Knowledge, Justification, and Reason-Based Belief

Egré, P.; Marty, P.; Renne, B.

DOI
10.48550/arXiv.1412.1862

Publication date
2014

Document Version
Other version

Knowledge, Justification, and Adequate Reasons

Paul Égré
CNRS / ENS / EHESS
PSL University

Paul Marty
ZAS Berlin

Bryan Renne
University of Saskatchewan

Abstract

Is knowledge definable as justified true belief ("JTB")? We argue that one can legitimately answer positively or negatively, depending on whether or not one's true belief is justified by what we call *adequate* reasons. To facilitate our argument we introduce a simple propositional logic of *reason-based belief*, and give an axiomatic characterization of the notion of adequacy for reasons. We show that this logic is sufficiently flexible to accommodate various useful features, including quantification over reasons. We use our framework to contrast two notions of JTB: one internalist, the other externalist. We argue that Gettier cases essentially challenge the internalist notion but not the externalist one. Our approach commits us to a form of infallibilism about knowledge, but it also leaves us with a puzzle, namely whether knowledge involves the possession of only adequate reasons, or leaves room for some inadequate reasons. We favor the latter position, which reflects a milder and more realistic version of infallibilism.

Keywords: Knowledge; Belief; Justification Logics; Adequate Reasons; Infallibilism; Externalism

1 Introduction

Can the ordinary concept of knowledge be defined in terms of justified true belief ("JTB")? Since Gettier's paper [16], the answer to this question is widely considered negative. Gettier produced two cases intended to show that a belief can be true, justified, and yet fall short of knowledge. The first concerns Smith, an applicant for a job who has "strong evidence" that Jones is the man who will get the job and also that Jones has ten coins in his pocket. Unknown to Smith, it turns out that Smith himself has ten coins in pocket and is actually the one selected for the job. Smith's belief that "the man who will get the job has ten coins in his pocket" is therefore true and justified, but it seems inappropriate to say that this belief constitutes knowledge. The second case is one in which Smith believes a false proposition p based on persuasive evidence for p and infers from p some true proposition $p \vee q$ by picking the true disjunct q at random. Here too, Smith justifiably believes $p \vee q$, but it seems incorrect to say that he knows $p \vee q$.

This is a substantially revised and expanded version of our earlier paper titled "Knowledge, Justification and Reason-Based Belief", a short version of which appeared in the *Proceedings of the Amsterdam Colloquium 2015*, ed. by T. Brochhagen, F. Roelofsen and N. Theiler (pp. 100-108). Paul Égré was funded by the ANR project "Trivalent Logics and Natural Language Meaning" (ANR-14-CE30-0010) and by the ANR program FrontCog ANR-17-EURE-0017 for research conducted at the Department of Cognitive Studies of ENS. Bryan Renne was funded by an Innovational Research Incentives Scheme Veni grant from the Netherlands Organisation for Scientific Research (NWO) and was employed by ILLC Amsterdam at the start of this project. Paul Marty was funded by the MIT Linguistics Department when this project started.

Our point of departure in this paper is the following: even though we agree with the force of Gettier's examples, we share with others (in particular [10, 26, 27, 8, 18, 29]) the intuition that those examples do not necessarily invalidate every analysis of knowledge in terms of justified true belief, depending on how the notion of justification is understood. Indeed, what Gettier's examples show is that an agent can have an internal justification for believing a proposition that is *plausible*, without that justification being properly *adequate* to the truth of the proposition in question. But if so, Gettier cases must only show that knowledge is not identical with JTB under an internalist conception of justification. The examples do not thereby rule out the existence of a more externalist notion of justification capable of sustaining the equation between knowledge and JTB. Define knowledge as true belief with an *adequate* justification, and it appears Gettier cases no longer have a bite. Our leading intuition in that regard is shared with the earlier analyses of knowledge by Chisholm in [8] and Sosa in [26, 27], who define knowledge as the possession of a *non-defective* justification, and by Dretske in [10], who defines knowledge as the possession of a *conclusive* reason.

