsistent* with each other (they cannot both be true). As we saw, if  $\phi$  is the *negation* of  $\psi$  (or vice-versa), then  $M$  is not a simple model, because the function *dox* never assigns value 1 to both  $\phi$  and  $\neg\phi$ . So this specific kind of inconsistent value–assignment is ruled out. But if  $\phi$  and  $\psi$  are *inconsistent without being contradictory*, would  $M$  count as a simple model?

That will depend on if the combination of the positive *dox*–valuation of inconsistent formulas with the state–transitions determined by  $i_s$  forces the semantics to make a formula of the type  $(B_s\phi \wedge B_s\neg\phi)$  true. Because that would mean that there is a state  $w' \in W$  and a formula  $\phi \in \Omega$  such that  $dox_s(w')(\phi) = 1$  and  $dox_s(w')(\neg\phi) = 1$  and, as we saw, the rules constraining *dox* does not allow this. So, as long as positive *dox*–valuations of inconsistent formulas does not lead to the type of situation that we just mentioned, we can have simple models  $M$  such that  $M, w \models B_s\phi$  and  $M, w \models B_s\psi$  but  $\phi$  and  $\psi$  are inconsistent with each other. The fact that *dox* assigns value 1 to inconsistent formulas  $\phi$  and  $\psi$  *does not guarantee* that  $M$  is not a simple model (assuming, again, that  $\phi$  is not the negation of  $\psi$  or vice-versa)<sup>152</sup>.

Now the question is: assuming that it is possible (in simple models) for subjects to

<sup>151</sup>A set of formulas  $\Gamma$  matches the input–variable  $\Sigma$  of an inferential schema  $\alpha$  when  $\Gamma = si_n(\Sigma)$ , for some  $n$ .

<sup>152</sup>Some epistemologists argue that rational agents can have justified beliefs (in the same state) toward propositions that form an inconsistent set. See, for example, Klein (1985) and Foley (1979). The conception of simple models is in the same spirit of this proposal.



have beliefs in inconsistent propositions, would not that make it the case that at least one of those beliefs is not reasonably held? Again, not necessarily. For it may be the case that, even if  $\phi$  and  $\psi$  are inconsistent and  $dox_s(w)(\phi) = 1$ ,  $dox_s(w)(\psi) = 1$ , there is no set of reasons  $\Gamma$  such that  $M, w \models R(B_s\neg\phi \mid \Gamma)$  (in particular, assume that it is not the case that  $M, w \models R(B_s\neg\phi \mid B_s\psi)$ ) and no set of reasons  $\Gamma$  such that  $M, w \models R(B_s\neg\psi \mid \Gamma)$  (in particular, assume that it is not the case that  $M, w \models R(B_s\neg\psi \mid B_s\phi)$ ). So, beliefs can be reasonably held in the same state even if their contents are inconsistent — at least as long as they do not rationalize beliefs in the negations of the contents of the other beliefs.

This is just one of the questions that are connected with our assumption that all doxastic attitudes held in our modelled states are *reasonably held*. But let us step these questions aside for now and move on to the validity of epistemic formulas in simple models given that assumption. All of the following formulas are SM-valid (the fact that these formulas are SM-valid is provable using our truth-conditions for attributions of relative-rationality plus the assumption that a relevant inferential schema is available in each state where a certain type of doxastic attitude holds — we will bracket the name of the inferential schema whose availability is used to prove the corresponding SM-valid formula):

- (6)  $\models_{SM} B_s\phi \rightarrow R(B_s\phi \mid B_s\phi)$  (use availability of *id*)
- (7)  $\models_{SM} (B_s(\phi \rightarrow \psi) \wedge B_s\phi) \rightarrow R(B_s\phi \mid \{B_s(\phi \rightarrow \psi), B_s\phi\})$  (use availability of *dmp*)
- (8)  $\models_{SM} B_s(\phi \wedge \psi) \rightarrow R(B_s\phi \mid B_s(\phi \wedge \psi))$  (use availability of *ce*)
- (9)  $\models_{SM} (B_s\phi \wedge B_s\psi) \rightarrow R(B_s(\phi \wedge \psi) \mid \{B_s\phi, B_s\psi\})$  (use availability of *ci*)
- (10)  $\models_{SM} B_s\phi \rightarrow R(B_s(\phi \vee \psi) \mid B_s\phi)$  (use availability of *di*)
- (11)  $\models_{SM} B_s\phi \rightarrow R(B_s\neg\neg\phi \mid B_s\phi)$  (use availability of *dn*)
- (12)  $\models_{SM} B_s\neg\phi \rightarrow R(B_s\neg(\phi \wedge \psi) \mid B_s\neg\phi)$  (use availability of *nd*)
- (13)  $\models_{SM} (D_s\phi \wedge D_s\psi) \rightarrow R(D_s(\phi \wedge \psi) \mid \{D_s\phi, D_s\psi\})$  (use availability of *dc*)
- (14)  $\models_{SM} (D_s\phi \vee D_s\psi) \rightarrow R(D_s(\phi \vee \psi) \mid \{D_s\phi, D_s\psi\})$  (use availability of *dd*)
- (15)  $\models_{SM} D_s\phi \rightarrow R(D_s\neg\phi \mid D_s\phi)$  (use availability of *d*)

Let us prove two of these as an illustration:

*Proof of (7).* Assume  $M, w \models B_s(p \rightarrow q) \wedge B_s p$  for some arbitrary simple model  $M$ ,  $w \in W$  and  $p, q \in \Omega$ . By the general truth-conditions for conjunctions, it follows that  $M, w \models B_s(p \rightarrow q)$  and  $M, w \models B_s p$ . Given that  $\{B_s(p \rightarrow q), B_s p\} = si_n(B(\phi \rightarrow \psi), B\phi)$ , for some  $n$ , it follows that there is a set of doxastic attitudes that hold in  $w$  such that they match the input-variable of  $dmp(B_s(\phi \rightarrow \psi), B_s\phi) = B_s\psi$ . By our assumption that for every such case there is a further state  $w'$  that is reachable from  $w$  via an application of  $dmp$ , where  $M, w' \models B_s q$ , and given our truth-conditions for attributions of relative rationality, it follows that  $M, w \models R(B_s q \mid \{B_s(p \rightarrow q), B_s p\})$ .

Since this type of derivation can be repeated for any simple model  $M$ , with arbitrary state  $w \in W$  and formulas  $\phi, \psi \in \Omega$ , it follows by our definition of SM-consequence that  $B_s(\phi \rightarrow \psi) \wedge B_s\phi \models_{SM} R(B_s\psi \mid \{B_s(\phi \rightarrow \phi), B_s\psi\})$ . Using the deduction theorem (**DT**), it follows that  $\models_{SM} (B_s(\phi \rightarrow \psi) \wedge B_s\phi) \rightarrow R(B_s\phi \mid \{B_s(\phi \rightarrow \psi), B_s\phi\})$ .  $\square$

*Proof of (13).* Assume  $M, w \models D_s p$ ,  $M, w \models D_s q$  for some arbitrary simple model  $M$ ,  $w \in W$  and  $p, q \in \Omega$ . Given that  $\{D_s p, D_s q\} = si_n(D\phi, D\psi)$ , for some  $n$ , it follows that there is a set of doxastic attitudes that hold in  $w$  such that it matches the input-variable of  $dc(D\phi, D\psi) = D(\phi \wedge \psi)$ . By our assumption that for every such case there is a further state  $w'$  that is reachable from  $w$  via an application of  $dc$ , where  $M, w' \models D_s(p \wedge q)$  for some arbitrary  $q \in \Omega$ , and given our truth-conditions for attributions of relative rationality, it follows that  $M, w' \models R(D_s(p \wedge q) \mid \{D_s p, D_s q\})$ .

