A Defense of Temperate Epistemic Transparency

Eleonora Cresto (eleonora.cresto@gmail.com)
CONICET, Argentina

Abstract. Epistemic transparency tells us that, if an agent S knows a given proposition p, then S knows that she knows that p. This idea is usually encoded in the so-called *KK principle* of epistemic logic. The paper develops an argument in favor of a moderate version of *KK*, which I dub *quasi-transparency*, as a normative rather than a descriptive principle. In the first part I put forward the suggestion that epistemic transparency is not a demand of ideal rationality, but of ideal epistemic responsibility, and hence that ideally responsible agents verify transparency principles of some sort; I also contend that their satisfaction should not be tied to an internalist epistemology. The central argument in favor of transparency is then addressed in the second part of the paper, through the development of a formal system. I show that, in a well-behaved formal setting, a moderate version of transparency is imposed upon us as a result of a number of independent decisions on the structure of higher-order probabilities, as long as we request that our probability and knowledge attributions cohere with each other. Thus I give a rationale to build a model for a hierarchy of languages with different levels of knowledge and probability operators; we obtain an analogous to *KK* for successive knowledge operators without actually demanding transitivity. The formal argument reinforces the philosophical intuition that epistemic transparency is an important desideratum we should not be too ready to dismiss.

Keywords: transparency, responsibility, self-knowledge, higher-order probabilities, epistemic logic

1. Introduction

Epistemic logic is typically concerned with the formalization of crucial aspects of the concept of knowledge. Once basic features of our pre-theoretical understanding of knowledge are captured by suitable axioms, we can go on to explore the formal properties of the system, which sometimes lead us to discover important consequences that might not be so evident at first sight. As with many theories (both formal and informal), we look for a reflective equilibrium between our prior intuitions and the claims that turn out to be valid in the theoretical framework. And, as with many other modal logics, we can discuss which particular system best expresses the features we have in mind.

In particular, we can wonder whether an adequate account of knowledge is such that, from the fact that an agent knows a given proposition p, we can infer that she also knows that she knows that p, *i.e.*, we can wonder whether we have the right to infer that she has *second-*

order knowledge. This is precisely the idea encoded in the so-called ***KK principle***, or $Kp \to KKp$. It is also referred to as a *principle of positive introspection*, *principle of knowledge-reflexivity*, *epistemic transparency*, *self-knowledge* or *luminosity*; in what follows I will refer to ***KK*** by any of these names indistinctly. In this paper I will attempt a defense of a temperate version of ***KK*** as a normative, rather than descriptive, epistemic principle.

Epistemic transparency can well be taken to be implicit in traditional theories of knowledge, from Plato to the first half of the Twentieth Century. Once epistemic logics appeared in the epistemological scene (starting with Hintikka's seminal book *Knowledge and Belief*, in 1962), the ***KK*** principle became explicitly stated in the formal literature. Then epistemic externalism broke in, and things began to change. Externalism tells us, precisely, that justification is not "internal" to our consciousness, but is a matter of placing ourselves in the right relationship with the environment, regardless of whether we are aware of this fact. Here's a delicate point, however, because to say that knowledge does not require any particular state of awareness regarding justification does not necessarily indicate that agents will typically know without knowing that they know, insofar as second-order knowledge could *also* be defined in an externalistic fashion. Thus, assuming an agent formed the corresponding second-order belief, such a belief could be justified by reasons to which, once again, she has no access.

Of course, externalism also allows for the idea that knowledge does not demand the corresponding second-order belief to begin with; clearly, in this case positive introspection fails. In any case, there are well-discussed, explicit rejections of transparency that proceed through more sophisticated paths. Many such arguments involve an attempt to escape from first-order skepticism: by rejecting ***KK*** we prevent second-order skeptical claims to infest the first level through *modus tollens*.

In recent years we have witnessed the development of new arguments against ***KK***. Timothy Williamson, in particular, articulates two fundamental lines of attack. On the one hand he suggests, mostly through carefully chosen examples, that no mental state is luminous, not even the most obvious ones – hence surely knowledge is not luminous either.[1] On the other hand, Williamson contends that knowledge requires a margin of error: if I am entitled to say that I know that the tree in front of me is not 0.5 meters, then the tree should not be 0.51 meters either. This requirement can be seen as a particular elaboration, for perceptual knowledge, of the so-called *safety* condition. The safety condition tells us that what the agent believes should be true not only in the actual world, but also in close worlds the agent cannot discriminate from the actual one. Williamson then shows, by means of a soritic argument, that if we accept ***KK***, the margin of error principle leads us to an inconsistency (cf. Williamson [30], chapter 5).

I believe Williamson's considerations on margin of error principles reveal us something important about the structure of first-order knowledge. However, there is still room to propose a limited defense of ***KK***.[2] I will argue in favor of a moderate version of ***KK*** indirectly, by showing that it is forced upon us as a result of a number of independent decisions on the structure of second-order probabilities and the way probability and knowledge attributions cohere with each other; it should be noted that my argument will not be tied to an internalist project about justification.

The article is organized as follows. It comprises two main parts, quite distinct from one another. In the first part (section 2) I consider briefly why we should focus on epistemic transparency in the first place. Notice that my answer to the question of why we care about epistemic transparency is not the central argument of the paper: it is just meant to motivate the search for the results this work seeks to obtain. The second part (sections 3 to 8) concerns the (indirect) argument for epistemic transparency properly speaking. I take as a starting point Williamson's formal proposal in "Very Improbable Knowing" (section 3); then I suggest an alternative setting that can be deemed philosophically satisfactory (sections 4 and 5); finally, I devote some space to discuss the philosophical consequences of the model (sections 6 and 7), as well as its connection with recent work on the topic (section 8).

First Part: Theoretical Motivations

2. Why Should We Care? Rationality, Responsibility and Reflection

Why is positive introspection important? One can be tempted to contend here that epistemic transparency is simply a feature of idealized reasoners, such that an epistemic system should validate transparency if and only if the system explicitly deals with ideally rational agents. According to this line of thought, attempting to defend a principle of epistemic transparency is either idle or impossible (because real agents do not satisfy it). This strikes me as a false dichotomy. I would like to suggest that introspective principles do much more (and, in a different sense, much less) than depict an ideal reasoner.

The assumption of ideal rationality deserves a careful discussion, of which here I can offer no more than an outline. We can put into question, for instance, whether epistemic models should always presuppose logical omniscience, where "logical omniscience" usually refers to consistency and deductive closure; sometimes the term is also meant to include probabilistic coherence, and perhaps even expected utility maximization. At one end of the spectrum we find philosophers who defend at least part of such requirements without reservations; we can say, for instance, that logical omniscience embodies a regulative ideal that we clearly wish to fulfill. Thus, as soon as we notice that a particular proposition follows deductively from some of our beliefs, we typically feel compelled to believe it as well – or to revise our background. Moreover, as deduction transmits truth, recognizing that we know the premises usually forces us to recognize that we also know the conclusion. This account seems particularly sensible if we understand beliefs as *commitments*, as some authors have proposed. At the opposite end of the spectrum we find defenders of so-called bounded rationality. Most supporters of bounded rationality rely on some version of the "ought-implies-can" maxim, and hence argue that we should not ask for what we are, as a matter of fact, unable to attain.

In any case, we could well distinguish the ideal of epistemic rationality that gets captured by logical omniscience from the type of idealization that springs from the acceptance of (some version of) epistemic transparency. Even if logical omniscience were deemed to be an appropriate demand (say, for reasons that ultimately go back to our pre-theoretical understanding of rationality), it does not follow from here that transparency demands will be found just as appropriate. Consider, for instance, the many formal proposals that rely on a Kripkean-based semantics – which forces agents to be logically omniscient – while at the same time rejecting the validity of ***KK***. I take it that such proposals are implicitly committed to the modeling of ideally rational agents; however, in such frames, knowing that one knows remains conceptually closer to empirical rather than to *a priori* knowledge, and hence it can be subjected to empirical doubts. Incidentally, recall that in general we do not ask agents to satisfy empirical omniscience to be rational.[3]

I would like to seek a middle ground between those who take transparency to be part of our ideal of a rational epistemic agent, and those who refuse to do so. The position favored in this paper is then intermediate: on the one hand, I agree that transparency is not an ideal of rationality in the same sense that consistency or deductive closure are; however, this is not to say that it is not an important ideal on its own right, albeit of a different type: it is an ideal of *epistemic responsibility*.[4]

How does the connection between transparency and responsibility go, exactly? Some authors have contended that the mere fact that we have certain beliefs already entitles us to talk about epistemic responsibility, in some deflated sense.[5] But, as I see it, responsibility is hardly an all-or-nothing affair. It might be correct to say that, in a minimal sense, all we require from agents in order to credit them with responsibility is to have certain doxastic attitudes. However, there are uses of the concept of epistemic responsibility that are not so minimal. In particular, there is a clearly identifiable sense of the term according to which we are reluctant to say that an agent is fully epistemically responsible for her belief that p unless she is very much aware of her having p, and perseveres on her belief that p on reflection. Part of the explanation for our reluctance is that agents who fully embrace their first-order mental states are perceived as more in control of themselves, and as having a more sophisticated epistemic life. This richer sense of responsibility is clearly a desideratum. To put it in a slogan: "make sure you own your own beliefs [and desires]". Transparency, under some suitable formulation to be discussed in the sections to follow, is then a requirement of fully idealized responsible agents: if S is a fully idealized responsible agent and S knows that p, then S has duly reflected on her (first-order) epistemic states and has found it to be the case that she knows that p.

Many philosophers exploited the link between agenthood and transparency in the past. A possible way to go would be to contend that having knowledge of our own intentional states is constitutive of the very idea of intentional state.[6] I don't think we need to commit ourselves to anything this strong, though. We can accept that full-fledged responsibility requires being in the right sort of reflective state, without pronouncing ourselves as to whether this fact is actually constitutive of intentional attitudes.[7] In any case, we have to be careful concerning what sense of reflective state is at stake. We might wonder, for instance, whether the reflective stance that I identified as necessary for full-fledged epistemic responsibility is not just a side effect of the demand for epistemic justification, particularly as understood by internalist epistemology. I take it that the answer is "no". According to the sense of reflection that gets vindicated by internalist epistemology, epistemic responsibility is a trait which, when possessed by the agent, prevents her from believing without an explicit assessment of the available evidence. According to the sense I am interested in, by contrast, epistemic responsibility is a trait that forces an agent to fully embrace her first-order beliefs: the agent can then be said to *ratify* them. Sometimes the two phenomena go together, but we can also have the second without the first.

To be successful, this picture has yet to tell us how the reflective stance involved in ratification (in the sense just explained) can give us second-order *knowledge*, in addition to second-order belief on our first-order attitudes. I will not elaborate on this idea here, but let me just give a hint as to how the argument would go. A crucial step is to recognize that justification is not always perceived as necessary for knowledge (and *self*-knowledge) attribution. If we pay due attention to the linguistic evidence we will notice that, in many circumstances, agents do not care about justification at all, and yet they do not abandon the concept of knowledge: in many circumstances, knowledge attribution *is* just attribution of true belief. A possible explanation for this fact is that justification becomes relevant when – but only when – we enter a very particular reflective stance, to wit, when we examine the relevant beliefs *under the light of a possible epistemic revision*.[8] Hence there is room to contend that the ability to have or lack justification does not preexist: the conceptual space for justification is *created* by placing ourselves in what we might call "a deliberation mood". Thus, reflecting on our beliefs, or on the beliefs of a third party agent, need not involve a deliberation mood, and hence not every reflective stance is a justification stance. In light of this, it is not generally true that full-fledged epistemic responsibility demands justification (although sometimes it certainly does). Sometimes the quest for justification just doesn't arise, but the agent could still be said to have knowledge, and second-order knowledge, of the relevant propositions – and hence still qualify as fully responsible, in the sense discussed here.[9]

I have contended that epistemically responsible beings satisfy transparency, under some suitably formulation. We might want to consider briefly, in addition, whether ideally responsible agents can also be taken *to know what they ignore*, *i.e.*, whether they can be taken to verify a principle of *wisdom*, or *negative introspection* ($\sim Kp \rightarrow K \sim Kp$). Nothing I have said so far forces us to conclude that negative introspection is indeed a reasonable demand. I have argued that epistemically responsible agents are expected to be aware of their own commitments, and fully embrace them. But they need not be equally expected to ratify their *lack* of commitments.[10] The explanation is simple: positive beliefs and knowledge states, when correctly identified as *the beliefs we stand by*, give us a sense of identity and agenthood, a sense of who we in fact *are*, in a way suspensions of judgment do not. To put it somewhat romantically, positive beliefs and knowledge states help us define the person we are now, whereas suspensions of judgment merely gesture towards the person we might become. Hence there is a clear asymmetry between demanding transparency and demanding wisdom. Of course, gaining awareness of our lack of knowledge can indeed be a desideratum in its own right, even if its desirability is not an obvious byproduct of ideal epistemic responsibility.