Admittedly, a definition of knowledge along those lines might not provide a noncircular or reductive analysis (see [33]): notions of "non-defectiveness", "conclusiveness", or indeed "adequacy" may ultimately have to be understood in ways that presuppose a prior grasp of the concept of knowledge. For example, if an adequate justification were to mean "a justification that is suitable to make the belief count as knowledge," then it would appear that we define knowledge in terms of itself. We agree with this objection, but the notion of adequacy may also turn out to not wholly depend on epistemic notions. Adequacy, for instance, may be at least partly characterizable in terms of truth-making, and the truth-making relation need not refer to prior epistemic notions. Or consider the relation between a fully formalized axiomatic proof and a mathematical statement derived in that proof: the "adequacy" of the proof as a vehicle for mathematical truth is a purely syntactic notion, with no epistemological concepts presupposed. These examples suggest to us that room remains for a fruitful investigation of the concepts of knowledge, belief, and justification that acknowledges the distinction between adequate and inadequate reasons.

The gist of our account lies in the distinction between reasons that *(merely) support* belief in a proposition and reasons that are not only supportive but are also what we call *adequate*. Our first goal in this paper is to give an axiomatic characterization of both concepts, and to use them to clarify the duality found in the concept of justified true belief. Our characterization treats adequate reasons basically as externalist noninferential justifications, following Fumerton's typology in [15]. They are noninferential in the sense that we do not require agents to be able to justify their plausibility by a further reason, and externalist in that adequacy is not necessarily a property an agent can ascertain. Like Chisholm and Sosa in their respective treatments of nondefective justification, or Dretske in his treatment of conclusive reasons, we moreover treat adequate reasons as being infallible, in the sense that they can only support true propositions. The latter property is not a definition of adequacy, however, but only a central property, as we shall explain.

A second goal we have on that basis is more technical: it consists, in the wake of work done in Justification Logic [2, 3], in giving an explicit treatment of reasons in epistemic logics but with specific emphasis on the notion of an adequate reason.[1] Our third goal is more philosophical. As

[1] [2, 3] uses a formal framework to track what goes wrong with specific lines of reasoning in the examples of Gettier, Goldman, and Kripke; and [4, 5] uses a related framework that has additional features from Belief Revision Theory to reason about the examples of Lehrer and Gettier. Our work here is different. First, while inspired by the Justification Logic approach to reasoning about justifications, our setting is in many respects simplified but at the same time includes certain novel features (see §3.5 for details). Second, our task here is different than that in [2, 3]:

already suggested, our account commits us to a version of infallibilism about knowledge, for an adequate reason supports only true propositions. However, our beliefs for a proposition often are grounded in two kinds of reasons, some of them adequate, and others inadequate. The question we are interested in concerns whether knowledge should be defined in terms of the possession of only adequate reasons, or whether it tolerates the inclusion of inadequate reasons. This is what we call *the problem of mixed reasons*. We address this problem and defend the view that knowledge should be made compatible with the possession of mixed reasons.

In order to articulate the distinction between two kinds of justification, in §2 we first present the basic concepts of our account of knowledge and justification, namely the concepts of reason, support, and of adequacy. We then introduce the *Logic of Reason-Based Belief* in §3. This logic provides an explicit representation of reasons to believe a proposition. We show that the logic is sufficiently flexible to accommodate quantifiers over reasons, limited or full closure of reason-based belief under implication, and an optional requirement that all beliefs be reason-based. In §4, we put this logic to work in the analysis of Gettier cases: first to tease apart two notions of justification, one internalist and the other externalist; and second to study the susceptibility of internal JTB and of external JTB to Gettier-type examples. Finally, in §5 we close the paper with a discussion of the problem of mixed reasons and with the discussion of an objection to our account, regarding whether adequacy is even a necessary condition for the ascription of knowledge. Technical notions and results that are not required for an understanding of the main text are relegated to an appendix.