Since this type of derivation can be repeated for any simple model  $M$  with arbitrary state  $w \in W$  and formulas  $\phi, \psi \in \Omega$ , it follows by our definition of SM-consequence that  $D_s\phi \wedge D_s\psi \models_{SM} R(D_s(\phi \wedge \psi) \mid \{D_s\phi, D_s\psi\})$ . By our deduction theorem (**DT**), it follows that  $\models_{SM} (D_s\phi \wedge D_s\psi) \rightarrow R(D_s(\phi \wedge \psi) \mid \{D_s\phi, D_s\psi\})$ .  $\square$

We leave the proofs of the other validities for the interested reader. A question may arise here if these SM-valid formulas are not *counterintuitive*. Not if we consider the particular properties of simple models. One might read **(9)**, for example, as saying that the fact that one believes two propositions is sufficient for the truth of the claim that it is relatively rational for one to believe the conjunction of those propositions — and that seems clearly wrong. But the beliefs that are supposed to hold in each state  $w$  of a simple model, remember, are supposed to be *reasonably held* by the modelled agent. Further, the consequents of **(6–15)** are formulas with the relative-rationality operator. These formulas do not say that it is rational for  $S$  to believe something *all things considered* (that is, all the reasons held by  $S$  in each state considered).

The SM-valid formulas we just presented express relevant relationships between available reasons, available inferential schemata, and the relative-rationality operator. All of them relate doxastic attitudes that are held by  $S$  with doxastic attitudes that are made relatively rational for  $S$ . In addition to this ‘left-to-right’ relationship between attributions of doxastic attitudes and attributions of relative-rationality, there is a general SM-valid formula expressing a relationship in the other direction:<sup>153</sup>

$$(16) \quad \models_{SM} R(B_s\phi \mid \Gamma) \rightarrow \bigwedge \Gamma$$

*Proof.* Assume  $M, w \models R(B_s p \mid \Gamma)$ , for some arbitrary simple model  $M$ , with  $w \in W$  and  $p \in \Omega$ . By the truth-conditions for formulas with the relative-rationality operator, there is an inferential schema  $\alpha$  in  $A$ , where  $\alpha(\Gamma) = B_s p$ , and there is a  $w' \in W$  such that  $w \xRightarrow[\alpha(\Gamma)]{=} w'$ . But  $w \xRightarrow[\alpha(\Gamma)]{=} w'$  only if  $M, w \models \bigwedge \Gamma$ . Since this applies for every  $w \in W$  and  $\phi \in \Omega$  such that  $M, w \models R(B_s\phi \mid \Gamma)$ , it follows that  $R(B_s\phi \mid \Gamma) \models_{SM} \bigwedge \Gamma$ . By **(DT)**, then, it follows that  $\models_{SM} R(B_s\phi \mid \Gamma) \rightarrow \bigwedge \Gamma$ .  $\square$

Now let us consider some SM-valid formulas with the absolute-rationality operator. Using the truth-conditions for formulas with our rationality operators, we easily obtain:

$$(17) \quad \models_{SM} R(B_s\phi) \rightarrow (\neg R(B_s\neg\phi) \wedge \neg R(D_s\phi))$$

$$(18) \quad \models_{SM} R(D_s\phi) \rightarrow (\neg R(B_s\phi) \wedge \neg R(B_s\neg\phi))$$

$$(19) \quad \models_{SM} R(B_s\phi) \rightarrow \neg R(B_s\neg\phi \mid \Gamma)$$

(A good way to prove these is to assume their antecedents to hold in some simple model and to assume, for *reductio*, the negations of their consequents to hold in that model).

One might wonder if there are ‘more interesting’ formulas involving the absolute-rationality operator to be proved here. In particular, one may ask if we can prove the validity of some version of a *closure principle*<sup>154</sup> involving that operator, or the validity of a version of *agglomeration* involving that operator. Let us begin with the latter.

Agglomeration of the absolute-rationality operator *is not* SM-valid, that is:

<sup>153</sup>In the following proof, although  $\Gamma$  is used as a *free-variable* in the schema, it is used as a *name* in the proof of the schema. We will adhere to this not strictly correct but harmless practice in other proofs involving the relative-rationality operator as well.

<sup>154</sup>For different versions of so-called ‘closure principles’ of justification in contemporary epistemological literature, see Hales (1995).

$$\not\models_{SM} (R(B_s\phi) \wedge R(B_s\psi)) \rightarrow R(B_s(\phi \wedge \psi))$$

We can see this by considering the possibility of there having a simple model  $M$  such that  $M, w \models R(B_s p) \wedge R(B_s q)$  but  $M, w \models \neg R(B_s(p \wedge q))$ , for some  $w \in W$  and  $p, q \in \Omega$ . Given  $M, w \models R(B_s p) \wedge R(B_s q)$ , we have  $M, w \models R(B_s p)$ . That means that there is a set of doxastic attitudes  $\Gamma$  such that  $M, w \models R(B_s p \mid \Gamma)$ . Therefore, there is an inferential schema  $\alpha$  such that  $w \xRightarrow{\alpha(\Gamma)} w'$  for some  $w' \in W$  where  $M, w' \models B_s p$ . Further, given  $M, w \models \neg R(B_s(p \wedge q))$  again, we have  $M, w \models R(B_s q)$  as well. That means that there is a set of doxastic attitudes  $\Gamma'$  such that  $M, w \models R(B_s q \mid \Gamma')$ . Therefore, there is an inferential schema  $\beta$  such that  $w \xRightarrow{\beta(\Gamma')} w''$  for some  $w''$  where  $M, w'' \models B_s q$ . Now, we could get  $M, w \models R(B_s(p \wedge q))$  if there were any guarantee that  $w' = w''$ , because then we could say that there is a further state  $u$  such that  $w' \xRightarrow{ci(\Sigma)} u$  (or, what would be the same, that  $w'' \xRightarrow{ci(\Sigma)} u$ ), where  $\Sigma = \{B_s p, B_s q\}$ . However, *there is no such guarantee*. As far as we can tell,  $S$  may not be able to reach a state from  $w$  where she believes *both*  $p$  and  $q$  ( $S$  may not be able to ‘put two and two together’), although  $S$  is able to reach a state from  $w$  where she believes  $p$  and  $S$  is able to reach a state from  $w$  where she believes  $q$ .

The *converse* of the principle of agglomeration, however, can be proved to be SM-valid. Before proving the converse, however, let us present the following SM-consequence schema:

$$(C1) \quad R(B_s(\phi \wedge \psi) \mid \Gamma) \models_{SM} R(B_s\phi \mid \Gamma) \wedge R(B_s\psi \mid \Gamma)$$

*Proof.* Suppose  $M, w \models R(B_s(p \wedge q) \mid \Gamma)$  for some arbitrary  $M, w \in W$  and  $p, q \in \Omega$ . That means that there is an inferential schema  $\alpha$  such that  $w \xRightarrow{\alpha(\Gamma)} w'$ , for some  $w' \in W$ . It follows, then, that  $M, w' \models B_s(p \wedge q)$ . By our assumption that  $ce(B_s(\phi \wedge \psi)) = \{B_s\phi, B_s\psi\}$  is always available, and given that  $B_s(p \wedge q) = sin(B_s(\phi \wedge \psi))$  for some  $n$ , it follows that there is a further state  $w''$  such that  $w' \xRightarrow{ce(\lambda)} w''$ , where ‘ $\lambda$ ’ represents  $B_s(p \wedge q)$ , and  $M, w'' \models B_s p$ ,  $M, w'' \models B_s q$ <sup>155</sup>. Given the transitivity of state-transitions under compositionality of inferential schemata, we have  $w \xRightarrow{ce(\alpha(\Gamma))} w''$ . By the truth-conditions for attributions of relative-rationality, we have  $M, w \models R(B_s p \mid \Gamma)$  and  $M, w \models R(B_s q \mid \Gamma)$ . Since this can be repeated for any model  $M$  with  $w, w', w'' \in W$  and arbitrary  $\phi, \psi \in \Omega$ , we have  $R(B_s(\phi \wedge \psi) \mid \Gamma) \models_{SM} R(B_s\phi \mid \Gamma) \wedge R(B_s\psi \mid \Gamma)$ .  $\square$