In what follows I will seek to show that we have formal reasons to argue in favor of some version of positive introspection, and hence in favor of a model that captures the concept of an ideally responsible agent. This result, I take it, will reinforce the claim that epistemic transparency is a desideratum we should not be too ready to dismiss.

Second Part: A Formal Argument in Favor of Epistemic Transparency

3. Williamson on Improbable Knowing

Let me start by recalling some of the axioms that are often discussed when we formulate an epistemic logic. It is usually accepted, for instance, that all tautologies from propositional logic should be valid in our system. Consider next axiom ***K***:

$$K(\phi \to \psi) \to (K\phi \to K\psi) \tag{K}$$

K amounts to saying that, if an agent knows both a material implication and its antecedent, then she also knows the consequent. **K** is valid in any normal system (in Kripke's sense); those who argue against deductive closure for knowledge, such as Robert Nozick [17], are bound to reject it. Consider also:

$$K\phi \to \phi \tag{T}$$

T says that if someone knows that ϕ, then ϕ is the case: we cannot know false things; ***T*** seems reasonable if we think, as most people do, that knowledge is factive, *i.e.*, that it involves truth. Finally, we may also wonder about the validity of a number of introspective principles, such as ***KK*** (the principle of positive introspection), as well as the principle of wisdom, or principle of negative introspection (NI):

$$K\phi \to KK\phi \tag{KK}$$

$$\sim K\phi \to K\sim K\phi \tag{NI}$$

Most authors have found negative introspection to be even more contentious than $\boldsymbol{KK}$; even though it is not the central topic of this paper, I will have a few more things to say about it in later sections.

On the other hand, in order to provide a semantic machinery for a modal epistemic system it is customary to rely on a set-theoretical structure that includes, at the very least, a set W of possible worlds and an accessibility relation R among worlds. Intuitively, if w_1 and w_2 are two worlds linked by R, then the agent cannot discriminate among them (for all he knows, if he is w_1 he might well be in w_2). The validity of certain axioms rather than others depends on the structure of the accessibility relation. Thus, for example, if R is reflexive (*i.e.*, if each world can be related to itself) we guarantee that $\boldsymbol{T}$ holds, whereas the transitivity of R is necessary and sufficient for $\boldsymbol{KK}$.

In "Very Improbable Knowing",[11] Williamson considers a frame $\langle W, R, P_{prior} \rangle$ for a single agent, where W is a set of worlds, $R \subseteq W \times W$ is an accessibility relation between worlds, and P_{prior} is a prior probability distribution defined over subsets of W. It is assumed that W is finite and that P_{prior} is uniform, in order not to add useless complications.[12] As usual, propositions are subsets of W, and, if ϕ is a proposition, then $K\phi$ is the set of all worlds connected with ϕ-worlds through R:

$$K\phi = \{w \in W : \forall x \in W(wRx \to x \in \phi)\}$$

Define also $R(w)$, for any $w \in W$, as the strongest proposition known in w:

$$R(w) = \{x \in W : wRx\}$$

It is easy to see that, by definition of $K\phi$, $R(w)$ is included in every proposition known by the agent.

Given that, by hypothesis, P_{prior} is a uniform probability measure and W is finite, for any proposition ϕ, $P_{prior}(\phi)$ will just amount to $\#[\phi]/\#W$. Consider now the definition of ϕ's probability in a given world w, or ϕ's *evidential probability* in w, which shall be written as $P_w(\phi)$. $P_w(\phi)$ is obtained by conditionalizing on what the agent knows in w, *i.e.*, on $R(w)$. Hence,

$$P_w(\phi) = P_{prior}(\phi \mid R(w)) = P_{prior}(\phi \cap R(w)) \ / \ P_{prior}(R(w))$$

This definition is in agreement with the much discussed E=K thesis, according to which the agent's evidence at a particular moment is no less than the totality of his or her knowledge. As we have a uniform prior distribution, in order to calculate ϕ's evidential probability in w we just consider how many of the $R(w)$-worlds are also ϕ-worlds. A natural consequence of this idea is that, for all w, $P_w(R(w)) = 1$.

We can also wonder about the extension of a proposition stating that ϕ's probability is r (for $r \in [0, 1]$); Williamson defines it as the set of worlds in which ϕ has evidential probability r:

$$[P(\phi) = r] =_{\text{df}} \{w \in W : P_w(\phi) = r\}$$

Within this setting, Williamson argues that the ***KK*** principle can be formulated in probabilistic terms: "The ***KK*** principle is equivalent to the principle that if the evidential probability of p is 1, then the evidential probability that the evidential probability of p is 1 is itself 1" (Williamson [28, p. 8]). It is easy to show that this claim is false if R is not transitive; in fact, when R is not transitive we can propose examples in which $P_w([P(R(w)) = 1])$ is as low as we want.[13]

4. Second Thoughts about Second-Order Probabilities

According to the intended interpretation of Williamson's proposal, whenever we assess the probability of proposition $[P(\phi) = r]$, for any ϕ, we are actually assessing a *higher-order* probability. However, I believe this assertion is problematic. The expression between square brackets is just a label to refer to a certain set of worlds; thus, it would not make any difference if our label were, say, "ψ", without any reference to P whatsoever – as long as the worlds remains the same. So there is a sense in which we might just as well be calculating a *first*-order probability. The root of the problem, I take it, is that propositions understood as sets of possible worlds are too coarse-grained to allow for what we want: second-order probabilities call for a more fine-grained representation device. Even if "χ" and "ψ" were logically equivalent, in the sense that they capture the same element in 2^W, intuitively, their probability might differ; as we shall see, the reason is ultimately that second-order probabilities demand that we conditionalize over second-order evidence. In light of this, in what follows I will propose a representation strategy that enables us to take into account not only propositions understood as sets of worlds, but also their *mode of presentation*, so to speak.[14]

Let me then suggest a model in which genuine second-order probabilities apply to well-regimented probabilistic statements, rather than to sets of worlds. Thus the probability of a set of worlds (or proposition) will depend crucially on the way we refer to it – in particular, on whether we refer to it through a probabilistic discourse or not.[15] We will then take the arguments of probability functions of our system to be sentences of a sequence of duly regimented languages. As usual, if $\underline{\phi^i}$ is a formula of L^i, $[\phi^i]$ will be the proposition, or set of worlds, in which $\underline{\phi^i}$ is true. In what follows, sentences and other linguistic items will always appear as underlined expressions.

To carry out this project we need to enrich the original frame with a function v that helps us assess the truth-values of sentences of a sequence of languages $L^0, L^1, \ldots L^n, \ldots$, with probability operators $\underline{\mathrm{P}^0}$, $\underline{\mathrm{P}^1}, \ldots$ $\underline{\mathrm{P}^n}, \ldots$ of increasingly higher levels.[16] We will consider also a sequence of functions $P^1_w, \ldots P^n_w, \ldots$ (for each $w \in W$), which will be applied to increasingly complex arguments. To put it somewhat sloppily, in each case P^i_w will take as arguments sentences of language $i-1$:

$$P^i_w : L^{i-1} \to \mathbb{R}$$

(A more careful presentation, as well as further details on language formations rules, will be given in section 5.) Expressions of the form $P_{prior}(\underline{\phi})$ or $P^i_w(\underline{\phi})$ will not be part of any language of the sequence $L^0, L^1, \ldots L^n, \ldots$, but they will belong to the metatheory. In this way we make sure we are not mixing up truths of the system with truths *in* the system.

Following Williamson's proposal, prior probability will amount to the cardinality of the set of worlds in which a sentence of some language is true, given function v, divided by the cardinality of W. On the other hand, it is natural to demand that, for all w, $\underline{\mathrm{P}^i(\phi) = r}$ be true in w iff $P^i_w(\underline{\phi}) = r$, where "$\underline{\mathrm{P}^i(\phi) = r}$" is a sentence of L^i, and "$\underline{\phi}$" is a sentence of L^{i-1}. The central problem now is how to define evidential probability i in a world – in other words, how to conditionalize.

Suppose we have information about the state of the weather tomorrow. We have read the forecast in the newspaper, watched the weather channel, etc. On the basis of all this, we conclude that the probability of rain tomorrow is r. Now suppose a friend asks us how probable it is that our rational degree of belief that there's rain tomorrow is in fact r. As I see it, in this case our friend is no longer interested in the probability of a proposition about meteorology, but in the probability of a proposition *about the degree of confirmation* possessed by our original meteorological statement. Which is the relevant evidence to answer this question, then? Intuitively, what we have to assess is how good we are at the time of engaging in confirmation theory. Thus the relevant total evidence is no longer $R(w)$, but a *second-order* evidence: the evidence for our second-order probability should consist in what we know about our capabilities to adequately confirm propositions. And the strongest proposition that expresses this idea is indeed $KR(w)$. Hence when we calculate a second-order evidential probability we should conditionalize on $KR(w)$. The proposal then generalizes to even higher-order levels. Notice, incidentally, that the idea of conditionalizing on higher-order evidence corpora when dealing with higher-order knowledge can be taken to be in perfect agreement with the K=E thesis, well understood.

How should we conditionalize, then? A first suggestion could be to apply the following rule:

For $i \geq 1$ and any $w \in W$:

$$\begin{aligned} &P^i_w(\underline{\mathrm{P}^{i-1}(\ldots \mathrm{P}^1(\phi) = r \ldots) = s}) = \\ &\qquad P^i_w(\underline{\mathrm{P}^{i-1}(\ldots \mathrm{P}^1(\phi) = r \ldots) = s}) \mid (\underline{K^{i-1} \ldots KR(w)}) = \\ &P_{prior}(\underline{\mathrm{P}^{i-1}(\ldots \mathrm{P}^1(\phi){=}r \ldots){=}s \;\; \& \;\; K^{i-1} \ldots KR(w)}) \;/\; P_{prior}(\underline{K^{i-1} \ldots KR(w)}) \end{aligned}$$

Here the sequence of $\underline{K}$s is simply the result of iterating the *same* $\underline{K}$ operator as many times as $\underline{\mathrm{P}}$-operators are in the nominator's argument. The language level to which the argument belongs (as indicated by $\underline{\mathrm{P}}$'s super-index), determines the order of the probability function whose value we are seeking to calculate, and fixes the number of $\underline{K}$s we'll have to iterate to make the calculation. If $i = 1$, we just have, as before:

$$P^1_w(\underline{\phi}) = P_{prior}(\underline{\phi \;\&\; R(w)}) \;/\; P_{prior}(\underline{R(w)})$$

This first proposal is not completely satisfactory, though, because it is easy to show that it leads us to divorcing probability 1 from knowledge: there will be models in which $P^2_w(\underline{\mathrm{P}^1(R(w)) = 1}) = 1$ and yet $KKR(w)$ is not true in w.[17]

To overcome this difficulty, we can enrich languages $L^0, L^1, \ldots L^n, \ldots$ with a sequence of knowledge operators $\underline{K^0}, \underline{K^1}, \ldots \underline{K^n}, \ldots$ that runs parallel to our sequence of *probability* operators.[18] We will need, then, a family of relations $R^1, \ldots R^n, \ldots$ for $\underline{K^1}, \ldots \underline{K^n}, \ldots$ Thus, the desire that probability and knowledge claims cohere with each other motivates us to propose a model with multiple K-operators. I will discuss the legitimacy of this motivation with some detail in section 6. But before that, let me describe the formal proposal more carefully.