2 Reasons, Support, and Adequacy

We distinguish between two kinds of reasons: those that merely "support" a proposition by inclining an agent to believe in it, and those that not only support a proposition but are also themselves "adequate." Our analysis of Gettier cases hinges on a precise understanding of this key distinction. In this section we first introduce the three basic concepts that we rely on in our epistemology beside the concept of belief, namely reasons, support, and adequacy. Regarding belief, suffice it to say that we treat belief as an explicit endorsement by an agent of either a proposition, or a reason providing evidence for a proposition. We say more about our characterization of belief in §3, where we further clarify the relation between beliefs and reasons.

2.1 Reasons

The first concept we need to clarify is the concept of a *reason*. As emphasized by Armstrong in [1], "talk of a man's reason may be to refer to a (certain sort of) belief-state of his, or it may be to refer to the proposition believed" (pg. 79). We handle reasons primarily as state-specific objects, and only derivatively as propositions. For us, a reason is some evidence on the basis of which we come to believe a proposition. A reason therefore does not have the type of a proposition, although it can be associated to a proposition in a systematic manner.[2] For example, my hearing voices

beyond providing a formal diagnosis of what goes wrong in certain Gettier-type examples, we define different kinds of JTB and discuss their relative susceptibility to such examples. This has some similarities with the work in [4, 5]; however, our logics are much simpler and we use them as part of a different analysis.

[2]We are indebted to an anonymous reviewer and to J. Dutant for urging us to clarify that matter. In a previous version of our work, we left open the possibility for reason symbols to directly denote propositions. Note that throughout the paper, and unless when it creates confusion, we use the word "proposition" both for syntactic objects (closed sentences of our language), and for the semantic objects they denote (viewed as sets of possible worlds).

outside might be a reason to believe that there are people outside. Semantically, we shall represent reasons by accessibility relations, thereby treating them as akin to belief-states. Given a reason r and a world w, we write $r(w)$ to represent the proposition determined by the reason r in w (that is the set of accessible worlds in virtue of the relation corresponding to r). So, each reason can be associated with a proposition at a world, but reasons *per se* are not propositions.[3] Reasons are related to the sort of answers produced by rational agents when they are asked, "Why do you believe this?" or "How do you know that?". If I am asked why I believe that Napoleon I is dead by now, I would answer that it is because I read that he was born in the eighteenth century, and he died in the nineteenth century. If I am asked why I believe that $2+2=4$, I would respond that my calculations confirm that, or that I accept a certain axiomatic system that I see as consistently deriving that fact. So reasons for us are not bare propositions, like the proposition "Napoleon died in the 19th century" or the proposition "$2+1=3$", or even bare arguments, but we take them to be related to evidential experience of some sort.

2.2 Support

The second concept we deal with is what we call *support* between a reason and a proposition. Syntactically, we treat the relation of support between a reason and a proposition as primitive: we will write "$r:\varphi$". Model-theoretically, we represent the support relation between a reason r and a proposition p by the fact that $r(w)$ entails p. This does not mean that we intend the relation of support to model deductive support. Rather, we do intend it to capture a very fundamental form of inductive support. That is, we view each reason as a basic Humean experience of some kind, creating an inclination to believe a proposition. If I hear voices outside the house, that is a supporting reason for the proposition that there are people outside the house. In principle, that might also be a reason to think there are loudspeakers broadcasting voices outside the house (with no actual person around). In our approach, however, we do not allow for reasons to support contradictory propositions. If r is a supporting reason for φ, then we will take r to be a supporting reason for any proposition entailed by φ. Consequently, a reason for thinking that there are loudspeakers broadcasting voices outside the house with no one around will have to be different from a reason to think that there are actual people outside the house.