Now we can show the converse of the principle of agglomeration to be SM-valid:

<sup>155</sup>It really makes no difference if we assume that  $B_s p$  holds in a state and  $B_s q$  holds in a different one — we will reach the very same conclusions.

$$(E) \quad \models_{SM} R(B_s(\phi \wedge \psi)) \rightarrow (R(B_s\phi) \wedge R(B_s\psi))$$

*Proof.* Suppose  $M, w \models R(B_s(p \wedge q))$  for some arbitrary  $M, w \in W$  and  $p, q \in \Omega$ . That means there is a set of doxastic attitudes  $\Gamma$  such that  $M, w \models R(B_s(p \wedge q) \mid \Gamma)$ . By (C1), it follows that  $M, w \models R(B_s p \mid \Gamma)$  and  $M, w \models R(B_s q \mid \Gamma)$ . Given our initial assumption that  $M, w \models R(B_s(p \wedge q))$ , it follows by the truth-conditions for formulas with the absolute-rationality operator that there is no set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg(p \wedge q) \mid \Sigma)$ . Therefore, there is no set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg p \mid \Sigma)$  (see proof below). And given, again, our assumption that  $M, w \models R(B_s(p \wedge q))$ , we have by the truth-conditions for formulas with the absolute-rationality operator that there is no set  $\Sigma$  such that  $M, w \models R(D_s(p \wedge q) \mid \Gamma \cup \Sigma)$ . It follows that there is not set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(D_s p \mid \Gamma \cup \Sigma)$  (see proof below). Therefore,  $M, w \models R(B_s p)$ . Since the same procedure can be repeated using  $q$ , we also have  $M, w \models R(B_s q)$ . Therefore,  $M, w \models R(B_s p) \wedge R(B_s q)$ . As this can be repeated for any model  $M$ , with  $w \in W$  and arbitrary  $\phi, \psi \in \Omega$ , it follows that  $R(B_s(\phi \wedge \psi)) \models_{SM} R(B_s\phi) \wedge R(B_s\psi)$ . Finally, by the deduction theorem (DT) it follows that  $\models_{SM} R(B_s(\phi \wedge \psi)) \rightarrow (R(B_s\phi) \wedge R(B_s\psi))$ .  $\square$

Now let us prove the ‘intermediary steps’ used in the proof above. First, let us prove that *if* there is no set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg(\phi \wedge \psi) \mid \Sigma)$ , for some  $M, w \in W$  and  $\phi, \psi \in \Omega$ , *then* there is no set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg\phi \mid \Sigma)$ . We will prove this by proving the converse of the above conditional: *if* there is a set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg\phi \mid \Sigma)$ , for some  $M, w \in W$  and  $\phi, \psi \in \Omega$ , *then* there is a set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg(\phi \wedge \psi) \mid \Sigma)$ . And this is obviously proven when we prove the following SM-consequence schema:

$$(C2) \quad R(B_s \neg\phi \mid \Sigma) \models_{SM} R(B_s \neg(\phi \wedge \psi) \mid \Sigma)$$

*Proof.* Suppose that there is a set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(B_s \neg p \mid \Sigma)$ , for some  $M, w \in W$  and  $p \in \Omega$ . That means that there is an inferential schema  $\alpha$  such that  $w \xrightarrow[\alpha(\Sigma)]{\implies} w'$ , where  $M, w' \models B_s \neg p$ . By our assumption that  $nd(B \neg\phi) = B \neg(\phi \wedge \psi)$  is always available, and given that  $B_s \neg p = si_n(B \neg\phi)$  for some  $n$ , it follows that there is a further state  $w''$  such that  $w' \xrightarrow[nd(\lambda)]{\implies} w''$ , where ‘ $\lambda$ ’ represents  $B_s \neg p$ , and  $M, w'' \models B_s \neg(p \wedge q)$  for some  $q \in \Omega$ . Given the transitivity of state-transitions under compositionality of inferential schemata, we have  $w \xrightarrow[nd(\alpha(\Sigma))]{\implies} w''$ . By the truth-conditions for formulas with the relative-rationality operator, we have  $M, w'' \models R(B_s \neg(p \wedge q) \mid \Sigma)$ . Since this can

be repeated for any model  $M$  with  $w, w', w'' \in W$  and arbitrary  $\phi, \psi \in \Omega$  we have  $R(B_s \neg \phi \mid \Sigma) \models_{SM} R(B_s \neg(\phi \wedge \psi) \mid \Sigma)$ .  $\square$

Next, we prove that *if* there is no set  $\Sigma$  such that  $M, w \models R(D_s(\phi \wedge \psi) \mid \Gamma \cup \Sigma)$ , for some  $M, w \in W$  and  $\phi, \psi \in \Omega$ , *then* there is not set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(D_s \phi \mid \Gamma \cup \Sigma)$ . We will prove this by proving the converse of the above conditional: *if* there is a set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(D_s \phi \mid \Gamma \cup \Sigma)$ , for some  $M, w \in W$  and  $\phi, \psi \in \Omega$ , then there is a set of doxastic attitudes  $\Sigma$  such that  $M, w \models R(D_s(\phi \wedge \psi) \mid \Gamma \cup \Sigma)$ . And this is obviously proven when we prove the following SM-consequence schema:

$$(C3) \quad R(D_s \phi \mid \Gamma) \models_{SM} R(D_s(\phi \wedge \psi) \mid \Gamma)$$

*Proof.* Suppose that there is a set of doxastic attitudes  $\Gamma$  such that  $M, w \models R(D_s p \mid \Gamma)$ , for some  $M, w \in W$  and  $p \in \Omega$ . That means that there is an inferential schema  $\alpha$  such that  $w \xrightarrow{\alpha(\Sigma)} w'$ , where  $M, w' \models D_s p$ . By our assumption that  $dc(D\phi) = D(\phi \wedge \psi)$  is always available, and given that  $D_s p = si_n(D\phi)$  for some  $n$ , it follows that there is a further state  $w''$  such that  $w' \xrightarrow{dc(\lambda)} w''$ , where ' $\lambda$ ' represents  $D_s p$ , and  $M, w'' \models D_s(p \wedge q)$  for some  $q \in \Omega$ . Given the transitivity of state-transitions under compositionality of inferential schemata, we have  $w \xrightarrow{dc(\alpha(\Gamma))} w''$ . By the truth-conditions for formulas with the relative-rationality operator, we have  $M, w'' \models R(D_s(p \wedge q) \mid \Gamma)$ . Since this can be repeated for any model  $M$  with  $w, w', w'' \in W$  and arbitrary  $\phi, \psi \in \Omega$  we have  $R(D_s \phi \mid \Gamma) \models_{SM} R(D_s(\phi \wedge \psi) \mid \Gamma)$ .  $\square$

That concludes our full proof of the SM-validity of schema **(E)**.

We just saw that agglomeration is not SM-valid, although its converse, schema **(E)**, is SM-valid. What about closure? The rough version  $(R(B_s \phi) \wedge (\phi \rightarrow \psi)) \rightarrow R(B_s \psi)$  is clearly not SM-valid, since its antecedent does not even make it true that  $S$  believes both,  $\phi$  and  $(\phi \rightarrow \psi)$  in some particular state, which would guarantee a reachable state where  $B_s \psi$  holds, in virtue of the availability of  $dmp$ . So let us consider two modified versions of that schema:  $(R(B_s \phi) \wedge R(B_s(\phi \rightarrow \psi))) \rightarrow R(B_s \psi)$  and  $R(B_s(\phi \wedge (\phi \rightarrow \psi))) \rightarrow R(B_s \psi)$ .