5. A Model for Temperate Transparency

Consider then the model $\mathscr{M} = \langle W, R^1, \ldots R^n \ldots, P_{prior}, v \rangle$, where:

1. W is a (finite) set of worlds.[19]

2. R^i is a reflexive relation over W, for all $i \geq 1$. Moreover, $R^i \subseteq R^{i-1} \ldots \subseteq R^1$.

3. There is a sequence of languages $L^0, L^1, \dots L^n, \dots$ such that:

   a) $\underline{p_1}, \dots \underline{p_n}$ are atomic formulas of L^0. Atomic formulas are well formed formulas (wff).

   b) If $\underline{\phi}$ is a wff of L^i, $\underline{\phi}$ is a wff of L^{i+1}.

   c) If $\underline{\phi}$, $\underline{\psi}$ are wff of L^i, so are $\underline{\sim\phi}$, $\underline{\phi \vee \psi}$, for any $i \geq 0$. (And, as usual, we have $\underline{\phi \rightarrow \psi} =_{\text{df}} \underline{\sim\phi \vee \psi}$, and $\underline{\phi \ \& \ \psi} =_{\text{df}} \underline{\sim(\sim\phi \vee \sim\psi)}$.)

   d) If $\underline{\phi}$ is a wff of L^0, $\underline{K^1\phi}$ belongs to the K-fragment of L^1. Formulas in the K-fragment of L^i are wff of L^i, for any $i \geq 1$.

   e) If $\underline{\phi}$ belongs to the K-fragment of L^i, so does $\underline{\sim\phi}$. Nothing else belongs to the K-fragment of L^i.

   f) If $\underline{\phi}$ belongs to the K-fragment of L^i, $\underline{K^{i+1}\phi}$ belongs to the K-fragment of L^{i+1}, for any $i \geq 1$.

   g) If $\underline{\phi}$, $\underline{\psi}$ are wff of L^0: $\underline{\mathrm{P}^1(\phi) = r}$, $\underline{\mathrm{P}^1(\phi \mid \psi) = s}$ belong to the P-fragment of L^1 (for any r, s in $[0,1]$). Formulas in the P-fragment of L^i are wff of L^i, for any $i \geq 1$.

   h) If $\underline{\phi}$, $\underline{\psi}$, belong to the P-fragment of L^i, so do $\underline{\sim\phi}$, $\underline{\phi \vee \psi}$. Nothing else belongs to the P-fragment of L^i.

   i) If $\underline{\psi}$ is in the P-fragment of L^i, then $\underline{\mathrm{P}^{i+1}(\phi) = r}$ belongs to the P-fragment of L^{i+1}, for any $i \geq 1$ and any r in $[0,1]$.

   j) If $\underline{\phi}$, $\underline{\psi}$ are wff of L^i, and either $\underline{\phi}$ or $\underline{\psi}$ belongs to the P-fragment of L^i, then $\underline{\mathrm{P}^{i+1}(\phi \mid \psi) = r}$ belongs to the P-fragment of L^{i+1}, for any $i \geq 1$ and any r in $[0,1]$.

   k) Nothing else is a wff of $L^0, L^1, \dots L^n, \dots$

To keep with the spirit of our prior terminology, at times it will be convenient to use "$\underline{R(w)}$" (for $w \in W$) as a shortcut for the relevant wff of L^0, such that for any $w \in W$, "$\underline{R(w)}$" is true in all worlds x such that wR^1x. Also, if "$\underline{\phi}$" is a sentence of $L^0, L^1, \dots L^n, \dots$, then $[\phi]$ is the set of worlds in which $\underline{\phi}$ is true.

4. v is a function that maps atomic formulas of L^0 into sets of worlds.

   Then the assessment of sentences in the model follows the usual pattern:

   For any $w \in W$:

$$\begin{array}{lll}
\models_w \underline{p_j} & \text{iff} & w \in v(p_j) \\
\models_w \underline{\sim\phi} & \text{iff} & \not\models_w \underline{\phi} \\
\models_w \underline{\phi \vee \psi} & \text{iff} & \text{either } \models_w \underline{\phi} \text{ or } \models_w \underline{\psi} \\
\models_w \underline{K^i\phi} & \text{iff} & \forall x \in W\text{: if } wR^ix, \text{ then } \models_x \underline{\phi}\text{[20]} \\
\models_w \underline{\mathrm{P}^i(\phi) = r} & \text{iff} & P^i_w(\underline{\phi}) = r\text{[21]} \\
\models_w \underline{\mathrm{P}^i(\phi \mid \psi) = r} & \text{iff} & P^i_w(\underline{\phi} \mid \underline{\psi}) = r\text{[22]}
\end{array}$$

5. For any $i \geq 0$, R^i satisfies the following property:

   $$(+) \qquad \forall w \forall x \in W(wR^{i+1}x \to x \in [K^i\ldots K^1R(w)])$$

   [In the next section we will discuss a rationale for demanding (2) and (5), as well as further possible requirements for the Rs.][23]

6. $P_{prior}(-)$ is a probability function on sentences of $L^0, L^1, \ldots L^n, \ldots$, and $P_{prior}(-|-)$ is a conditional probability function on pairs of sentences of $L^0, L^1, \ldots L^n, \ldots$, such that

   1. $P_{prior}(\underline{\phi}) = \#\{w : \models_w \underline{\phi}\} \ / \ \#W$; and
   2. $P_{prior}(\underline{\phi} \mid \underline{\psi}) = \#\{w : \models_w \underline{\phi} \ \& \models_w \underline{\psi}\} \ / \ \#\{w : \models_w \psi\}$

7. For any $i > 1$ and any $w \in W$, $P^i_w(\underline{\phi})$ is an unconditional probability function on the P-fragment of L^{i-1}, such that $P^i_w(\underline{\phi}) = P_{prior}(\underline{\phi} \mid \underline{K^{i-1}\ldots K^1R(w)})$.

   If $i = 1$, $\underline{\phi}$ belongs to L^0, and we have $P^1_w(\underline{\phi}) = P_{prior}(\underline{\phi} \mid \underline{R(w)})$.

8. For any $i > 1$ and any $w \in W$, $P^i_w(\underline{\phi} \mid \underline{\psi})$ is a conditional probability function, where $\underline{\phi}$ and $\underline{\psi}$ belong to L^{i-1}, and at least one of them belongs to the P-fragment of L^{i-1}, such that $P^i_w(\underline{\phi} \mid \underline{\psi}) = P_{prior}(\underline{\phi} \mid \underline{\psi \ \& \ K^{i-1}\ldots K^1R(w)})$.[24]

   If $i = 1$, both $\underline{\phi}$ and $\underline{\psi}$ belong to L^0, and we have $P^1_w(\underline{\phi} \mid \underline{\psi}) = P_{prior}(\underline{\phi} \mid \underline{\psi \ \& \ R(w)})$.

A few comments are in order. To simplify, at times we will use the notation "R^+" to refer to higher-order R^is (for $i > 1$). In addition, let $\mathscr{P}^i$ be the set of wff of L^i, for any i; let $\mathscr{P}^i_{\mathrm{K}}$ be the set of wff of the K-fragment of L^i, and let $\mathscr{P}^i_{\mathrm{P}}$ be the set of wff of the P-fragment of L^i. We then have $\mathscr{P}^i = Cn(\mathscr{P}^{i-1} \cup \mathscr{P}^i_K \cup \mathscr{P}^i_P)$ (where "Cn" is the Tarskian operator for logical consequence), as well as $\mathscr{P}^0 \subset \mathscr{P}^1 \subset \ldots$ Notice that $\underline{K^i} : \mathscr{P}^{i-1}_K \to \mathscr{P}^i_K$ (for $i \geq 1$), so $\underline{K^i}$ is not strictly speaking an "operator" and $\mathscr{P}^i_K$ is not closed under Boolean connectives. This is, I think, as it should be, considering the intended meaning of the formalism (see below). In any case, to keep the terminology simple, I will continue to refer to the sequence of $\underline{K^i}$s as a sequence of knowledge operators, in a loose sense. A similar point applies to probabilistic sentences within the sequence of languages in the model. We actually have $\underline{\mathrm{P}^i}(-) : \mathscr{P}_{\mathrm{P}}^{i-1} \to \mathscr{P}_{\mathrm{P}}^{i}$, as well as $\underline{\mathrm{P}^i}(-|-) : (\mathscr{P}_{\mathrm{P}}^{i-1} \times \mathscr{P}^{i-1}) \cup (\mathscr{P}^{i-1} \times \mathscr{P}_{\mathrm{P}}^{i-1}) \to \mathscr{P}_{\mathrm{P}}^{i}$. As with their knowledge counterparts, I will speak loosely of "probability operators" to refer to the $\underline{\mathrm{P}^i}$s.

Let me stress that, according to the intended interpretation, an expression such as "S has second-order knowledge that p" (*i.e.*, (*) "$\underline{K^2 p}$") does not make sense. To have second-order knowledge means that, *on reflecting on our beliefs*, we find it to be the case that we know or do not know that p. Hence it is just appropriate to require that "$\underline{K^2}$" always be followed by a first-order operator or its negation; the interpretation of higher-order levels of knowledge follows the same spirit. This is in strike contrast with the intended meaning of first-order knowledge: to have first-order knowledge that p (*i.e.*, "$\underline{K^1 p}$") means that, on reflecting *on the world*, we find it to be the case that p.[25] In short, the existence of different K-operators highlights the intuition that, when an agent reflects on her mental states, she is not dealing with the same type of phenomenon as when, say, she sees a tree in front of her. Incidentally, note that the attitude we take towards ignorance in the first-order case is typically very different from the attitude we take towards ignorance in higher-order levels. Intuitively, ideal agents could well be assumed to be aware of their knowledge states (*i.e.*, of what they positively know), whereas they are very rarely assumed to be empirically omniscient. This reinforces the motivation for having different knowledge operators.

We can of course discuss how much higher up in the hierarchy actual agents are able to grasp well formed sentences. This is an empirical question, and one we should not worry about in this context. Clearly, we should not put *a priori* limitations to the levels agents could reach on careful reflection.

The present framework validates *Modus Ponens* and *Generalized Necessitation*: if $\vdash \underline{K^i \phi}$, then $\vdash \underline{K^{i+1} K^i \phi}$ (for any $i \geq 0$), as well as the following axioms:

- **K**: $\underline{K^i(\phi \to \psi) \to (K^i\phi \to K^i\psi)}$

  [Notice that $\underline{K^i(\phi \to \psi)}$, is a wff only for $i = 1$ or, trivially, for $i = 0$.]

- **Generalized T**: $\underline{K^i\phi \ \to \ \phi}$ [for any $i \geq 1$]

- **KK$^+$**: $\underline{K^i \ldots K^1\phi \ \to \ K^{i+1}K^i \ldots K^1\phi}$ [for any $i \geq 1$]

(See the Appendix, propositions **1** to **3**.)

Notice that we have obtained a variation of the standard ***KK*** principle for all levels *without requesting transitivity.*[26] What is doing the trick is the weaker property (+) (from clause (5)), which actually amounts to demanding that R^{i+1} composed with R^i be included in R^i. Indeed, clauses (2) and (5), which regulate the behavior of the Rs, are tailored to satisfy the self-imposed constraint that our attributions of probability and knowledge coexist in a coherent way. Other requirements can be discussed as well; intuitively, different restrictions will correspond to different degrees of idealization of the epistemic agent involved (but more on this in section 6). Notice, for example, that there might be more than one way to satisfy property (+). The most conservative strategy would be to strengthen (+) to a biconditional, in which case $[K^{i+1}K^{i}\ldots R(w)] = [K^{i}\ldots R(w)]$, for any w. Then R^{i+1} differs from R^i as little as possible without violating (+). For the least conservative way to comply with (+), just let R^{i+1} be the identity relation (Id); in particular, we might want to have Id as soon as possible – *i.e.*, at level 2:

$$(*) \qquad \forall w \forall x \in W(wR^2x \rightarrow w = x)$$

In this case $[K^2K^1R(w)]$ might be a proper subset of $[K^1R(w)]$, for some w. Notice that $R^j = Id$ is a fixed point for the model.

More modestly, we could seek to weaken (+), and demand instead that R^i satisfy:

$$(\blacklozenge) \qquad \text{For all positive } i, \exists w \forall x \in W(wR^{i+1}x \rightarrow x \in [K^{i}\ldots K^1R(w)])$$

I.e., we could demand that, at each level, $[K^i \ldots K^1R(w)]$ be non-empty at least for some w, but not necessarily for all worlds. I will provide a rationale for these requirements in the next section.