A delicate issue is whether the support relation ought to be treated as subjective (internal to an agent's beliefs) or as objective (external to the agent's beliefs). That is, could an agent believe that r supports φ without r actually supporting φ? Conversely, could r support φ without the agent believing that r supports φ? Regarding the latter, in our system any reason supports every logical truth, but we see it as possible for an agent not to believe that, in particular because an agent may not automatically see the truth of every logical truth or be aware of every reason. Hence, a reason may support a proposition without an agent believing that.

The question whether an agent may believe a reason to support a proposition without that reason supporting the proposition is more delicate. One way in which this could happen is if reasons can be treated as mere propositions. For example, an agent who accepts a fallacious mathematical argument may believe that true premises *deductively support* a false conclusion. In that case one may say that the agent incorrectly believes the premises to deductively support the conclusion. Again we find it better not to let propositions be reasons *per se*, but to keep to reasons

[3]Note also that given a sentence p, p is supposed to express the same proposition relative to every world. In our framework a reason r may express different propositions $r(w)$ depending on the world w.

being evidence of a sort, related to some experience. Take an agent who miscalculates the result of some mathematical equation. He accepts the premise that a bat and a ball cost \$1.10, that the bat costs \$1 more than the ball, and wrongly computes from that that the ball costs \$0.10 (it actually costs \$0.05).[4] On an objective reading of reasons, letting p_1 and p_2 stand for the two premises of the argument, and q for the false conclusion that the ball costs \$0.10, we would write: $B((p_1 \wedge p_2) : q)$ to express that the agent believes the premises to deductively support the conclusion. But we find more appropriate to write instead that $B(r : (p_1 \wedge p_2 \rightarrow q))$, where r stands for the calculation made by the agent, to represent the fact that the agent sees his calculations to support a certain deductive relation between premises and conclusions. As a result, we treat the relation of deductive support between bare propositions by means of the usual logical resources available in our system, and not in terms of the colon operator.

Because we see reasons as evidence-based propositions, we might wish to endorse the success principle $B(r : \varphi) \rightarrow r : \varphi$. That is, the agent believing some kind of experience to support a proposition would suffice to make the experience in question a supporting reason for φ. A rational agent may therefore not necessarily be aware of all his or her reasons, but could not be mistaken about reasons that he or she explicitly endorses. One may object that even rational agents may be deluded about their reasons. For example, an agent might be mistaken about her own experiences. She has a memory of an encounter with a famous actor, but that encounter never happened, and the memory is no real experience. When asked to justify why she believes this actor is blond, she gives as a reason that she remembered seeing him to be blond during that encounter. Arguably, the reason does not support the proposition in that case, because the reason is not grounded in any true experience. Such cases invite us to guard against the aforementioned success principle. Because of that, we will propose a weakening of that principle, intended to secure more generality, and to not rule out such cases.

2.3 Adequacy

What about *adequacy*? Adequacy is by far the most central concept in our approach. There are various ways in which the notion can be thought of. One option is to think of an adequate reason as a *reliable* reason (see [18]). On that approach, an adequate reason is a reason that produces true beliefs most of the time. Like other critics of reliabilism, we do not view this characterization as strong enough. A stronger characterization might be in terms of Dretske's notion of *conclusive* reason. According to Dretske, r is a conclusive reason for p if and only if r would not be the case if p were not the case ([10]). Another approach is in terms of Chisholm's or Sosa's respective conceptions of *non-defectiveness* [8, 26], whereby a justification is non-defective provided it is not the basis of any false proposition. Consistently with those views, our use of the term "adequate" is also related to Spinoza's understanding in *Ethics*, where the notions of adequate idea (*adaequata idea*) and adequate knowledge (*adaequata cognitio*) are meant to imply truth.[5]

We draw inspiration from the latter set of approaches but an important caveat about our approach is that we do not propose to give an explicit definition of adequacy in terms of necessary and sufficient conditions. Instead, we propose to characterize adequacy in terms of two necessary conditions, and in terms of the interaction with the notions of support and belief we have in our ontology. The first central condition we impose on a reason for it to be adequate is very close to

[4]This example is from S. Frederick and D. Kahneman ([20]). A variant is discussed by Sorensen in [25].
[5]See [28] Part II, in particular Def. 4, Prop. 34 and Prop. 41.