Now, the same feature of simple models that leads to the SM-invalidity of agglomeration leads also to the SM-invalidity of the first schema:

$$\not\models_{SM} (R(B_s \phi) \wedge R(B_s(\phi \rightarrow \psi))) \rightarrow R(B_s \psi)$$

That is, *there is no guarantee* that the state that is reachable for  $S$  where  $B_s \phi$  holds is the same state as the one that is reachable for  $S$  where  $B_s(\phi \rightarrow \psi)$  holds. When the antecedent

of the above schema is satisfied, there is no guarantee that there is a single state where both  $B_s\phi$  and  $B_s(\phi \rightarrow \psi)$  are true and, therefore, there is no guarantee that a state is reachable for  $S$  where  $B_s\psi$  is true. Here is another way to put it: given the SM-invalidity of agglomeration, the antecedent of the schema above,  $R(B_s\phi) \wedge R(B_s(\phi \rightarrow \psi))$ , does not entail  $R(B_s(\phi \wedge (\phi \rightarrow \psi)))$ . And if this is not entailed, we have no reason to think that  $S$  will *necessarily* ‘put two and two together’ in some relevant state.

Perhaps, then, if there were a guarantee that  $R(B_s(\phi \wedge (\phi \rightarrow \psi)))$  holds in a state, we could draw the desired conclusion  $R(B_s\psi)$  by using the availability of  $dmp$ . But not even that can be proven to always be the case. In order to see why, suppose we were trying to prove the following (Warning! We are not stating that the schema below is SM-valid):

$$\models_{SM} R(B_s(\phi \wedge (\phi \rightarrow \psi))) \rightarrow R(B_s\psi)$$

We should proceed as before. First, we assume an instantiation of the antecedent to be true for some  $M, w \in W$  and  $p, q \in \Omega$ . Then, we conclude that  $M, w \models R(B_s(p \wedge (p \rightarrow q)) \mid \Gamma)$  for some  $\Gamma$  and, therefore, that there is a state  $w'$  reachable from  $w$  where  $M, w' \models B_s(p \wedge (p \rightarrow q))$ . By using **(E)** and the availability of  $dmp$ , we conclude that there is a state  $w''$  where  $M, w'' \models B_sq$ . Up to this point, we can conclude that  $M, w \models R(B_sq \mid \Gamma)$ , and *this is all we can conclude* about the epistemic status of  $B_sq$ .

That is so because, in order for us to prove that  $M, w \models R(B_sq)$ , we would need to prove that there is no  $\Sigma$  such that  $M, w \models R(B_s\neg q \mid \Sigma)$  or  $M, w \models R(D_sq \mid \Gamma \cup \Sigma)$ . And the only way to do that is to prove the following intermediary step: if there is no  $\Sigma$  such that  $R(B_s\neg(\phi \wedge (\phi \rightarrow \psi)) \mid \Sigma)$ , then there is no  $\Sigma$  such that  $R(B_s\neg\psi \mid \Sigma)$ . We would get this by proving, as before, the converse of the above conditional: if there is a  $\Sigma$  such that  $R(B_s\neg\psi \mid \Sigma)$ , then there is a  $\Sigma$  such that  $R(B_s\neg(\phi \wedge (\phi \rightarrow \psi)) \mid \Sigma)$ . But that would require the following inferential schema to be always available:

$$nc(B\neg\psi) = B(\neg(\phi \wedge (\phi \rightarrow \psi)))$$

The inferential schema  $nc$ , however, is not in the list of inferential schemata that are ‘always available’.

The criterion that we used to list those ten inferential schemata is that they are relatively simple, and it seems that every minimally rational creature capable of having beliefs and doubts is able to instantiate them. But  $nc$  is a more complex inferential schema, and it is not one whose availability seems to be required for a minimally rational creature. This, of course, is mere speculation based on our ‘sense of simplicity’, and it appears to add some *ad hoc-ness* to the structure of simple models — our choice of those ten inferential

schemata is not backed up by up-to-date cognitive psychology or anything like that.

But at this point we are not looking for *empirical adequacy* when it comes to which inferential schemata should count as available to minimally rational creatures. We are expecting our semantics for attributions of epistemic rationality to have, remember, another kind of adequacy: the one we explained in the previous chapter using our three *criteria* (roughly, rationality must be a function of reasons and abilities, there should be no contradictory rationality judgments, and there should be a substantive overlap between the ‘judgments’ of our formal semantics and the judgments of competent speakers of English).

So, in virtue of our choice of the inferential schemata that are always available, the closer we get to a closure principle of rationality is **(7)** presented above. Of course, *when*  $nc$  makes a state  $w'$  available, that is, when  $\langle w, w' \rangle \in i_s(nc(B\neg q))$  for some  $w$  and  $q$ , we may prove that  $M, w \models R(B_s(p \wedge (p \rightarrow q))) \rightarrow R(B_s q)$ . But that will hold only in particular situations where  $nc$  is available (that is, only for particular simple models  $M$ , states  $w, w' \in W$  and formulas  $p, q \in \Omega$ ).

There are many more things that we can prove — for example, we can prove that  $\models_{SM} \neg R(B_s(\phi \wedge \neg \phi))$ . We cannot prove, however, versions of **(6–15)** with the absolute-rationality operator in the antecedent and the consequent, for example:

$$R(B_s \phi) \rightarrow R(B_s(\phi \vee \psi)) \quad (\text{modified version of (10)})$$

$$R(D_s \phi) \rightarrow R(D_s(\phi \wedge \psi)) \quad (\text{modified version of (13)})$$

etc.

We could prove these only if other, more sophisticated inferential schemata were assumed to be always available. In fact, we could get a whole system with valid conditional formulas of the forms  $R(B_s \ ) \rightarrow R(B_s \ )$  and  $R(D_s \ ) \rightarrow R(D_s \ )$  with a one-to-one correspondence with valid conditional formulas of  $\mathcal{PL}$  (the blank spaces in the schemata above would be filled with the antecedents and consequents of the valid conditional formulas of  $\mathcal{PL}$ ). But that would create an inconsistent framework (one that outputs both, schemas and negations of schemas) — unless we make the further requirement that *dox*-valuations must be ‘consistent’ (the requirement would be: for no inconsistent pair of formulas  $\phi$  and  $\psi$   $dox(\phi) = 1$  and  $dox(\psi) = 1$ ; for all inconsistent pair of formulas  $\phi$  and  $\psi$ , if  $dox(\phi) = 0$  then  $dox(\psi) = 0$ ). Actually, in this way we would get both, the validity of the agglomeration schema:  $(R(B_s \phi) \wedge R(B_s \psi)) \rightarrow R(B_s(\phi \wedge \psi))$ , and the validity of the closure schemas:  $(R(B_s \phi) \wedge (\phi \rightarrow \psi)) \rightarrow R(B_s \psi)$  and  $(R(B_s \phi) \wedge R(B_s(\phi \rightarrow \psi))) \rightarrow R(B_s \psi)$ .

In this case, we would not even need the relative-rationality operator, and our possible-states semantics would start looking more like a canonical possible-worlds semantics. That would be, of course, one alternative. But it is not an alternative in view of our commitment



with the idea of modelling ‘resource–bounded’ reasoners, that is, reasoners whose reasoning capacities are not complete in relation to a given axiomatic system. One can see our simple models as attempts to model minimally rational reasoners — reasoners whose inferential abilities do not transcend those ten simple inferential schemata (that does not mean, of course, that simple models *cannot* assign a non–empty set of ordered–pairs of states to other inferential schemata than those ten).

### 5.3 Level of idealization and adequacy of simple models

There is a number of *idealizations* in our simple models. Let us consider each of these in turn. First, simple models are models over *optimal* inferential schemata only. That means that *non–optimal* inferential schemata are not in the scope of our models. It is plausible to think, however, that there are *non–optimal* inferential schemata available to most real–world reasoners (and that real–world reasoners instantiate such *non–optimal* schemata). As a result, we are modelling *good* (although limited) reasoners only — *bad* reasoners are not modelled by simple models. Call this the ‘*competence idealization*’.