As for the intended meaning of higher-order probabilities, recall that, as opposed to *first*-order probabilities, an evidential probability claim of second-order degree is the evidential probability *of a probability statement* (a wff of a P-fragment of a language in the model). A conditional evidential probability of second-order degree, on the other hand, is either (i) the evidential probability of a well formed (first-order) probability statement conditional on another well formed probability statement, or (ii) the evidential probability of any well formed statement conditional on a probability statement, or perhaps (iii) the evidential probability of a probability statement conditional on another well formed statement that need not be itself probabilistic. Hence the restrictions on the arguments of P^i_w, as found in clauses (7) and (8). Recall that, in the present proposal, the (meta-theoretic) claims we can prove *about* the model are not among the statements expressible by languages *in* the model. Thus probability functions that help us express truths about the model (say, from the theoretician's perspective) do not conflate with probability operators of $L^1, \dots L^n, \dots$

As it should be clear from (7) and (8), our conditionalization rule will now incorporate sentences with operators $\underline{K^1}, \dots \underline{K^n}, \dots$ whose behavior is regulated by $R^1, \dots R^n, \dots$ To put it more explicitly, we will have

For $i \geq 1$:

$$P^i_w(\underline{\mathrm{P}^{i-1}(\dots \mathrm{P}^1(\phi) = r \dots) = s}) =$$

$$P_{prior}(\underline{\mathrm{P}^{i-1}(\dots \mathrm{P}^1(\phi) {=} r \dots) {=} s \ \& \ K^{i-1} \dots K^1 R(w)}) \ / \ P_{prior}(\underline{K^{i-1} \dots K^1 R(w)})$$

As before, if $i = 1$, $P^1_w(\underline{\phi}) = P_{prior}(\underline{\phi \ \& \ R(w)}) \ / \ P_{prior}(\underline{R(w)})$. As opposed to our first proposal (in section 4), the relevant sentences can no longer contain iterations of the same $\underline{K}$ operator, but they will include $i - 1$ higher-order $\underline{K}$-operators. The language level to which the argument belongs (as indicated by $\underline{\mathrm{P}}$'s super-index), determines the order of the evidential probability function whose value we are seeking to calculate, and fixes $\underline{K^{i-1}}$'s degree. *Mutatis mutandis* for conditional evidential probability.

6. Discussion

Let me discuss some of the consequences that follow from demanding specific requirements for our sequence of Rs. First of all, I would like to address a prior worry on the structure of R^1. Someone might wonder why not ask that R^1 be an equivalence relation, and avoid any further complication. But we do not want to impose such a restrictive structure on R^1. Indeed, the failure of transitivity for R^1 seems just as appropriate, given, among other things, Williamson's convincing considerations on margin of error principles for (first-order) knowledge (cf. section 1). The fact that I know that p in a close world w_1 need not mean that I still know that p in a world w_n that is utterly different from the actual one, only by virtue of there being a chain of words between w_1 and w_n, any of which differs from its neighbors only slightly. In other words, "old fashioned" violations of the (unqualified) ***KK*** Principle seem very well-motivated to me.

Now, by demanding nested Rs in higher levels we may lose ordered pairs as we go up (we let worlds be progressively more "isolated", so to speak). By demanding property (+), moreover, we ensure that we will have an analogous to $\boldsymbol{KK}$ for successive operators. I would like to stress here that these demands are not *ad hoc*.

I have argued that, in order to calculate a second-order evidential probability at world w, the agent should conditionalize over $\underline{KR(w)}$, rather than over $\underline{R(w)}$. To make this move possible, I have suggested that genuine higher-order probabilities apply to probabilistic statements rather than to sets of worlds. This idea led us to define a sequence of languages with increasingly complex probabilistic claims. As a result of this strategy we obtain that the second-order probability claim stating that the probability of $\underline{R(w)}$ in world w is 1 is also 1. Thus, in order to have knowledge and probability concepts that fit with each other we should also say that the agent *knows* that $\underline{KR(w)}$ in w. This, in turn, forces us to define a sequence of knowledge operators. To put it briefly, we want to have a sequence of Ks that reflects, with a non-probabilistic vocabulary, the idea that we have probability 1 at higher-order levels (when we do have it). The choice of the structure of the Rs is then the result of seeking that probability and knowledge claims complement each other in a coherent way.

So far I have been assuming that, if S's total knowledge allows S to give probability 1 to a particular statement (perhaps even a probabilistic statement), then we should be entitled to say that S *knows* the truth of that statement. This is indeed the crucial assumption that motivates us to define a sequence of knowledge operators, and which finally leads us to the vindication of transparency principles of some sort. Many philosophers, however, have been reluctant to accept this assumption, mostly for reasons related to the nature of infinite domains, where probability 1 is not certainty. If p's probability can be 1 and still p be false, then the inference from probability 1 to knowledge should surely fail. Or so the objection goes.

The objection, however, needs to be seriously qualified. What the objection actually does is provide us with a strong reason to distinguish between different types of probability 1 in infinite models. For an infinite W, $P^1_w(\underline{\phi}) = 1$ should not always force the agent to have $\models_w \underline{K\phi}$, *but sometimes this is indeed required*, namely, when $[R(w)] \subseteq [\phi]$. Hence, were we working with an infinite W, it would be advisable to make a distinction between cases of $P^1_w(\underline{\phi}) = 1$ in which $[R(w)] \subseteq [\phi]$, and those in which $[R(w)] \not\subseteq [\phi]$; *mutatis mutandis*, this observation applies to higher-order levels as well.

Therefore, in an infinite model the motivation for having multiple K-operators still holds. Only, when W is infinite we are no longer entitled to make a rhetorical move from $\models_w \underline{\mathrm{P}^{i+1}(\ldots(\mathrm{P}^1(\phi)=1)\ldots)=1}$ to $\models_w \underline{K^{i+1}\ldots K^1\phi}$, but rather from $[K^i\ldots K^1R(w)] \subseteq [K^i\ldots K^1\phi]$ (which *a fortiori* enables us to have $\models_w \underline{\mathrm{P}^{i+1}(\ldots(\mathrm{P}^1(\phi)=1)\ldots)=1}$, as is obvious) to $\models_w \underline{K^{i+1}\ldots K^1\phi}$. Here I will not make further comments on how the structure of such an infinite model might go; a more detailed account will be left for further work. But, in the meantime, these remarks should suffice to appease some worries.

Are clauses (1) to (8) from section 5 enough to secure that the model has all the consequences we would like it to have? Yes, for the most part – although some observations are in order.

The satisfaction of property (+) guarantees that no set of the form $[K^i\ldots K^1R(w)]$ will ever be empty, for any w and any i, regardless of the structure of the original R^1. This, in turn, guarantees the existence of evidential probabilities beyond level 2, which will be obtained by conditionalization on higher order evidence corpora, expressed by $\underline{K^i\ldots K^1R(w)}$. From a philosophical point of view, the resulting system can be said to describe an ideally responsible agent, insofar as epistemically responsible agents are expected to know all they know, as discussed in part A of this paper. In such a system we obtain the validity of principle $\underline{K^i\ldots K^1\phi \rightarrow K^{i+1}K^i\ldots K^1\phi}$, as we have already mentioned, for all $\underline{\phi}$ in L^0 and any $i \geq 1$. We have called it ***the KK$^+$ Principle*** (see the Appendix, proposition **3**). For reasons to be discussed shortly, we might also want to consider the restriction of the generalized version of ***KK***$^+$ to the second level, or $\underline{K^1\phi \rightarrow K^2K^1\phi}$; let me dub it ***the KK^{+2} Principle***. Likewise, we might also want to refer to the restriction of ***KK***$^+$ for $i > 1$ as ***the KK^{++} Principle***.

Moreover, if R^2 is the identity relation (*i.e.*, if our model fulfills property (*)), we obtain an even stronger assumption on the agent's introspective capabilities. Indeed, property (*) does much more than validate the ***KK***$^+$ principle: it also validates a version of negative introspection, to wit: $\sim K^i\phi \rightarrow K^{i+1}\sim K^i\phi$; just notice that, for any $w \in W$ and any $i > 1$, if $R^+ = Id$, then $\{w\} = [K^i\ldots K^1R(w)] \subseteq [K^1R(w)]$. Unconditional probability claims do not mandate property (*) (though of course they do not exclude it either), but we will have more to say on this point in section 7, at the time of considering *conditional* evidential probabilities.

If R^2 does not satisfy (+) (hence clearly it does not satisfy (*) either), we relax the coherence demand imposed on our system, because there could be a world w such that $P^2_w(\underline{\mathrm{P}^1(\phi)=r})=1$ and yet $\not\models_w \underline{K^2K^1R(w)}$. As is obvious, if $\underline{K^2K^1R(w)}$ is not true in w, P^3_w will not be defined – nor will any other upper-level evidential probability

P^i_w, for $i \geq 3$. We can interpret this result, from a philosophical point of view, as evidence that, when we are less than ideally epistemically responsible, higher-order probability assessments will not always be meaningful (say, because our epistemic capabilities are limited). When higher-order probabilities are undefined in w, we will say that *the agent does not have responsible knowledge of* $R(w)$.

A system in which the agent does not have responsible knowledge in *any* world is a system in which the coherence between probabilities and knowledge is extremely poor. In the absence of property (+), requirement ($\blacklozenge$) from section 5 is enough to guarantee that there is at least some world in which probability and knowledge claims do not part ways. In such a world the agent's beliefs are "in the region of responsibility", so to speak. Were we to enrich a system of this type with **S5** alethic operators, as long as the accessibility relation for possibility were in fact identical to $W \times W$, we would obtain the validity of $\diamond(K^i \ldots K^1\phi \ \rightarrow \ K^{i+1}K^i \ldots K^1\phi)$. Let us call it ***the KK$^\diamond$ Principle***. Once again, we could choose to pay attention to a particular instance of ***KK$^\diamond$***, restricted to the second level, to wit: $\diamond(K\phi \rightarrow K^2K\phi)$; I will refer to it as ***the KK$^{\diamond 2}$ Principle***. Let me emphasize here that, as we have just seen, relaxing the internal coherence of the system has immediate consequences for the degree of idealization of the epistemic agents we seek to model. Thus, by relaxing a bit the internal coherence of the system we can get to model less than perfectly idealized responsible agents.

As I have already mentioned, it is worth noticing that we have obtained the validity of transparency principles of sorts without demanding transitivity for the Rs. Property (+) is weaker, in the sense that transitivity at level 1 suffices for (+) to hold (at all levels), whereas the converse is not true: property (+) does not impose transitivity, at any level (see the Appendix, propositions **6** and **7**). Moreover, property (+) cannot be satisfied at a single specific stage (for some $i > 0$) without simultaneously being satisfied at all levels.[27] This is of course not true for transitivity. Given that we may lose pairs as we go up, the Rs can become or stop being transitive without major restrictions – though, of course, in the limit, the smallest possible R^i is the identity relation, and hence trivially transitive. Notice, in addition, that transitivity at level 1 will suffice for all formulas $K^i \ldots K^1\phi \ \rightarrow \ K^{i+1}K^i \ldots K^1\phi$ to hold, for $i \geq 1$ (see the Appendix, **Corollary** to proposition **6**); however, if transitivity starts at higher-order stages it will not have the desired effect on conditionals of lower-level languages (see proposition **8**, in the Appendix, for a sketch of how to obtain straightforward counterexamples). To put it differently, the effect of transitivity on transparency statements percolates all way up, but not down.

A worry we might have at this point is whether principle (+), though weaker than transitivity at level 1, is not in fact still too strong. We might want our system to draw a difference between linking knowledge claims between level 1 and level 2, and progressing from level 2 to still higher-order levels. In other words, we might want to distinguish $\boldsymbol{KK}^{+2}$ from $\boldsymbol{KK}^{++}$, in case we think there is room to contend that $\boldsymbol{KK}^{++}$ should be valid even if $\boldsymbol{KK}^{+2}$ is not. The philosophical motivation for this suggestion might go as follows. Suppose, first, that we are interested in modeling less-than-perfectly-responsible subjects, but still *minimally* responsible. In this scenario $\boldsymbol{KK}^{+2}$ fails, whereas $\boldsymbol{KK}^{\Diamond 2}$ can be assumed to hold. However, we might also contend that the general (unrestricted) $\boldsymbol{KK}^{\Diamond}$ principle is too weak to account for the behavior of the system at higher-order stages, and hence that we should request $\boldsymbol{KK}^{++}$. The argument could be that, although knowing that one knows is certainly hard to achieve, once we achieve it, knowing that one knows that one knows comes for free (*mutatis mutandis* for even higher-order claims): once we succeed in being able to reflect on our beliefs and find out that we know that p (when we do know it), then we are already in the business of introspection, so to speak, thus still higher-order reflections on what we positively know do not add anything essentially different, phenomenologically speaking, to our prior experience of consciously finding out that we know that p. According to this picture, $\boldsymbol{KK}^{++}$ would no longer be a normative principle rooted in responsibility considerations, but a factual claim – or perhaps a normative claim stating a conceptual link between different levels of introspective knowledge.[28]

I am not convinced whether this is indeed the best way to go; in any case, it is nice to notice that the present proposal can be suitably adapted to deal with this possibility. Thus, if we want to represent less-than-perfectly-responsible beings while nonetheless assuming that higher-order transparency is conceptually required, we should request that our model validates $\boldsymbol{KK}^{\Diamond 2}$ and $\boldsymbol{KK}^{++}$, but not $\boldsymbol{KK}^{+2}$. In order to obtain this result, we can exploit the fact that transitivity, unlike property (+), can be restricted to higher-order levels. In particular, we can replace (+) by the weaker property ($\blacklozenge$) and set transitivity for levels $i > 2$, but not for R^1.[29]

Some results concerning probability claims are also worth mentioning here; we will have more to say about probability in the next section. We can prove that $P^2_w(\underline{\mathrm{P}^1(\phi)=r})$ will be always 1 when r is either 0 or 1, whereas (unconditional) evidential probability at level 2 will be guaranteed to be 1 or 0 if R^1 is an equivalence relation. It is interesting to notice that if R^1 is not transitive and the evidential probability at level 1 for an arbitrary proposition $[\phi]$ is r, for $0 \neq r \neq 1$, then the corresponding evidential probability at level 2 (*i.e.*, the evidential probability that the first level probability is r) need not be 1, which means that going up in the hierarchy will not always be trivial (see the Appendix, propositions **9** to **11**).