Chisholm's, Sosa's, or indeed Spinoza's respective conceptions. We say that a reason is adequate only if the propositions it supports are true (axiom (A) below). But we also see adequacy as putting a requirement on the support relation. In particular, we consider that if an agent wrongly believes a reason r to support a proposition φ (that is, believes so although r actually fails to support φ), then r cannot be adequate (axiom (AS) below). So, an adequate reason is such that it secures not only the propositions that it actually supports, but also such that it secures the correctness of the belief in the associated support relation. This second requirement basically corresponds to the weakening of the success condition evoked before, that $B(r:\varphi)$ should always entail $r:\varphi$. It says that this is so provided r is adequate.

Let us make four important remarks on our treatment of adequacy:

1. Our central requirement on adequacy ensures that if φ is not true, and r supports φ, then r cannot be adequate. This property of adequacy therefore comes close to Dretske's, except that we rule out the possibility that a reason might be adequate evidence for one proposition and inadequate for another (we discuss that aspect in §5). That first requirement implies that reasons that do not validly support a given conclusion are inadequate.

2. In our system $B(r:\varphi)$ does not entail that Br, which says that an agent does not automatically deem adequate any reason she has in support of a proposition. For example, in the Müller-Lyer illusion, an agent who trusts her perceptual experience that two lines differ in length has a reason in support of the proposition that the lines are not equal. But she can have a different and overriding reason, her measuring of the lines (reason s), telling her that the lines do not differ in length. We would represent the case as:

$$B(r:p) \wedge B(s:\neg p) \wedge Bs \wedge \neg Br \ .$$

3. We do not identify the adequacy of a reason with the validity or even the soundness and validity of an argument. Consider the following example of what Sorensen [25] calls "paralemmatic reasoning". Our agent has 10 cents in his pocket (premise 0). He infers from that, and from the information that the bat and ball cost \$1.10 (premise 1) and the bat costs \$1 more than the ball (premise 2), that he has enough money to buy the ball (conclusion). Let us assume he infers the conclusion really because he infers that the ball costs 10 cents. Let us represent by p_0, p_1, p_2 the premises of that argument, by q the correct conclusion that the agent has enough money to buy the ball, and by p the incorrect lemma that the ball costs 10 cents. Let us call r the agent's calculations. The agent believes that $r:(p_0 \wedge p_1 \wedge p_2 \rightarrow q)$. The argument $p_0 \wedge p_1 \wedge p_2 \rightarrow q$ has sound premises, and the conclusion validly follows. In effect, however, what r supports in the agent's mind is a more specific argument, namely the argument that $p_1 \wedge p_2 \rightarrow p$, combined with the argument that $p \wedge p_0 \rightarrow q$. Because of our adequacy constraint, r cannot be an adequate reason here, because r supports an incorrect argument. Sorensen in [25] writes that "paralemmas are precise counterexamples to bare evidentialism – the view that we know a proposition simply by virtue of having adequate evidence for it." By adequate evidence we take Sorensen to refer to some *intrinsic* property of the evidence, holding irrespective of the agent's beliefs. This is not our characterization of adequacy, precisely because for us a reason can support a logically perfect argument without being adequate. We are not bare evidentialists as a consequence.