Second, the epistemic judgments made true by simple models are restricted to cases where doxastic attitudes are made justified for subjects in virtue of simple logical relations — the ones studied in classical *propositional logic* — between the contents of pre–inferential beliefs and the contents of inferential beliefs. This is clearly an undesirable idealization, since most types of inferential schemata we are interested in using and representing are not included in the scope of the function  $i_s$  of simple models. For example, we cannot properly assess inferential schemata whose contents are logically related at the intra–sentential level (relations studied by *first–order logic*), or inferential schemata whose contents form inductive, probabilistic or explanatory arguments, or inferential schemata whose contents include modal operators, etc. Call this the ‘*propositional–level idealization*’.

Third, we are assuming that  $dox_s$  does not make value–assignments to the members of  $\Omega$  in such a way as to give rise to *explicitly conflicting* doxastic attitudes ( $B_s\phi$  explicitly conflicts with both  $B_s\neg\phi$  and  $D_s\phi$ ). We have a requirement of ‘doxastic non–contradiction’ here. As a result, in no state of a simple model an agent forms a belief in a contradiction. Of course, that does not put simple models on a par with canonical possible–worlds models for doxastic operators. We have no closure of belief in simple models, and no requirement of consistency for every possible (reachable) state is required. So, here, we have less idealization than usual. However, one might think that it is at least possible for an agent to form a belief in a contradiction, or that it is at least possible for an agent to believe and disbelieve a certain proposition. Simple models exclude that possibility, though. Call

this the ‘*no-contradiction idealization*’.

Fourth, the ‘defeat relations’ taken into account by the truth-conditions for attributions of absolute-rationality in simple models is problematic. (Call this the ‘*defeat idealization*’). Consider again our truth-conditions for formulas with the absolute-rationality operator:

$M, w \models R(B_s\phi)$  if and only if (i) there is a set of formulas  $\Gamma$  such that  $M, w \models R(B_s\phi \mid \Gamma)$  and (ii) there is no further set  $\Sigma$  such that  $M, w \models R(B_s\neg\phi \mid \Sigma)$  or  $M, w \models R(D_s\phi \mid \Gamma \cup \Sigma)$ .

The relevant defeat relations are taken into account by condition (ii). There are three problems with (ii). The first one is that the ‘defeaters’ figuring in (ii) must themselves be *reachable*. So if it is relatively rational for  $S$  to believe  $\phi$  and  $S$  has a reason to believe  $\neg\phi$  in  $w$  but  $S$  is not able to infer the latter (there is no state reachable from  $w$  in which  $S$  forms a belief in  $\neg\phi$  on the basis of  $S$ ’s reasons in  $w$ ), then our semantics will make true the formula  $R(B_s\phi)$  in  $w$ . One might object to this however, since it may appear that counter-evidence for  $\phi$  can defeat  $S$ ’s justification to believe  $\phi$  even if  $S$  is not able to infer  $\neg\phi$ . So a defense of the truth-conditions presented above would require philosophical argumentation to explain away that intuition.

The second problem is that (ii) rules out attributing absolute-rationality to  $B_s\phi$  when  $S$  has a *defeated defeater* for believing  $\phi$ : the presence of a single reachable defeater (it does not matter if it is ‘neutralized’ or not) suffices to take away justification. Assume that  $\Sigma$  is the *total set* of reasons available to  $S$  in state  $w$ . Now suppose that there is a subset  $\Gamma \subset \Sigma$  such that  $M, w \models R(B_s p \mid \Gamma)$  and a subset  $\Gamma' \subset \Sigma$  such that  $M, w \models R(B_s \neg p \mid \Gamma')$ . According to the truth-conditions presented above, we have  $M, w \not\models R(B_s p)$ . But there are at least two scenarios that could be conceived here such that they would make it the case that *it is* rational for  $S$  to believe  $\phi$ . In the first scenario, the *degree of justification* that  $\Gamma$  confers upon  $B_s p$  is substantially bigger than the degree of justification that  $\Gamma'$  confers upon  $B_s \neg p$  (whatever interval is being used to measure degrees of justification). In the second one,  $S$  has a further set of reasons  $\Gamma'' (\neq \Gamma)$  such that it ‘neutralizes’ the justification conferred upon  $B_s \neg p$  by  $\Gamma'$ , in the sense that while  $R(B_s \neg p \mid \Gamma')$  is true, it is also true that  $R(D_s \neg p \mid \Gamma' \cup \Gamma'')$ . Simple models cannot take the first scenario into account because there is no way of computing *degrees of justification* using simple models. The second scenario *can* be taken into account, but at the cost of adding a significant layer of complexity to the decidability of formulas of the type  $R(B_s \_)$ , because then we would have to determine not only if there are undefeated defeaters, but also if there are undefeated undefeated defeaters and so on.

Finally, we have the ‘*reasonability idealization*’ that we already mentioned above: every doxastic attitude held in a state of a simple model is *reasonably held*. So we have five important idealizations here: the *competence idealization*, the *propositional-level idealization*, the *no-contradiction idealization*, the *defeat idealization* and the *reasonability idealization*. How these idealizations relate to our *criteria* of adequacy for a semantics for rationality attributions that were fleshed out in the previous chapter? Let us consider each of those *criteria*.

Our first criterion of adequacy is, remember:

- (*Cr1*) The truth or falsity of formulas that attribute/deny rationality to doxastic attitudes for a certain subject  $S$  in a certain state should be a function of two things: the reasons available to  $S$  in that state and the inferential abilities possessed by  $S$  in that state.

Does the type of possible-states semantics constituted by simple models satisfy (*Cr1*)? It is not obvious that, in simple models, rationality is conceived as a function of available reasons *in the way we want*. We want the relevant reasons to have a content that *gives support to* the content of the target belief. From the perspective of simple models, however, the relation among reasons and what they rationalize is purely formal: the only type of support that is taken into account here is (classical) logical consequence determined by relations between *propositional forms* (not between the propositions themselves, with a particular probability). Given that simple models are models over both, formulas of  $\mathcal{PL}$  and inferential schemata whose parameter-language is  $\mathcal{PL}$ , whatever lies outside *propositional logic*, when it comes to support relations, is not captured by simple models.

It is also not obvious that, in simple models, rationality is conceived as a function of inferential abilities *in the way we want*. Abilities to reason about objects and their properties using quantifiers and abilities to reason about probabilities/modalities are not modelled by simple models. So, when it comes to (*Cr1*), we can say that a possible-states semantics based on simple models fails to include all the relevant support relations and inferential abilities (not that it goes wrong by including the support relations and abilities that it does). We could change that and get other families of models (to constitute other types of *ps-semantics*) by giving up the propositional-level idealization.

Now let us consider our second criterion of adequacy:

- (*Cr2*) A semantics for rationality attributions should not make true both, a formula that attributes rationality to a doxastic attitude for a certain subject in a certain state *and* its negation.

Suppose  $M, w \models R(B_s\phi)$  where  $M$  is a simple model with  $w \in W$  and  $\phi$  is an arbitrary formula in  $\Omega$ . It follows by the truth-conditions for formulas with the absolute-rationality operator that there is a set of doxastic attitudes  $\Gamma$  such that  $M, w \models R(B_s\phi \mid \Gamma)$  and there is no further set  $\Sigma$  such that  $M, w \models R(B_s\neg\phi \mid \Sigma)$  or  $M, w \models R(D_s\phi \mid \Gamma \cup \Sigma)$ . Now suppose that  $M, w \models \neg R(B_s\phi)$ , that is, that  $M, w \not\models R(B_s\phi)$ . If that is the case, then either there is no set  $\Gamma$  such that  $M, w \models R(B_s\phi \mid \Gamma)$ , which contradicts the first assumption, or there is such a set but there is also a further set  $\Sigma$  such that  $M, w \models R(B_s\neg\phi \mid \Sigma)$  or  $M, w \models R(D_s\phi \mid \Gamma \cup \Sigma)$ , which also contradicts the first assumption. Therefore, our *ps-semantics* for formulas of  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$  constituted by simple models is in accordance with (Cr2).