We shall say that the general principles $\boldsymbol{KK}^{+}$ and $\boldsymbol{KK}^{\Diamond}$, as well as the more restricted $\boldsymbol{KK}^{\Diamond 2}$, $\boldsymbol{KK}^{+2}$ and $\boldsymbol{KK}^{++}$, are **quasi-transparency principles**, which describe different degrees of idealization we can demand from agents. Notice that the validity of $\boldsymbol{KK}^{+}$, $\boldsymbol{KK}^{\Diamond}$, $\boldsymbol{KK}^{\Diamond 2}$, $\boldsymbol{KK}^{+2}$ or $\boldsymbol{KK}^{++}$, in each case, was not imposed from the outside, as it were, but was obtained as a consequence of the natural injunction to conditionalize over increasingly higher orders of evidence, while at the same time attempting to adjust probability-language and knowledge-language in a progressively coherent way.

7. Some Remarks on Probabilistic Reflection

Some of the results highlighted in section 6 referred to the way lower- and higher-level unconditional probabilities relate to each other in the model. In this section I will discuss briefly some ways to link lower- and higher-level *conditional* probabilities. In particular, we may wonder about the correction of the so-called *Reflection Principle* in probability. To distinguish it from $\boldsymbol{KK}$ let me refer to it as *the Probabilistic Reflection Principle*, or PRP. A standard formulation of PRP goes as follows:

$$P(\alpha | P(\alpha) = r) = r$$

(where "α" is a proposition and $r \in [0, 1]$). There are many possible interpretations of the principle in the literature, which shall not be discussed here.[30]

Consider, first, how to translate it to our present notation. Clearly, the relevant probability function should be (at least) a second-level function. Thus, PRP has it that the (second-order) evidential conditional probability of a sentence $\underline{\phi}$ of L^0 in a given world w, given the truth of the sentence stating that the probability of $\underline{\phi}$ is r, is itself r:

$$P^2_w(\underline{\phi} \mid \underline{\mathrm{P}^1(\phi) = r}) = r \text{ (for } w \in W),$$

which, in turn, should be rendered as:

$$P_{prior}(\underline{\phi} \mid \underline{\mathrm{P}^1(\phi) = r \;\;\&\;\; K^1 R(w)}) = r$$

Moreover, our present framework enables us to formulate PRP for higher levels, such as:

$$P^3_w(\underline{\phi} \mid \underline{\mathrm{P}^2(\phi \mid \mathrm{P}^1(\phi) = r) = r}) = r.$$

More generally, we can have

$$P^i_w(\underline{\phi} \mid \underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ \ldots)\ldots) = r}) = r,$$

or, equivalently:

$$P_{prior}(\underline{\phi} \mid \underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ \ldots)\ldots) = r \ \ \& \ \ K^{i-1}\ldots KR(w)}) = r;$$

I will refer to it as *Iterated PRP*. We will say that Iterated PRP is a theoretical truth of a model $\mathscr{M}$ that satisfies clauses (1) to (8) from section 5 iff for every w in W in which the relevant probabilities exist, $P^i_w(\underline{\phi} \mid \underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ \ldots)\ldots) = r}) = r$, for any $i \geq 2$. More flexible combinations can be discussed as well. For example, Iterated PRP could hold for certain levels only (say, for some $i \geq 2$), or just for some worlds of W. In case it is true for some worlds and not for others, we may label such worlds as particularly desirable, from an epistemic point of view – as we did in the previous section with epistemically optimal worlds that guarantee the existence of higher-order unconditional probabilities.

Is *Iterated PRP* a theoretical truth of $\mathscr{M}$? Or, at the very least, does it hold for some $i \geq 2$? And, in case the answer is negative, how bad is this result? Clearly, the truth of probabilistic reflection depends on the structure of the Rs. As we shall see in a moment, without additional requirements, a model fulfilling clauses (1) to (8) does not make PRP true, and cannot guarantee the truth of Iterated PRP for any finite i. However, with a few additional demands – some of them already discussed in previous sections – some version of the principle can be secured.

In any case, let me point out first that it is not clear to me whether we are entitled to ask that PRP be satisfied *when working with evidential probabilities*. Recall that “the evidential probability that p”, for an agent S, is actually rendered as: “the probability that p, given all S knows”. Then, “the evidential probability that p, given the truth of the sentence stating that the probability that p is r” is equally rendered as “the probability that p, given that $\underline{\mathrm{P}^1(p) = r}$ *and* given that, on reflecting on her beliefs, S finds it to be the case that she knows that…”. But the second conjunct may well affect how confident S is in “$\underline{\mathrm{P}^1(p) = r}$”; in the limit, S can even be sure in w that “$\underline{\mathrm{P}^1(p) = r}$” is false, for some w, in which case $P^2_w(\underline{p} \mid \underline{\mathrm{P}^1(p){=}r})$ will be undefined. Thus, PRP should only hold when S's knowledge does not have the chance to affect S's confidence in the truth of "$\underline{\mathrm{P}^1(p) = r}$". Assuming all S knows is $R(w)$, this will be the case only when $[K^1R(w)] \subseteq [P^1(p) = r]$, *i.e.*, when $P^2_w(\underline{\mathrm{P}^1(p) = r}) = 1$.

As it happens, Williamson has proven, for his own setting (a frame $\langle W, R, P_{prior} \rangle$ with a single accessibility relation) the following proposition: whenever the relevant probabilities are not undefined, the (probabilistic) reflection principle holds for any arbitrary proposition iff the frame is quasi-reflexive, quasi-symmetric and transitive.[31] As R can be assumed to be reflexive (due to the factivity of knowledge), this amounts to saying that R should be an equivalence relation. We can easily translate this result to our present system, and we will obtain that, when the relevant conditional probabilities are not undefined, PRP holds (at the lower level) for any arbitrary sentence of L^0 iff R^1 is an equivalence relation; the proof can be generalized for higher-order levels as long as higher-order functions remain the same (*i.e.*, as long as $R^+ = R^1$).

I do not think, however, that we should demand symmetry and transitivity at the lower level *just in order to guarantee that PRP holds* – it should be clear that violations of the principle motivated by the fact that S's knowledge affects S's confidence in $\underline{\mathrm{P}^1(p) = r}$ need not be a symptom of S's irrationality. Moreover, I have already argued why it is not sensible to demand that, for every well-behaved epistemic model, R^1 be transitive – hence it is not sensible to demand that it be an equivalence relation.

At higher levels, however, other considerations become important, as we have seen – after all, we already accepted several restrictions on R^+, motivated by our search for coherence. Shall we demand that higher-order R^is be equivalence relations, then, even though R^1 is not? Interestingly, this might not be enough to make Iterated PRP a theoretical truth of $\mathscr{M}$ (for $i > 2$). The reason is that, in our system, successive relations can become increasingly smaller as we go up. Hence, having successive equivalence accessibility relations beyond level 2 *is not sufficient* for the satisfaction of higher-order Iterated PRP. We should also demand that consecutive equivalence relations be identical, *i.e.*, that they do not lose further ordered pairs. More precisely, we can prove that:

For any given $i \geq 2$:

> If $R^{i-1} \in \mathscr{M}$ is an equivalence relation, and $R^i = R^{i-1}$ (*i.e.*, if we do not lose any ordered pair when progressing from level $i - 1$ to level i), then
>
> for all $w \in W$ and all $\underline{\phi} \in L^0$ such that the relevant probabilities are not undefined:

$$P_w^{i+1}(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ \ldots) = s) = r}) = r.$$

(see the Appendix, proposition **12**)

As stated, the conditional tells us that Iterated PRP holds at any given level $i+1$ (for $i \geq 2$), if certain restrictions on both R^i and R^{i-1} are satisfied. Iterated PRP, however, may hold at level $i+1$ and still fail for lower levels; hence s and r (in "$P_w^{i+1}(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ \ldots) = s) = r})$ $= r$") need not coincide.

This result tells us that Iterated PRP becomes true once a symmetric and transitive R^+ stabilizes. Hence, *if satisfying Iterated PRP is taken to be important on independent grounds*[32] we have good reasons to ask for a stable R^+ as soon as possible; thus, we have a strong motivation to adopt $R^2 = R^+ = Id$. Clearly, Iterated PRP is fulfilled in this case, for all $i > 2$. As we can see, securing the satisfaction of Iterated PRP has important consequences for the strength of introspective capabilities we should be ready to demand from an ideally responsible agent.[33]

8. Relation to Other Work

There are a number of recent papers that also attempt to vindicate introspective principles within epistemic logic. In particular, there are clear links between the model I have just presented and the formalism proposed by Jérôme Dokic and Paul Egré in [6], and by Egré in [7]. One of their main motivations is to deactivate Williamson's soritic argument against luminosity; in order to achieve their goal they distinguish between *perceptual* and *reflective* knowledge, each of which gets captured by a different operator. Their model then validates

$$(\mathrm{KK}') \qquad\qquad \mathrm{K}_\pi\phi \rightarrow \mathrm{KK}_\pi\phi$$

where "$\mathrm{K}_\pi\phi$" stands for "the agent has perceptual knowledge of ϕ". The authors then show that Williamson's soritic argument is blocked once operators K and K_π are suitably distinguished from one another. Thus, transparency failures at the perceptual level do not generalize to the reflective level.

There are, however, some obvious differences between the two approaches. Dokic/Egré do not offer a probabilistic framework, and they are concerned with reflections on perceptual knowledge, exclusively – hence knowledge operators beyond the second level are not allowed. Finally, $\boldsymbol{KK'}$ is not a consequence of independent decisions on the formal structure of the system. In any case, the model for quasi-transparency presented here can be seen as a refinement of the system proposed by Dokic and Egré.[34]

On a different line, notice that the present formalism can offer a straightforward solution to various epistemic paradoxes, such as Fitch's paradox, provided we enrich the model with alethic modalities and quantifiers ranging over propositions. In a nutshell, Fitch [9] has shown that the knowability principle (all truths are knowable) collapses with omniscience (all truths are known), on the assumption that the K-operator is factive and conjunction-distributive; as is well known, a crucial step in Fitch's proof is the derivation of the necessity of a statement denying that an agent can know an unkonwn truth, to wit: "$\underline{\sim K(\phi \ \& \sim K\phi)}$". But this is no longer a well formed formula in our model, so the proof cannot go through. More generally, Moorean-like conjuncts are not proper objects for knowledge or ignorance in the first place, unless the second conjunct expresses the attitude of a *different* agent; this result holds for principled reasons that are independent of the worries raised by epistemic paradoxes.[35]

9. Conclusions

In the first part of this paper I suggested that epistemic transparency is not a demand of rationality, but of ideal responsibility, and hence that ideally responsible agents verify transparency principles. I also argued that the appropriate reflective stance required by ideal responsibility need not collapse with a justification stance, so the satisfaction of reflective principles is not meant to be tied to an internalist epistemology.

In any case, the central argument of the paper in favor of transparency was addressed along sections 3 to 8, and proceeded indirectly through the development of a formal system. The core of the formal argument relied on an attempt to make probabilistic and knowledge claims fit with each other smoothly. I showed that, once we understand that higher-order evidential probabilities require conditionalization over higher-order bodies of evidence, a coherent epistemic framework will lead us to validate several introspective principles; I have dubbed them *quasi-transparency* principles. It should be emphasized that quasi-transparency principles were not just assumed to hold, but they have been obtained as a result of implementing a number of natural constraints on the structure of the system. Thus, formally speaking they behave quite differently from presuppositions of consistency or deductive closure.