4. Finally, we do not view the two conditions we impose on adequacy, namely (A) and (AS), as sufficient conditions for a reason being adequate, but only as necessary conditions. For

us, and as our semantics will show, adequacy is fundamentally a property of a reason and a world. Or put otherwise, the adequacy of a reason represents the property of a world being a *good case* for the reason in question (see [33]). The same reason, with the same intrinsic properties, could turn out to be adequate in a good case and inadequate in a different case that is bad. Being a good case is not a notion we think we can explicitly define and exhaustively characterize. Intuitively, being a good case (relative to a reason) means being a case at which the reason is used in a way that is not lucky, that is safe, or that has the right fit, whatever those notions might mean. We consider our two axioms on adequacy to capture part of that intuition, but some of the cases we discuss (particularly Goldman cases, see below) are arguably cases in which both of our conditions are fulfilled, but which we do not want to characterize as adequate in the externalist sense we think is relevant.

Our axiomatic treatment of adequacy therefore implies that the adequacy of a reason does not lead to any false conclusion, and it also implies that the agent is not deluded about the experience reported in the reason. But to rehearse our main point, those are only two necessary conditions on adequacy. We view adequacy as an even stronger property of a reason and a world, namely as the property for a world to secure that the reason is not merely internally but also externally warranted.

3 A Simple Logic of Reason-Based Belief

Having laid out the basic ingredients in our ontology, in this section we present a simple logic of reason-based belief. We first present our system axiomatically and then give its underlying model theory. We then present various extensions of the basic system, in particular to allow for quantification over reasons, which will be needed in our analysis of justified true belief in the next section.

3.1 Reason-Based Belief

Fix nonempty sets P of propositional letters and R of reason symbols (also called "reasons"). F is the set of formulas φ defined by the following grammar:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (r : \varphi) \mid r \mid B\varphi$$
$$p \in P,\ r \in R$$

Other Boolean connectives are defined as abbreviations.

- $r : \varphi$ says, "r supports φ".
- r says, "r is an adequate reason."

  For the sake of clarity, we might have introduced a unary operator A in the language, and written Ar to express "r is an adequate reason". We choose to write r instead of Ar to save symbols. Our logic will guarantee that every proposition supported by an adequate reason is true (i.e., we will have $r : \varphi \to (r \to \varphi)$ for all formulas φ). Note that, as we define it, adequacy is not relativized to a specific proposition. This makes our treatment of adequacy different from Dretske's treatment of conclusiveness. For Dretske, a reason is not conclusive *per se* but conclusive *for* a proposition ([10]). In our setting, when a reason is adequate, it makes every proposition it supports true. Hence no reason can be such that it is adequate relative to one proposition it supports, and inadequate relative to another. This feature of our account may eventually prove too strong (we return to this issue in the last section of the paper), but our proposal is to think of adequacy as a fundamental property of truth-conduciveness.

- $B\varphi$ says, "the agent believes φ."

  The formula Br is therefore read, "the agent believes r is an adequate reason." Sometimes it will be convenient to say that "the agent accepts reason r" to mean that Br holds. So believing reason r is adequate and accepting reason r will be considered to mean the same thing.

To reduce the number of parentheses while ensuring unique readability of formulas, we adopt the convention that the colon operator binds more strongly than any Boolean connective. For example,

$$r\!:\!\varphi \to \psi \quad \text{denotes} \quad (r\!:\!\varphi) \to \psi \quad \big[\text{and not } r\!:\!(\varphi \to \psi)\big] \ .$$

The theory RBB of *Reason-Based Belief* appears in Table 1. We write $\vdash \varphi$ to mean that φ is derivable in RBB. The negation is written $\nvdash \varphi$.