Now, we take it that, assuming that a certain semantics for attributions of rationality satisfies (Cr1) and (Cr2) (even if satisfaction of (Cr1) is restricted to a specific language and type of support), the really important question is if it satisfies our third criterion of adequacy:

- (Cr3) A semantics for attributions of rationality should make true/false formulas that attribute/deny rationality to doxastic attitudes in such a way as to agree with the judgments of competent speakers of English about *non-problematic* cases.

This criterion will show that the particular *ps-semantics* we have been building is highly inadequate. Of course, from the armchair we are able to see that simple models will make attributions of rationality true in several situations where competent speakers would attribute rationality — a certain level of agreement between our semantics and the judgments of English-speakers is guaranteed to obtain. For example, when  $S$  has  $B_s(p \wedge q)$  as an undefeated reason, the semantics of simple models will make  $R(B_s p)$  true. However, there is a wide range of cases about which our semantics will get things wrong. Given the *defeat idealization*, we will have disagreement between the ‘judgments’ of this particular *ps-semantics* and the judgments of competent English-speakers over a wide range of cases (cases where one has a ‘defeated defeater’ for a certain belief). Further, in virtue of the *propositional-level idealization* we will also have disagreement over a whole class of beliefs that are taken to be justified in virtue of inductive, probabilistic or explanatory relations of support by competent English-speakers.

In both cases we will have disagreement between our semantics and the judgments of competent speakers, in that several cases that are taken to be cases of rational belief by the relevant speakers are not taken to be cases of rational belief by the semantics. But this is not the only type of ‘overlap failure’. There may be disagreement in the other direction as well: some cases that are judged to be cases of rational belief in the relevant *ps-semantics* may not be regarded as cases of rational belief by competent English-speakers. That may be so because, in simple models, we always assume that the doxastic attitudes that are

held in a certain doxastic are *reasonably held*, and that such a status suffices for a doxastic attitude to count as a reason for believing something else (see SM-valid formulas (6–15)). But some speakers may judge that reasonability *is not* a sufficient epistemic status for a doxastic attitude to count as a (good) reason for other doxastic attitudes: there are many beliefs for which one has no *counter-evidence*, but also no *positive evidence* whatsoever. It is doubtful that some such doxastic attitudes can play the role of being ‘justifiers’ for further doxastic attitudes.

So, in principle, we would be entitled to conclude that this particular family of *ps-semantics* (the one built from simple models) for attributions of epistemic rationality is *inadequate* (according to the *criteria* of adequacy we fleshed out in *Chapter 4*). Given that much, let us explore what can we conclude about the *type* of semantics (*ps-semantics*) that we have been considering.

#### 5.4 Inconclusive conclusion about our *ps-semantics*

We just concluded that a particular version of a *ps-semantics* — one based built from *simple models* — is not adequate. But perhaps it would be more accurate to say: ‘this particular type of *ps-semantics* (the one built from simple models) is inadequate *as a completely general semantics* for attributions of epistemic rationality’. Notice that simple models are models for formulas of  $\mathcal{E}_{\mathcal{PL}}$ , *and those formulas only*. It is reasonable to judge, then, that models for  $\mathcal{E}_{\mathcal{PL}}$  must be assessed in comparison with attributions of rationality in virtue of logical relations at the propositional level — not with attributions of rationality in virtue of *any* type of support relation.

Here is another way to put it: simple models are models of simple, not necessarily sophisticated reasoners that can reason on the basis of ‘coarse-grained’ relations among formulas with boolean operators. Call these reasoners ‘boolean-reasoners’. So *we should not* use simple models to model situations of agents whose doxastic attitudes are made rational or justified in virtue of more complex and richer inferential relations. Accordingly, there would be versions of *possible-states semantics* with models for ‘predicate-reasoners’ (the contents of the doxastic attitudes are represented by formulas of *first-order logic* ( $\mathcal{FOL}$ ), and the models range over inferential schemata whose parameter-language is  $\mathcal{FOL}$ ), ‘probability-reasoners’ (the contents of the doxastic attitudes are represented by formulas of the *probability calculus* ( $\mathcal{PC}$ ), and the models range over inferential schemata whose parameter-language is  $\mathcal{PC}$ ), etc.

So, as long as we are dealing with boolean-reasoners only, or with reasoners whose epistemic situation is properly represented by means of a simple model, the fact that the

---

judgments of this particular *ps-semantics* is not in full agreement with the judgments of English-speakers across a wide variety of situations (including situations that are not properly represented by simple models) is not seen as an indicator of the inadequacy of that semantics. That is because simple models are meant to represent just one specific type of situation — not all of them.

What we fleshed out in this chapter is just a first attempt to offer a semantics for attributions of epistemic rationality. We just did this for  $\mathcal{E}_{\mathcal{P}\mathcal{L}}$ . That is, we presented just one type of *ps-semantics* (what we called here ‘simple models’ constitute just one family of *ps-semantics*). The adequacy of *possible-states semantics* in general will be open to a more comprehensive scrutiny as soon as we have models for formulas in other versions of  $\mathcal{E}_{\mathcal{L}}$ , where  $\mathcal{L}$  can be substituted by  $\mathcal{FOL}$ ,  $\mathcal{PC}$  or what have you. Maybe by getting richer languages and embedding richer models into our models we will get more accurate and more comprehensive results. So, it is an open question if there is a satisfactory version of a general *ps-semantics* of the type developed here. In future research, we aim to develop this type of semantics in more detail and to ground axiomatic systems that we will call ‘*Rationality Logics*’ by means of it.

# Conclusion

In the first part of this work we argued that *ex ante* rationality is not just a function of reasons to believe (or disbelieve, or doubt). In support of that judgment we presented (i) cases of ‘unreachable’ beliefs (beliefs whose content gains support from the content of the reasons available to a certain subject *S*, but such that *S* is not able to competently form them), (ii) conceptual considerations about attributions of rationality or justification (that being *ex ante* justified in believing something requires *being in a position* to competently form the relevant belief). We suggested, then, that rationality is not only a function of available reasons, but also of a certain kind of procedural knowledge: *knowledge of how to reason*.

We explicated the notion of *knowledge of how to reason* (or *knowledge of how to perform an inference*) using the notion of *availability of an inferential schema*. There is much more work to be done here: we need to explain in more detail what is the nature of inferential schemata, and to check in a more precise way how exactly our explication fits with current views about knowledge-how in general. While the reader may find the present work wanting in these respects, so we take it, the general point (that it is rational for one to form a certain belief only if one has the ability to do so) is not hampered by the lack of answers to those challenges.

Using the notion of *knowledge of how to reason*, we explicated the concept of *ex ante* rationality, and we showed how our proposal differs from similar theories already advanced in the literature (by Turri and Goldman). We offered a counterexample to these theories. Further, we argued that the ‘adding beliefs’ strategy for dealing with the problem of unreachable beliefs does not work: having more reasons does not necessarily put one in a position to perform an epistemically approvable inference.

In the second part we began to develop a more formal work. We developed both, a formal language to represent attributions of epistemic rationality and a version of a certain type of semantics for that language. Our formal developments here constitute the beginning of a bigger project — that of building an adequate model-theoretic semantics

---

for attributions of rationality in general, and using such a semantics to ground what we will call '*Rationality Logics*'. If the project will prove being theoretically fruitful is an open question.

So, *Part 1* of our dissertation is conclusive: we have a theory of *ex ante* rationality that, as far as our arguments go, is better than rival accounts. *Part 2* is not conclusive, however. Its conclusion lies in the future.