Once we arrive at quasi-transparency principles because of formal reasons, we can check whether the system is adequate, from a philosophical perspective. I believe it clearly is. Notice, for example, that, in order to make knowledge claims sensitive to different orders of probabilistic statements, we ended up with a proposal in which iterated knowledge operators belong to increasingly richer languages; hence the framework as a whole vindicates the idea that higher-order knowledge is crucially different from first-order knowledge. I have argued that this is in agreement with a number of independent intuitions. In the first place, the attitude we adopt towards the fact that agents are typically more or less ignorant of the world (at the first level of knowledge) is normally very different from the attitude we adopt towards their ignorance at higher levels. We take ideally responsible agents to be aware of their own knowledge states – we *demand* them to be so aware – whereas we neither assume nor demand that epistemically responsible agents be empirically omniscient. Second, resorting to different operators is in agreement with the intuition that higher-order knowledge does not make room for "margin of error" principles (as Dokic and Egré [6] were right to point out). It is also in agreement with the idea that second-order knowledge is basically concerned with the possible "ratifiability" of first-order states.

Most importantly, a system that validates quasi-transparency principles vindicates the central suggestion put forward in section 2: even though second- (and higher-) order knowledge is not a demand of rationality, it is nonetheless an important desideratum of ideal epistemic subjects. It is an ideal we seek to fulfill to conceive of ourselves, not merely as rational creatures, but as full-fledged *agents.*

Appendix

Let $\mathscr{M} = \langle W, R^1, \dots R^n \dots, P_{prior}, v\rangle$ satisfy clauses (1) to (8) from section 5, unless otherwise noted.

Proposition 1. **(*K*)** For $\underline{K^1(\phi \to \psi)}$, $\underline{K^1(\phi)}$, and $\underline{K^1(\psi)} \in L^1$:

$\underline{K^1(\phi \to \psi) \to (K^1\phi \to K^1\psi)}$ is valid in $\mathscr{M}$.

Proof. Trivial from the definition of true-in-a-world for sentences with K^i-operators, and the fact that $\mathscr{M}$ validates *Modus Ponens.*

Notice that $\underline{K^i(\phi \to \psi) \to (K^i\phi \to K^i\psi)}$ holds vacuously in $\mathscr{M}$ for any $i > 1$, given that, for $i > 1$, $\underline{K^i(\phi \to \psi)}$ is not a wff of $L^0, \dots L^n, \dots$

Thus, well-discussed instances of $(\boldsymbol{K})$ in other normal systems, such as $(*)\underline{K^2(Kp \to p) \to (K^2Kp \to K^2p)}$, are not instances of $(\boldsymbol{K})$ in this model, insofar as neither $(*)\underline{K^2(Kp \to p)}$ nor $(*)\underline{K^2p}$ are wff in $\mathscr{M}$.

Proposition 2. $(\boldsymbol{T})$ For any $i \geq 1$ and any $\underline{K^i\phi} \in L^i$: $\underline{K^i\phi \to \phi}$ is valid in $\mathscr{M}$.

Proof. Trivial from the fact that R^i is reflexive, for any $i \geq 1$.

Proposition 3. $(\boldsymbol{KK^+})$ For any $i \geq 1$ and any $\underline{\phi} \in L^0$:

$\underline{K^i\ldots K^1\phi \to K^{i+1}K^i\ldots K^1\phi}$ is valid in $\mathscr{M}$.

Proof. Suppose both $\models_w \underline{K^i\ldots K^1\phi}$ and $\not\models_w \underline{K^{i+1}K^i\ldots K^1\phi}$, for some $w \in W$ (for *reductio*). Then we can show by induction on i that there is some chain Γ of worlds (not necessarily distinct) $wR^{i+1}xR^i\ldots R^1z$ such that $\not\models_z \underline{\phi}$, *i.e.*, we can arrive at a not-$\underline{\phi}$ world in $i+1$ steps, through potentially distinct $i+1$ relations. As we have $wR^{i+i}x$ at the initial stage of the chain, by property (+) we obtain $\models_x \underline{K^i\ldots K^1R(w)}$. Hence for any chain of worlds such that $xR^iyR^{i-1}\ldots R^1s$, we have $\models_s \underline{R(w)}$ (again, by induction on i); in particular, this holds for chain Γ, in which case $s = z$, and thus $\models_z \underline{R(w)}$. This means that we have wR^1z, and, as $\underline{\phi}$ is false in z, $\models_w \underline{\sim K^1\phi}$; as the Rs are reflexive for any level, we finally obtain $\models_w \underline{\sim K^i\ldots K^1\phi}$, which contradicts the assumption. Hence $\models_w \underline{K^i\ldots K^1\phi \to K^{i+1}K^i\ldots K^1\phi}$, for any $w \in W$.

Corollary. For all $i \geq 1$, and any $\underline{\phi} \in L^0$: $\underline{K^i\ldots K^1\phi \leftrightarrow K^{i+1}K^i\ldots K^1\phi}$ is valid in $\mathscr{M}$, even though $R^1\ldots, R^{i+1}\ldots$ may well be distinct.

Proof. Straightforward from $(\boldsymbol{KK^+})$ and $(\boldsymbol{T})$

Proposition 4. Let $R^+ = R^2$ be the identity relation (Id). Then property (+) is satisfied, and hence $(\boldsymbol{KK^+})$ holds.

Proof. Suppose $wR^{i+1}x$. We have to show that $x \in [K^i\ldots K^1R(w)]$, *i.e.*, that for all chains $xR^iyR^{i-1}\ldots R^1z$, we have wR^1z. As R^+ is the identity relation, any such chain is in fact $wR^iwR^{i-1}\ldots wR^1z$, so the result holds trivially.

Proposition 5. $(\boldsymbol{KK}^{\Diamond})$ Suppose $\mathscr{M}^*$ satisfies clauses (1) to (4) from section 5, but not clause (5) (so property (+) does not hold). Let $R^i \in \mathscr{M}^*$ also satisfy the following property, for all $i \geq 1$:

$$(\blacklozenge) \qquad \exists w \forall x \in W(wR^{i+1}x \to x \in [K^i\ldots K^1R(w)])$$

Then for any $\underline{\phi} \in L^0$ and any $i \geq 1$ there is some $w \in W$ such that $\models_w \underline{K^i\ldots K^1\phi \to K^{i+1}K^i\ldots K^1\phi}$.

Proof. Straightforward from **3**.

Proposition 6. Assume R^i is a reflexive relation over W, for all $i \geq 1$, and transitive at least for $i = 1$. Assume moreover that $R^i \subseteq R^{i-1} \ldots \subseteq R^1$, for all $i \geq 1$, whereas clauses (3) and (4) on language formation and valuation are as before. Then property (+) holds, *i.e.*, for any worlds $w, x \in W : wR^{i+1}x \to x \in [K^i\ldots K^1R(w)]$.

Proof. Take any $w, x \in W$ such that $wR^{i+1}x$, and suppose (for *reductio*) that $x \notin [K^i\ldots K^1R(w)]$. Thus $\underline{K^i\ldots K^1R(w)}$ is false in x, which means that there is some chain of worlds $wR^ixR^{i-1}\ldots R^1z$, (not necessarily distinct) such that $\not\models_z \underline{R(w)}$. However, as higher-order relations never add pairs, we also have $wR^1xR^1\ldots R^1z$; by transitivity, we obtain wR^1z, and hence $\models_z \underline{R(w)}$. Contradiction.

Proposition 7. Property (+), together with the assumption that the R^is are reflexive and nested, does not impose transitivity on any R^i.

Proof. We can build a straightforward counterexample. Consider a model in which:

$$R^1 = \{(w, w), (x, x), (y, y), (z, z), (w, x), (x, z), (y, x), (y, z)\}$$

$$R^+ = \{(w, w), (x, x), (y, y), (z, z), (x, z), (y, x)\}$$

It is easy to check that property (+) is satisfied; notice in particular that no set of the form $[K^i\ldots K^1R(-)]$ is empty. More precisely, we will have, for any $i > 1$:

$$\{w\} = [K^i\ldots K^1R(w)] = [K^1R(w)] \subset [R(w)] = \{w, x\}$$

$$\{x, z\} = [K^i \ldots K^1 R(x)] = [K^1 R(x)] = [R(x)]$$

$$\{y, x\} = [K^i \ldots K^1 R(y)] \subset [K^1 R(y)] = [R(y)] = \{y, x, z\}$$

$$\{z\} = [K^i \ldots K^1 R(z)] = [K^1 R(z)] = [R(z)]$$

However, none of the Rs is transitive.

Proposition 8. Suppose $\mathscr{M}^*$ satisfies requirements (1) to (4) from section 5, but not property (+). Suppose moreover that $R^i \in \mathscr{M}^*$ is only transitive for $i \geq n > 1$. Then $\mathscr{M}^*$ does not guarantee the validity of sentences of the form $\underline{K^i \ldots K^1 \phi \to K^{i+1} K^i \ldots K^1 \phi}$, for $i < n$.

Proof. Consider the following blueprint for an (infinite) family of models $\mathscr{M}^*_n$, one for each $n > 1$, where each $\mathscr{M}^*_n$ provides counterexamples to the validity of sentences $\underline{K^i \ldots K^1 \phi \to K^{i+1} K^i \ldots K^1 \phi}$ for all $i < n$. Let $W \in \mathscr{M}^*_n$ contain distinct $n+1$ worlds; we will assume worlds can be ordered, so as to have $x_1 \ldots x_n, x_{n+1}$. Next we define $R^i \in \mathscr{M}^*_n$ as follows:

(a) For $i < n$:

Let $(v, w) \in W \times W$. Then $(v, w) \in R^i$ iff:

1. $v = w$; or
2. $v = x_k$ and $w = x_{k+1}$ $(1 \leq k \leq n)$; or
3. $n > 2$; $v = x_1$ and $w = x_1 \vee \ldots \vee x_{i-1}$ $(2 \leq i \leq n-1)$; or
4. $i = 1$; $v = x_1$ and $w = x_n$.

(In other words, R^i is reflexive and we have $x_1 R^i x_2 R^i x_3 \ldots R^i x_{n+1}$, for any $i < n$. Moreover, world x_1 progressively loses its connections with distant worlds as we go up. At the starting point, in R^1, world x_1 can reach every other world in W except for x_{n+1}. At R^2 we lose both (x_1, x_n) and (x_1, x_{n-1}); R^3 no longer has (x_1, x_{n-2}), etc. At the final step, in R^{n-1}, world x_1 can only reach x_2 and itself.)

(b) For $i \geq n$: $R^i = Id \cup \{(x_1, x_2)\}$ (hence $R^{i \geq n}$ is transitive).

Then sentence $\underline{K^{n-m} \ldots K^1 R(x_1) \to K^{n-m+1} K^{n-m} \ldots K^1 R(x_1)}$ will be false at world x_m, for $1 \leq m \leq n-1$. The proof proceeds by induction on n, and is left to the reader. Just notice that we will have, for all $\mathscr{M}^*_n (n > 1)$:

$$\varnothing = [K^n \ldots K^1 R(x_1)] \subset [K^{n-1} \ldots K^1 R(x_1)] \subset \ldots \subset [K^1 R(x_1)] \subset [R(x_1)].$$

Proposition 9. For any $\underline{\phi} \in L^0$: if $R^1 \in \mathscr{M}$ is an equivalence relation, then for any $w \in W$, and any $r \in [0,1]$, $P^2_w(\underline{\mathrm{P}^1(\phi) = r})$ is either 1 or 0.

Proof. If R^1 is an equivalence relation over W, then for any $x \in [R(w)]$: $[R(x)] = [R(w)] = [K^1R(w)] = [K^1R(x)]$, hence $P^1_w(\underline{\phi}) = P_{prior}(\underline{\phi} \mid \underline{R(w)}) = P_{prior}(\underline{\phi} \mid \underline{R(x)}) = P^1_x(\underline{\phi})$. Moreover, we also have $P^2_w(\underline{\mathrm{P}^1(\phi) = r}) = P_{prior}(\underline{\mathrm{P}^1(\phi) = r} \mid \underline{K^1R(w)}) = P_{prior}(\underline{\mathrm{P}^1(\phi) = r} \mid \underline{K^1R(x)})$ $= P^2_x(\underline{\mathrm{P}^1(\phi) = r})$. Now, $P_{prior}(\underline{\mathrm{P}^1(\phi) = r} \mid \underline{K^1R(w)}) = \#\{y \in W : \models_y \underline{\mathrm{P}^1(\phi) = r}$ & $\models_y \underline{K^1R(w)}\}$ / $\#\{y \in W : \models_y \underline{K^1R(w)}\}$. But, as we have seen, all y in $[K^1R(w)]$ coincide in the probability they give to $\underline{\phi}$; in particular, either all of them give $\underline{\phi}$ probability r, or none does. Hence, either all y in $[K^1R(w)]$ make $\underline{\mathrm{P}^1(\phi) = r}$ true, or none does. Thus $P^2_w(\underline{\mathrm{P}^1(\phi) = r})$ is either 1 or 0.