AXIOM SCHEMES

(CL)	Axiom Schemes of Classical Propositional Logic
(RK)	$r\!:\!(\varphi \to \psi) \to (r\!:\!\varphi \to r\!:\!\psi)$
(A)	$r\!:\!\varphi \to (r \to \varphi)$
(BRK)	$B(r\!:\!(\varphi \to \psi)) \to (B(r\!:\!\varphi) \to B(r\!:\!\psi))$
(BA)	$B(r\!:\!\varphi) \to (Br \to B\varphi)$
(AS)	$B(r\!:\!\varphi) \to (r \to (r\!:\!\varphi))$
(D)	$B\varphi \to \neg B\neg\varphi$

RULES

$$\frac{\varphi \to \psi \quad \varphi}{\psi}\ (\text{MP}) \qquad \frac{\varphi}{r\!:\!\varphi}\ (\text{RN}) \qquad \frac{\varphi \leftrightarrow \psi}{B\varphi \leftrightarrow B\psi}\ (\text{E})$$

Table 1. The theory RBB

Regarding the axioms and rules of RBB from Table 1, (CL) and (MP) say that RBB is an extension of classical propositional logic. (RK) says that reasons are closed under material implication, and (RN) says that reasons support all derivable formulas. (A) says that if r supports φ and r is an adequate reason, then φ is true. (BRK) says that if the agent believes r supports a conditional, and believes r supports the antecedent, then the agent believes r supports the consequent. (BA) says that if the agent believes r supports φ and the agent believes that r is an adequate reason, then the agent believes φ. (AS) says that if the agent believes r to support a proposition, then r does not support the proposition unless r is adequate. (D) says that the agent's beliefs are consistent: she cannot have contradictory beliefs (i.e., believe both φ and $\neg\varphi$ for some φ). Finally, (E) says that the agent's beliefs do not distinguish between provably equivalent formulas.

As for mnemonics, (CL) is "Classical Logic," (MP) is "Modus Ponens," (RK) is Kripke's axiom K of modal logic (used here for reasons), (RN) is "Reason Necessitation," (A) is "Adequacy," (BRK) is "the Belief version of RK," (BA) is "the Belief version of adequacy," (AS) is "Adequate Support," (D) is a well-known axiom from modal logic [7], and (E) is a well-known rule from minimal modal logic [7].

3.2 Semantics for RBB

We now present a semantics for our system. The semantics is based on two main components: a neighborhood semantics for belief, intended to make belief as weak as possible, and a Kripke semantics for reasons, this time intended to capture the closure properties on reasons.[6] We justify both desiderata in the next section.

We construct the models of RBB in a couple of stages. We first begin with *pre-models*, which are structures $M = (W, [\cdot], N, V)$ having:

- a nonempty set W of "possible worlds,"
- a function $[\cdot] : R \to \wp(W \times W)$ mapping each reason $r \in R$ to a binary relation $[r] \subseteq W \times W$ on the set of possible worlds,
- a "neighborhood function" $N : W \to \wp(\wp(W))$ mapping each world w to a collection $N(w)$ of sets of worlds ("propositions") that the agent believes at w,
- a propositional valuation $V : W \to \wp(P)$ mapping each world w to the set $V(w)$ of propositional letters that are true at w.

To indicate that the components W, $[\cdot]$, N, and V come from pre-model M, we may write W^M, $[\cdot]^M$, N^M, and V^M (respectively). Letting $w \in W$ be a world, $r \in R$ be a reason, and $X \subseteq W$ be a set of reasons (i.e., a "proposition"), we introduce the following abbreviations:

- $r(w) := \{v \in W \mid (w, v) \in [r]\}$ is the set of all worlds that are r-accessible from w (i.e., accessible by the relation $[r]$).
- $r^\circ := \{v \in W \mid (v, v) \in [r]\}$ is the set of all worlds that are r-accessible from themselves. These are the worlds at which the reason r is said to be *reflexive*. As we will see shortly, a reason that is reflexive at a world will be adequate at that world. So reflexivity and adequacy are equivalent notions, and hence we may conflate the two, which ought not cause confusion.

To indicate the sets $r(w)$ and r° arise from worlds in pre-model M, we may write $r^M(w)$ and $r^{M\circ}$ (respectively). Though not mentioned above, we do require that all pre-models satisfy the following property:

(pr) For each $x \in P \cap R$, we have $x \in V(w)$ if and only if $w \in x(w)$.