## References

- ACHINSTEIN, Peter. **The Book of Evidence**. New York: Oxford University Press, 2001. 304 p.
- ADAMS, Ernest. Four probability-preserving properties of inferences. **Journal of Philosophical Logic**, v. 25, n. 1, p. 1–24, 1996.
- ALSTON, William. **Beyond ‘Justification’**. New York: Cornell University Press, 2005. 274 p.
- AUDI, Robert. **The Structure Of Justification**. Cambridge: Cambridge University Press, 1993. 496 p.
- BEALL, Jc; RESTALL, Greg. Logical Consequence. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Winter 2013 Edition). URL = <<http://plato.stanford.edu/archives/win2013/entries/logical-consequence/>>, 2005.
- \_\_\_\_\_. **Logical Pluralism**. Oxford: Clarendon Press, 2006. 152 p.
- BLACKBURN, Patrick; RIJKE Maarten de; VENEMA Yde. **Modal Logic** (Cambridge Tracts in Theoretical Computer Science, 53). Cambridge: Cambridge University Press, 2001. 578 p.
- BOGHOSSIAN, Paul. Epistemic Rules. **Journal of Philosophy**, v. 105, n. 9, p. 472–500, 2008.
- \_\_\_\_\_. What is inference? **Philosophical Studies** (online-first articles), p. 1–18, 2012.
- BROWN, Bryson. On Paraconsistency. In: JACQUETTE, Dale (ed.), **A Companion to Philosophical Logic**. Cornwall: Blackwell Publishers, p. 628–650, 2002.
- CARNAP, Rudolf. **Logical Foundations of Probability** (2 ed). Chicago: University of Chicago Press, 1962. 613 p.
- CARROL, Lewis. What the Tortoise Said to Achilles. **Mind**, v. 4, n. 14, p. 278–280, 1895.
- CHRISTENSEN, David. Disagreement as evidence: The epistemology of controversy. **Philosophy Compass**, v. 4, n. 5, p. 756–767, 2009.
- CHRISTENSEN, David; LACKEY, Jennifer (eds.). **The Epistemology of Disagreement: New Essays**. Oxford: Oxford University Press, 2013. 288 p.
- COHEN, Jonathan. Can Human Irrationality Be Experimentally Demonstrated? **Behavioral and Brain Sciences**, v. 4, n. 3, p. 317–370, 1981.
- CONEE, Earl; FELDMAN, Richard. Evidentialism. **Philosophical Studies**, v. 48, p. 15–34, 1985. Reprinted with an Afterword in Conee and Feldman (2004), p. 83–107.
- \_\_\_\_\_. Internalism Defended. In: KORNBLITH, Hilary (ed.), **Epistemology –**

- Internalism and Externalism.** Cornwall: Blackwell Publishers, p. 231–260, 2001.
- \_\_\_\_\_. **Evidentialism – Essays in Epistemology.** New York: Oxford University Press, 2004. 336 p.
- \_\_\_\_\_. Evidence. In: SMITH, Quentin (Ed.), **Epistemology: New Essays.** New York: Oxford University Press, p. 83–104, 2008.
- CRUPI, Vincenzo. Confirmation. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Spring 2014 Edition). URL = <<http://plato.stanford.edu/archives/spr2014/entries/confirmation/>>, 2013.
- DAVIDSON, Donald. A Coherence Theory of Truth and Knowledge. In: DAVIDSON, Donald (ed.), **Subjective, Intersubjective, Objective.** New York: Oxford University Press, 2001. 256 p.
- DOUGHERTY, Trent (ed.). **Evidentialism and Its Discontents.** New York: Oxford University Press, 2011. 352 p.
- DUC, Ho Ngoc. Reasoning about Rational, but not Logically Omniscient Agents. **Journal of Logic and Computation**, v. 7, n. 5, p. 633–648, 1997.
- DUNN, Michael. Relevance Logic and Entailment. In: GABBAY, Dov; GUENTHNER, Franz (eds.), **Handbook of Philosophical Logic**, v. III. Dordrecht: Reidel, p. 117–224, 1986.
- EARMAN, John. **Bayes or Bust?** Cambridge: MIT Press, 1992. 288 p.
- FANTL, Jeremy. Knowledge How. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Winter 2012 Edition). URL = <<http://plato.stanford.edu/archives/win2012/entries/knowledge-how/>>, 2012.
- FAGIN, Ronald; HALPERN, Joseph; MOSES, Yoram; VARDI, Moshe. **Reasoning about Knowledge.** Cambridge: MIT Press, 1995. 491 p.
- FODOR, Jerry. **LOT 2: The Language of Thought Revisited.** Oxford: Oxford University Press, 2008. 240 p.
- FOLEY, Richard. Justified Inconsistent Beliefs. **American Philosophical Quarterly**, v. 16, p. 4, p. 247–257, 1979.
- \_\_\_\_\_. **Working without a Net: A Study of Egocentric Epistemology.** Oxford: Oxford University Press, 1992. 224 p.
- GOLDMAN, Alvin. What is Justified Belief? In: PAPPAS, George (ed.), **Justification and Knowledge.** Dordrecht: D. Reidel, 1979. Reprinted in Goldman (2012), p. 29–49.
- \_\_\_\_\_. **Epistemology and Cognition.** Cambridge: Harvard University Press, 1986. 446 p.
- \_\_\_\_\_. Internalism, Externalism, and the Architecture of Justification. **The Journal of Philosophy**, v. 106, n. 6, p. 309–338, 2009. Reprinted in Goldman (2012), p. 123–150.

- \_\_\_\_\_. Toward a Synthesis of Reliabilism and Evidentialism. In: DOUGHERTY, Trent (ed.) **Evidentialism and Its Discontents**. New York: Oxford University Press, 2011. Reprinted in Goldman (2012), p. 123–150.
- \_\_\_\_\_. **Reliabilism and Contemporary Epistemology**. New York: Oxford University Press, 2012. 336 p.
- GRECO, John. Virtues and vices of virtue epistemology. **Canadian Journal of Philosophy**, v. 23, n. 3, p. 413–432, 1993.
- \_\_\_\_\_. The nature of ability and the purpose of knowledge. **Philosophical Issues**, v. 17, n. 1, p. 57–69, 2007.
- HAILPERIN, Theodore. **Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications**. Bethlehem: Lehigh University Press, 1996. 304 p.
- HALES, Steven. Epistemic Closure Principles. **Southern Journal of Philosophy**, v. 33, n. 2, p. 185–202, 1995.
- HANNA, Robert. **Rationality and Logic**. Cambridge: MIT Press – A Bradford Book, 2006. 344 p.
- HARMAN, Gilbert. **Change in View: Principles of Reasoning**. Cambridge: MIT Press – A Bradford Book, 1986. 162 p.
- JOHNSON–LAIRD, Philip. Toward a Unified Theory of Reasoning. **Psychology of Learning and Motivation**, v. 59, p. 1–42, 2013.
- KELLY, Thomas. ‘Evidence’, In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Fall 2008 Edition). URL = <<http://plato.stanford.edu/archives/fall2008/entries/evidence/>>, 2006.
- KLEENE, Stephen Cole. **Mathematical logic**. New York: Dover Publications, 1967. 432 p.
- KLEIN, Peter. **Certainty: A Refutation of Skepticism**. Minneapolis: University of Minnesota Press, 1981. 264 p.
- \_\_\_\_\_. The Virtues of Inconsistency. **The Monist**, v. 68, n. 1, p. 105–135, 1985.
- \_\_\_\_\_. What Makes Knowledge the Most Highly Prized Type of Belief? In: BLACK, Tim; BECKER, Kelly (eds.), **The Sensitivity Principle in Epistemology**. Cambridge: Cambridge University Press, p. 152–172, 2012.
- KORCZ, Keith Allen. The Epistemic Basing Relation. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Spring 2010 Edition). URL = <<http://plato.stanford.edu/archives/spr2010/entries/basing-epistemic/>>, 2002.
- KVANVIG, Jonathan. Justification and Proper Basing. In: OLSSON, Erik (ed.), **The Epistemology of Keith Lehrer**. Dordrecht: Kluwer Publishing Company – Philosophical Studies Series (Book 95), p. 43–62, 2003.