Proposition 10. For any $\underline{\phi} \in L^0$ and any $w \in W$: If $P^1_w(\underline{\phi}) = r = 1$ or 0, then $P^2_w(\underline{\mathrm{P}^1(\phi) = r}) = 1$ (regardless of how R^1 is).

Proof. Suppose $P^1_w(\underline{\phi}) = 1$. Then we have $\models_x \underline{\phi}$, for all $x \in [R(w)]$. Moreover, for every $y \in [K^1R(w)]$ and all z such that yRz, $\models_z \underline{R(w)}$, *i.e.*, $z \in [R(w)]$. Hence $\models_z \underline{\phi}$, and $\models_y \underline{\mathrm{P}^1(\phi) = 1}$. As this is the case for every y in $[K^1R(w)]$, $P^2_w(\underline{\mathrm{P}^1(\phi) = 1}) = 1$.

Symmetrically, suppose now $P^1_w(\underline{\phi}) = 0$. Then for all $x \in [R(w)]$, $\not\models_x \underline{\phi}$. Again, for every $y \in [K^1R(w)]$ and all z such that yRz, $\models_z \underline{R(w)}$, hence $\not\models_z \underline{\phi}$, and $\models_y \underline{\mathrm{P}^1(\phi) = 0}$. As this holds for every y in $[K^1R(w)]$, $P^2_w(\underline{\mathrm{P}^1(\phi) = 0}) = 1$.

Corollary. For any $\underline{\phi}$ in the P-fragment of L^{i-1} and any $w \in W$: If $P^i_w(\underline{\phi}) = r = 1$ or 0, then $P^{i+1}_w(\underline{\mathrm{P}^i(\phi) = r}) = 1$.

Proof. Straightforward from **10**.

Proposition 11. Suppose $P^2_w(\underline{\mathrm{P}(\phi) = r}) = s$, for $0 \neq r \neq 1$. Then, if $R^i \in \mathscr{M}$ is not transitive, s need not be either 1 or 0.

Proof. For a counterexample, suppose $W = \{w, x, y, z\}$, and suppose R^1 is reflexive and symmetric, but not transitive; let (x, w), (x, y) and (y, z) be in R^1, but not (x, z). Let also $[\phi] = \{w\}$. Then $P^1_x(\underline{\phi}) = 1/3$, and $P^2_x(\underline{\mathrm{P}(\phi) = 1/3}) = 1/2$.

Proposition 12. **Sufficient conditions for Iterated PRP**
For any given $i \geq 2$:
If $R^{i-1} \in \mathscr{M}$ is an equivalence relation, and $R^i = R^{i-1}$, then for all $w \in W$ and all $\underline{\phi} \in L^0$ such that $P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r})$ is not undefined: $P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) = r$.

Proof. We assume that $R^{i-1} \in \mathscr{M}$ is an equivalence relation, and that $R^i = R^{i-1}$. Assume also $P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r})$ is not undefined. We have to show that $P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r})$ is equal to r.

1) By definition,

$$P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) =$$
$$P_{prior}(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots)= s) = r \ \ \& \ \ K^i\ldots K^1R(w)}) =$$
$$\frac{\#\ ([\phi]\ \cap\ [\mathrm{P}^i(\phi|\mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r]\ \cap\ [K^i\ldots K^1R(w)])}{\#\ ([\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r]\ \cap\ [K^i\ldots K^1R(w)])}$$

2) As R^i is an equivalence relation and the conditional probability in (1) is defined, we have $P^{i+1}_w(\underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) = 1$ (by proposition **9**) and hence $[K^i\ldots K^1R(w)] \subseteq [\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r]$. Thus,

3) From 1 and 2:

$$P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) =$$
$$\#\ ([\phi]\ \cap\ [K^i\ldots K^1R(w)])\ /\ \#\ [K^i\ldots K^1R(w)]$$

4) Moreover, as $R^i = R^{i-1}$: $[K^i\ldots K^1R(w)] = [K^{i-1}\ldots K^1R(w)]$. Hence,

5) From 3 and 4:

$$P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) =$$
$$\#\ ([\phi]\ \cap\ [K^{i-1}\ldots K^1R(w)])\ /\ \#\ [K^{i-1}\ldots K^1R(w)]$$

6) As R^{i-1} is an equivalence relation, we have, by proposition **9**,

$$P^i_w(\underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s}) = 1$$

and hence

$$[K^{i-1}\ldots K^1 R(w)] \ \subseteq\ [\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi| \ldots)\ldots) = s]$$

7) Hence, from 5 and 6

$$P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) = \frac{\#\ ([\phi]\ \cap\ [\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s]\ \cap\ [K^i\ldots K^1 R(w)])}{\#\ ([\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s]\ \cap\ [K^i\ldots K^1 R(w)])}$$

8) However, by definition:

$$P^i_w(\underline{\phi} \mid \underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s}) = \frac{\#\ ([\phi]\ \cap\ [\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s]\ \cap\ [K^i\ldots K^1 R(w)])}{\#\ ([\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s] \cap [K^i\ldots K^1 R(w)])}$$

9) Hence, from 7 and 8

$$P^i_w(\underline{\phi} \mid \underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s}) = P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r})$$

10) But, by line 2, we have $\models_y \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}$, for all y such that $\models_y \underline{K^i\ldots K^1 R(w)}$. In particular, as reflexivity guarantees that $w \in [K^i\ldots K^1 R(w)]$, we have $\models_w \underline{\mathrm{P}i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}$.

11) Hence, from 10

$$P^i_w(\underline{\phi} \mid \underline{\mathrm{P}^{i-1}(\phi \mid \mathrm{P}^{i-2}(\phi \mid \ldots)\ldots) = s}) = r$$

12) Hence, from 9 and 11

$$P^{i+1}_w(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \mathrm{P}^{i-1}(\phi \mid \ldots) = s) = r}) = r.$$

Notice that demanding equivalence alone is not sufficient for Iterated PRP. Suppose $R^i \neq R^{i-1}$, even though they are both equivalence relations. We can prove that, if $R^i \subset R^{i-1}$, there is some world in which, if Iterated PRP is well defined, Iterated PRP is false for $i+1$, for some sentence of L^0: if R^{i-1} is an equivalence relation but $R^i \neq R^{i-1}$, there are some w, x such that $x \in [K^{i-1}\ldots K^1R(w)]$ but $x \notin [K^i\ldots K^1R(w)]$. Just take $[\phi] = \{x\}$. Then, if it exists, $P_w^{i+1}(\underline{\phi} \mid \underline{\mathrm{P}^i(\phi \mid \ldots) = r}) = s$, where $r = 1 / \#[K^{i-1}\ldots R(w)]$, and $s = 0$ (because there is no world in $[K^1\ldots R(w)]$ in which $\underline{\phi}$ is true).

Notes

[1] Thus, for instance, a hypochondriac could falsely think she is in pain (cf. Williamson [29, p. 535]; cf. also Sosa [23], chapter 2). For an interesting discussion on this point see Leitgeb [14].

[2] An important antecedent of what I intend to do here can be found in Dokic and Egré [6], and Egré [7]. I will come back to their proposal in Section 8 of this paper.

[3] For a discussion of this point cf. Christensen [5], chapter 6.

[4] Thus, according to the view I favor, epistemic responsibility and epistemic rationality can very well come apart – and they often do. Most authors simply assume that they amount to more or less the same thing (cf. for example Owens's description of what he dubs the "juridical theory of responsibility", in Owens [18]).

[5] Cf. Engel [8], Hieroyimi [12], or Owens [18], among others. According to this tradition, discussions about epistemic responsibility lead us to the problem of epistemic voluntarism. Thus we could be tempted to reason as follows: beliefs are voluntary only in a much deflated sense, but nonetheless we can be subjected to criticism for having a particular doxastic corpus rather than other; hence the necessary conditions for epistemic responsibility cannot be too stringent. In Owens [18] we find an interesting attempt to preserve a deflated sense of responsibility even under the assumption that agents have no freedom whatsoever regarding their own doxastic states.

[6] Cf. for example Bilgrami [1, 2]. Bilgrami argues that we should follow Strawson in thinking of freedom as not purely metaphysical, but normative. Freedom is then defined by the "reactive attitudes" (blame, criticism, resentment) we find in ourselves and in others; moreover, according to Bilgrami it is also defined by the normative reactions we can *justify* with our values. He then contends that we cannot justify our criticism of an agent's beliefs or desires unless we assume the agent to have self-knowledge of her own intentional states. Thus, self-knowledge is a necessary condition for responsibility, and the following conditional holds: "To the extent that an intentional state is in the region of responsibility, *i.e.*, to the extent that an intentional state is the rational cause of an action which is the object of justifiable reactive attitudes, or to the extent that an intentional state is itself the object of a justifiable reactive attitude, then that intentional state is known to its possessor." ([1, p. 218]). Exceptions to self-knowledge are precisely signs of the inapplicability of the normative conditions specified in the antecedent.

[7] Cf. Chapter 1 of Foley [10] for an alternative account on epistemic responsibility that also puts reflection at the center stage.

[8] Several philosophers explored this path before, particularly within the pragmatist tradition. Isaac Levi, for one, has written extensively on the so-called "Belief-Doubt model", of Peircean roots, according to which we should not devote our

energies to justify prior beliefs, but to justify belief *changes*; cf. chapter 1 of Levi [15], among other places. From a somewhat different perspective – but still in a similar spirit – Michael Williams has argued that justification possesses a default-and-challenge structure; cf. for example Williams [27]; cf. also Brandom [4].

[9] Just to clarify, the claim is not that, if S knows that p, then S will always know that she knows that p, regardless of whether her second order belief is or is not justified. Rather, the claim is: if S is ideally responsible and S knows that p, then (i) S has duly reflected on whether p; (ii) S notices that she is convinced that 'p' is true and approves of her being so convinced (hence she believes that she knows that p); and (iii) either S is aware of having good reasons to believe that p (thus S is justified in believing that she knows that p), or S's beliefs, including her second-order belief that she knows that p, is not in the conceptual space required for justification to be meaningful in the first place.

[10] Notice that a statement such as "S does not know that p" is systematically ambiguous between (i) p is false and S knows it to be false; and (ii) S suspends judgment on whether p. The second meaning is clearly the one we should focus on for the present discussion.

[11] Williamson [28].

[12] Hence P_{prior} is regular, in the sense that $P_{prior}(\phi) = 0$ iff $\phi = \varnothing$. Williamson takes priors in his system to refer to the intrinsic plausibility of worlds prior to our gathering any evidence (cf. also his [30], chapters 9 and 10). If we feel uncomfortable with this extremely objectivist picture, we can always take priors to embody the personal measures of the theoretician – who can in turn be conceptualized as the subject who seeks to make knowledge attributions to third party agents.

[13] For a straightforward illustration, let $W = \{x, y_1, \ldots, y_n, z\}$, and let R be reflexive and non-transitive, with $(x, y_1), \ldots (x, y_n), (y_1, z), \ldots (y_n, z) \in R$; notice that $(x, z) \notin R$. Then $R(x) = \{x, y_1, \ldots, y_n\}$. As always, $P_x(R(x)) = 1$, since, by definition, all worlds in $R(x)$ are reachable from to x. However, as all y_k are connected to z, and z is not in $R(x)$, $P_{y_k}(R(x)) < 1$, for all y_k $(1 \leq k \leq n)$. Hence the proposition $[\mathrm{P}(R(x)) = 1]$ is just the singleton $\{x\}$. Hence $P_x(\mathrm{P}[(R(x)) = 1)]) = P_{prior}([\mathrm{P}(R(x)) = 1] \cap R(w)) \;/\; P_{prior}(R(x)) = 1/n + 1$, because there are $n + 1$ worlds in $R(x)$. This amounts to a sort of probabilistic failure of ***KK***, according to Williamson. Notice that $K(R(x))$ is true in x, whereas $KKR(x)$ is not, *i.e.*, in world x the agent does not know that she knows that $R(x)$. Even worse, if we consider frames with an increasingly larger number of y-worlds, intuitively, as $P_x([\mathrm{P}(R(x)) = 1)])$ approaches 0, not only does the agent ignore in x that she knows that $R(x)$, but she also takes her knowledge of $R(x)$ to be *very improbable.*

[14] We might rather choose to refine W and allow for metaphysically impossible worlds – say, worlds in which "χ" and "ψ" are neither both true nor both false, in spite of being logically equivalent (thanks to Timothy Williamson for this suggestion). But the approach adopted in this paper seems more natural, and respects the intuition that purely linguistic differences sometimes matter, even to fully rational agents.