- \_\_\_\_\_. Truth is Not the Primary Epistemic Goal. In: STEUP, Matthias; TURRI, John; SOSA, Ernest (eds.), **Contemporary Debates in Epistemology**. 2 ed. Oxford: Blackwell Publishing, p. 352–362, 2013.
- KYBURG, Henry. **Probability and the Logic of Rational Belief**. Middletown: Wesleyan University Press, 1961. 346 p.
- LEHRER, Keith. How Reasons Give Us Knowledge, Or The Case Of The Gypsy Lawyer. **The Journal of Philosophy**, v. 68, n. 10, p. 311–313, 1971.
- LEWIS, David. Elusive Knowledge. **Australasian Journal of Philosophy**, v. 74, p. 549–567, 1996.
- MAHER, Patrick. Confirmation Theory. In: BORCHERT, Donald (ed.), **The Encyclopedia of Philosophy**. 2 ed. London: MacMillan Publishers, 2005.
- \_\_\_\_\_. Explication defended. **Studia Logica**, v. 86, n. 2, p. 331–341, 2007.
- MAIER, John. Abilities. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Fall 2011 Edition). URL = <<http://plato.stanford.edu/archives/fall2011/entries/abilities/>>, 2010.
- MAYES, Randolph. Theories of Explanation. In: **Internet Encyclopedia of Philosophy**. URL = <<http://www.iep.utm.edu/explanat/>>. Accessed in 03/02/2014.
- MENZEL, Christopher. Possible Worlds. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Spring 2014 Edition). URL = <<http://plato.stanford.edu/entries/possible-worlds/>>, 2013.
- NETA, Ram. What evidence do you have? **British Journal for the Philosophy of Science**, v. 59, n. 1, p. 89–119, 2008.
- NOË, Alva. Against Intellectualism. **Analysis**, v. 65, n. 288, p. 278–290, 2005.
- PACUIT, Eric. Dynamic Epistemic Logic I: Modeling Knowledge and Belief. **Philosophy Compass**, v. 8, n. 9, p. 798–814, 2013.
- POLLOCK, John. **Cognitive Carpentry: A Blueprint for How to Build a Person**. Cambridge: MIT Press, 1995. 291 p.
- POLLOCK, John; CRUZ, Joseph. **Contemporary Theories of Knowledge**. 2 ed. Lanham: Rowman & Littlefield, 1999. 288 p.
- POSTON, Ted. Internalism and Externalism in Epistemology. In: **Internet Encyclopedia of Philosophy**. URL = <[www.iep.utm.edu/int-ext/](http://www.iep.utm.edu/int-ext/)>. Accessed in 03/02/2014.
- PRIEST, Graham; SMILEY, Timothy. Can Contradictions be True? **Proceedings of the Aristotelian Society**, v. 68(Supplement), p. 17–54, 1993.
- PRIEST, Graham; TANAKA, Koji; WEBER, Zach. Paraconsistent Logic. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Fall 2013 Edition). URL = <<http://plato.stanford.edu/archives/fall2013/entries/logic-paraconsistent/>>, 1996.
- PRIEST, Graham. Logicians Setting Together Contradictories: A Perspective on Rele-

- vance, Paraconsistency and Dialetheism. In: JACQUETTE, Dale (ed.), **A Companion to Philosophical Logic**. Cornwall: Blackwell Publishers, p. 651-664, 2002.
- PYLYSHYN, Zeno. Rules and representations: Chomsky and representational realism. In: KASHER, Asa (ed.), **The Chomskyan Turn**. Oxford: Blackwell Pub, 1991. 416 p.
- RIPS, Lance. **The Psychology of Proof: Deductive Reasoning in Human Thinking**. Cambridge: MIT Press, 1994. 463 p.
- ROSA, Luis. Justification and Cognitive Algorithms. **Philosophia** (*online-first articles*). URL = <<http://link.springer.com/article/10.1007/s11406-013-9506-6>>, 2013.
- RYLE, Gilbert. Knowing How and Knowing That: The Presidential Address. **Proceedings of the Aristotelian Society**, v. 46, p. 1–16, 1945.
- SHAPIRO, Stewart. Necessity, Meaning, and Rationality: The Notion of Logical Consequence. In: JACQUETTE, Dale (ed.), **A Companion to Philosophical Logic**. Cornwall: Blackwell Publishers, p. 227–240, 2002.
- SMITH, Edward; LANGSTON, Christopher; NISBETT, Richard. The Case for Rules in Reasoning. **Cognitive Science**, v. 16, p. 1–40, 1992.
- SORENSEN, Roy. Epistemic Paradoxes. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Spring 2014 Edition). URL = <<http://plato.stanford.edu/archives/spr2014/entries/epistemic-paradoxes/>>, 2006.
- SOSA, Ernest. **A Virtue Epistemology: Apt belief and reflective knowledge**. Volume I. New York: Oxford University Press, 2007. 128 p.
- \_\_\_\_\_. How Competence Matters in Epistemology. **Philosophical Perspectives**, v. 24, n. 1, p. 465–475, 2010.
- \_\_\_\_\_. **Knowing full well**. Princeton: Princeton University Press, 2011. 176 p.
- STALNAKER, Robert. Intellectualism and the objects of knowledge. **Philosophy and Phenomenological Research**, v. 85, n. 3, p. 754–760, 2012.
- STANLEY, Jason; WILLIAMSON, Timothy. Knowing How. **Journal of Philosophy**, v. 98, n. 8, p. 411–444, 2001.
- STANLEY, Jason. **Know How**. New York: Oxford University Press, 2011. 224 p.
- TALBOTT, William. Bayesian Epistemology. In: ZALTA, Edward (ed.), **The Stanford Encyclopedia of Philosophy** (Fall 2013 Edition). URL = <<http://plato.stanford.edu/archives/fall2013/entries/epistemology-bayesian/>>, 2001.
- THAGARD, Paul. **Mind: Introduction to Cognitive Science**. 2 ed. A Bradford Book, 2005. 280 p.
- TURRI, John. The ontology of epistemic reasons. **Nous**, v. 43, n. 3, p. 490–512, 2009.
- TOLLIVER, Josph. Basing Beliefs on Reasons. **Grazer Philosophische Studien**, v. 15, p. 149–161, 1982.
- TURRI, John. The ontology of epistemic reasons. **Nous**, v. 43, n. 3, p. 490–512, 2009.

- \_\_\_\_\_. On the relationship between propositional and doxastic justification. **Philosophy and Phenomenological Research**, v. 80, n. 2, p. 312–326, 2010.
- TVERSKY, Amos; KAHNEMAN, Daniel. Availability: A heuristic for judging frequency and probability. **Cognitive Psychology**, v. 5, n. 2, p. 207–232, 1973.
- VAN BENTHEM, Johan. **A Manual of Intensional Logic**. Stanford: CSLI Publications, 1988. 74 p.
- WEDGWOOD, Ralph. Rational ‘ought’ implies ‘can’. **Philosophical Issues**, v. 23, n. 1, p. 70–92, 2013.
- WIGGINS, David. Practical Knowledge: Knowing How To and Knowing That. **Mind**, v. 121, n. 481, p. 97–130, 2012.
- WILLIAMSON, Timothy. **Knowledge and its Limits**. New York: Oxford University Press, 2002. 352 p.
- ZAGZEBSKI, Linda. The inescapability of Gettier problems. **Philosophical Quarterly**, v. 44, n. 174, p. 65–73, 1994.