[15] For other well-known frameworks that deal with higher-order probability, cf. Skyrms [22], Gaifman [11], Samet [21], or van Fraassen [26].

[16] For systematization purposes, I find it convenient to keep open the possibility to write formulas of L^0 with no probability operators as "$\underline{\mathrm{P}^0\phi}$", for any $\underline{\phi}$ in L^0.

[17] Suppose $R = \{(w, w), (w, x), (x, x), (x, y), (y, y)\}$. Then we have $[R(w)] = \{w, x\}$, $[KR(w)] = \{w\}$, and $[KKR(w)] = \varnothing$. However, as we conditionalize on $[KR(w)]$,

$P^2_w(\underline{\mathrm{P}^1(R(w)=1)}) = P_{prior}(\underline{\mathrm{P}^1(R(w)=1)\ \&\ KR(w)})\ /\ P_{prior}(\underline{KR(w)}) = \#\{w\}\ /\ \#\{w\} = 1$.

[18] As with probability operators, for systematization purposes I find it convenient to keep open the possibility to write formulas of L^0 with no knowledge operators as "$\underline{K^0\phi}$", for any $\underline{\phi}$ in L^0.

[19] I follow Williamson in demanding that W be finite, but this restriction can of course be abandoned – in which case some of the clauses that follow would need to be appropriately amended.

[20] In other words, for all i, $[K^i\phi] = \{y \in W : \forall x \in W(yR^ix \rightarrow x \in [\phi])\}$. Notice that, for $i > 1$, "$\underline{K^i\phi}$" is a wff only if "$\underline{\phi}$" is of the form "$\underline{K^{i-1}\psi}$" or "$\underline{\sim K^{i-1}\psi}$", in agreement with clauses (3.d), (3.e) and (3.f).

[21] Notice that "$\underline{\mathrm{P}^i(\phi) = r}$" is well formed only if "$\underline{\phi}$" belongs to the P-fragment of L^{i-1}.

[22] Notice that "$\underline{\mathrm{P}^i(\phi|\psi) = r}$" is well formed only if both "$\underline{\psi}$" and "$\underline{\phi}$" belong to L^{i-1}, and either "$\underline{\psi}$" or "$\underline{\phi}$" belongs to the P-fragment of L^{i-1}.

[23] Notice that, if $i = 0$, then (+) becomes the trivial claim that, for any w, x such that wR^1x, $x \in [R(w)]$, which is of course true by definition of $[R(w)]$.

[24] This flexibility regarding the nature of the arguments of conditional evidential functions will enable us to establish some important links between lower- and higher-order probabilities, as we shall see in section 7.

[25] Related to this, notice that, even though "$\underline{q\ \&\ K^1q}$" is expressible in L^2 (by clauses (3.b) and (3.c)), (*)"$\underline{K^2(q\ \&\ K^1q)}$" is not a wff of L^2. This should not be counterintuitive, once we consider what a second-order knowledge operator means in the model. It could be objected that agents sometimes do express sentences such as "I know both that the train has just arrived and that I know it has come from abroad". But it can well be argued that in such cases what the agent actually wants to convey is *both* that she knows that the train has just arrived, *and* that she has second-order knowledge that she knows that the train has come from abroad. The correct way of rendering this idea would then be "$\underline{K^1q\ \&\ K^2K^1q}$", which is a perfectly well formed formula in the model. Similar remarks hold for cases in which an agent S reflects on the knowledge possessed by third party agents (which, from S's point of view, is actually akin to reflecting on features *of the world*), and compares it with S's own knowledge. In a similar fashion, more complex sentences that appear to involve ideas not expressible in any language of the sequence need to be suitably reinterpreted.

Is this a reasonable strategy? I believe it is. Arguably, there is no such thing as *the* correct logical form of an English assertion: what counts as an adequate formalization depends in each case on the goals and interests of our system. Ultimately, successful theorization always involves some trade-off; the unconvinced reader can take the (eventual) non-standard formalization as the cost to pay in order to get other positive features. Among other things, the chosen rules for language formation enable us to formulate knowledge claims while keeping an eye on different levels of probabilistic discourse. Moreover, distinguishing among levels of K-operators captures important philosophical intuitions on the concept of knowledge, which we should be reluctant to sacrifice.

[26] Thanks to an anonymous referee for pointing this out to me.

[27] Just notice that, as the Rs are nested and reflexive, every time we have $wR^{i+1}x$, we will also have both $wR^{i+1}wR^ix$ and $wR^{i+2}wR^{i+1}x$; thus, for any chain of worlds $x_1R^{i+1}x_2R^i \ldots R^1x_{i+2}$, the satisfaction of property (+) for R^{i+1} guarantees that we

can reach any world x_n in the chain from any other world x_m (for $m \leq n$) in any finite number of steps – hence property (+) is satisfied at all levels.

[28] Bonnay and Egré [3] offer an interesting defense of this perspective. "[E]ven if one upholds the view that knowing that one knows is essentially more difficult than simply knowing, one might still consider possible that iterations of knowledge stop making a difference at some point beyond two or more iterations. A hint that this may be so is provided by the difficulty of ascribing knowledge beyond two levels of iteration in ordinary language (thus, a sentence like "he knows that he knows, but he does not that he knows he knows" sounds nearly contradictory)" [3, 3.2]. Alternatively, we could take this phenomenon to be a "side-effect of some limited capacity to compute metarepresentations." [3, 3.2]. Either case, they propose to deal with it by means of their *token semantics*; as I hope to show, the present account offers an alternative way to obtain the same result.

[29] Notice that we could always restrict transitivity to even higher stages, as is obvious.

[30] The interested reader is referred to Skyrms [22].

[31] Corollary 5, informal communication [additional notes and proofs to be (perhaps) incorporated to Williamson [28]]. Similar results hold for a regular countable additive probability distribution over a serial frame (Corollary 7). Thanks to Horacio Arló-Costa for pointing out these results to me.

[32] For example, it has been contended that violations of (certain instances of) PRP make agents susceptible to Diachronic Dutch Books (DDB) (cf. in particular van Fraassen [24, 25, 26]). It should be noted, however, that DDB arguments are even more controversial than PRP itself, so the rhetorical move from DDB to PRP needs to resort to additional considerations to be effective.

[33] Recall that, if $\mathscr{M}$ is such that $R^2 = Id$, a version of negative introspection becomes valid in $\mathscr{M}$.

[34] See also Bonnay and Egré [7] for an alternative mechanism to assess higher-order knowledge claims that bears some resemblance with the present proposal; cf. footnote 28.

[35] In Linsky [16] we find a detailed account of how a stratified approach to knowledge can deal with several epistemic paradoxes. The present proposal differs from the one favored by Linsky in many crucial respects; most importantly, our proposal can hardly be accused of being *ad hoc*. Its successful treatment of Fitch's paradox is a happy *consequence* of our framework, rather than a motivation for adopting it. See also Paseau [19].

Acknowledgements

Special thanks are due to Horacio Arló-Costa, Paul Egré, Ignacio Viglizzo, Timothy Williamson, and an anonymous referee for *The Journal of Philosophical Logic*. I am also deeply grateful to Ramiro Caso for his help in editing this document. Previous versions of this paper have been presented at talks and workshops held at the Instituto de Investigaciones Filosóficas, UNAM (Mexico City, Mexico); Universidad de Bahía Blanca (Bahía Blanca, Argentina); Universidad Torcuato Di Tella (Buenos Aires, Argentina); GAF, Universidad de Buenos Aires

(Buenos Aires, Argentina); UFESP (Sao Paulo, Brazil); University of Southern California (Los Angeles, US; presented at *The Formal Epistemology Workshop 2011*, with comments by Eric Pacuit), and Universität Konstanz (Konstanz, Germany). I would like to thank the participants and the commentator for their discussions and suggestions.

References

1. Bilgrami, A.: 1999, 'Why is self-knowledge different from other kinds of knowledge?'. In: L. E. Hahn (ed.): *The Philosophy of Donald Davidson*, Library of Living Philosophers. Chicago: Open Court, pp. 211–224.
2. Bilgrami, A.: 2006, *Self-Knowledge and Resentment.* Harvard: Harvard University Press.
3. Bonnay, D. and P. Egré: 2009, 'Inexact Knowledge with Introspection'. *Journal of Philosophical Logic* **38**, 179–227.
4. Brandom, R.: 1994, *Making It Explicit.* Cambridge, Ma.: Harvard University Press.
5. Christensen, D.: 2004, *Putting Logic in its Place: Formal Constraints on Rational Belief.* Oxford: Clarendon Press.
6. Dokic, J. and P. Egré: 2009, 'Margin for Error and the Transparency of Knowledge'. *Synthese* **166**, 1–20.
7. Egré, P.: 2008, 'Reliability, Margin for Error, and Self-Knowledge'. In: V. Hendricks and D. Pritchard (eds.): *New Waves in Epistemology.* New York: Pagrave Macmillan, pp. 215–250.
8. Engel, P.: 2009, 'Epistemic Responsibility without Epistemic Agency'. *Philosophical Explorations* **12**(2), 205–219.
9. Fitch, F.: 1963, 'A Logical Analysis of Some Value Concepts'. *Journal of Symbolic Logic* **28**, 135–152.
10. Foley, R.: 2001, *Intellectual Trust in Oneself and Others.* Cambridge: Cambridge University Press.
11. Gaifman, H.: 1986, 'A Theory of Higher-Order Probabilities'. In: B. Skyrms and W. Harper (eds.): *Causation, Chance, and Credence.* Dordrecht: Kluwer Academic Publisher, pp. 191–219.
12. Hieronymi, P.: 2005, 'The Wrong Kind of Reason'. *Journal of Philosophy* **102**(9), 427–457.
13. Hintikka, J.: 1962, *Knowledge and Belief.* Ithaca: COrnell University Press.
14. Leitgeb, H.: 2002, 'Critical Study of *Knowledge and its Limits*'. *Grazer Philosophischen Studien* **65**, 195–205.
15. Levi, I.: 1997, *The Covenant of Reason.* Cambridge: Cambridge University Press.
16. Linsky, B.: 2009, 'Logical Types in some Arguments about Knowability and Belief'. In: J. Salerno (ed.): *New Essays on the Knowability Paradox.* Oxford: Oxford University Press, pp. 163–179.
17. Nozick, R.: 1981, *Philosophical Investigations.* Cambridge: Cambridge University Press.
18. Owens, D.: 2000, *Reason without Freedom: The Problem of Epistemic Normativity.* London: Routledge.

19. Paseau, A.: 2008, 'Fitch's Argument and Typing Knowledge'. *Notre Dame Journal of Formal Logic* **49**(2), 155–176.
20. Rescher, N.: 2005, *Epistemic Logic: A Survey of the Logic of Knowledge.* Pittsburgh: University of Pittsburgh Press.
21. Samet, D.: 1997, 'On the Triviality of High-Order Probabilistic Beliefs'. Game Theory and Information 9705001, EconWPA.
22. Skyrms, B.: 1980, 'Higher Order Degrees of Belief'. In: *Prospects for Pragmatism. Essays in Honor of F. P. Ramsey*, Vol. 1. Oxford: Clarendon Press, pp. 109–137.
23. Sosa, E.: 2007, *A Virtue Epistemology: Apt Beliefs and Reflective Knowledge*, Vol. 1. Oxford: Clarendon Press.
24. van Fraassen, B.: 1984, 'Belief and the Will'. *Journal of Philosophy* **86**, 235–256.
25. van Fraassen, B.: 1989, *Laws and Symmetry.* Oxford: Oxford University Press.
26. van Fraassen, B.: 1995, 'Belief and the Problem of Ulyses and the Sirens'. *Philosophical Studies* **77**, 7–37.
27. Williams, M.: 2001, *Problems of Knowledge: A Critical Introduction to Epistemology.* Oxford: Clarendon Press.
28. Williamson, T., 'Very Improbable Knowing'.
29. Williamson, T.: 1995, 'Is Knowing a State of Mind?'. *Mind* **104**, 533–565.
30. Williamson, T.: 2000, *Knowledge and its Limits.* Oxford: Oxford University Press.