This says that for propositional letters that are also reasons, the truth assignment given to x by the valuation V agrees with the reflexivity of x. This ensures that there is no ambiguity in the assignment of truth to propositional letters that are also reasons.[7]

---

[6] For other systems dealing with belief in terms of a neighborhood semantics, see [30] and [11, 12].

[7] By definition, the nonempty sets P and R are not necessarily disjoint. Therefore, our language allows for the possibility that there are objects that are both propositional letters and reasons. The choice as to which possibility to realize is up to the user, who may decide to take $P \cap R = \emptyset$ or not as per her preference.

A *pointed pre-model* is a pair (M, w) consisting of a pre-model M and a world w in M. We write $M, w \models \varphi$ to say that φ is true at the pointed pre-model (M, w), and we write $M, w \not\models \varphi$ for the negation. We define the satisfaction relation $\models$ and the set

$$[\![\chi]\!]_M := \{v \in W \mid M, v \models \chi\}$$

of worlds in the pre-model M at which the formula χ is satisfied as follows.

- $M, w \models p$ means that $p \in V(w)$.
- $M, w \models \neg\varphi$ means that $M, w \not\models \varphi$.
- $M, w \models \varphi \vee \psi$ means that $M, w \models \varphi$ or $M, w \models \psi$.
- $M, w \models r : \varphi$ means that $r(w) \subseteq [\![\varphi]\!]_M$.
- $M, w \models r$ means that $w \in r(w)$.
- $M, w \models B\varphi$ means that $[\![\varphi]\!]_M \in N(w)$.

We note that $\models$ is well-defined: for each $x \in P \cap R$, we have $x \in V(w)$ if and only $w \in x(w)$ by (pr), and therefore $M, w \models x$ is well-defined.

We say that a pre-model $M = (W, [\cdot], N, v)$ is a *model* if and only if M satisfies each of the following additional properties:

(brk) If $[\![r : (\varphi \to \psi)]\!]_M \in N(w)$ and $[\![r : \varphi]\!]_M \in N(w)$, then $[\![r : \psi]\!]_M \in N(w)$.

This says that if the agent believes r supports the implication $\varphi \to \psi$ and its antecedent φ, then she believes r supports the consequent ψ as well.

(ba) If $[\![r : \varphi]\!]_M \in N(w)$ and $r^\circ \in N(w)$, then $[\![\varphi]\!]_M \in N(w)$.

This says that if the agent believes r supports φ and she believes r is reflexive, then she also believes φ.

(as) If $[\![r : \varphi]\!]_M \in N(w)$ and $w \in r(w)$, then $r(w) \subseteq [\![\varphi]\!]_M$.

This says that if the agent believes r supports φ and r is reflexive, then r does in fact support φ.

(d) $X \in N(w)$ implies $W - X \notin N(w)$.

This says that if the agent believes X at world w, then she does not believe the complement $W - X$ at world w.

(brk), (ba), (as), and (d) are the model-theoretic analogs of the axioms (BRK), (BA), (AS), and (D), respectively. A *pointed model* is a pair (M, w) consisting of a model M and a world w in M.

Given a pre-model M, to say that φ is *valid in* M, written $M \models \varphi$, means that $[\![\varphi]\!]_M = W$. To say that φ is *valid*, written $\models \varphi$, means that $M \models \varphi$ for every model M. That is, validity is taken over the class of models (and not the larger class of pre-models). It is shown in Theorem A.1 that RBB is sound and complete for this semantics: we have $\vdash \varphi$ if and only if $\models \varphi$. Unless explicitly noted otherwise, our focus in what follows will generally be on the concept of model (and not the concept of pre-model).

The following terminology will be useful for what follows: given a reason r and a pointed model (M, w) representing the key features of a particular situation of reason-based belief, to say

- "r is adequate at (M, w)" means $M, w \models r$;
- "r is veridical for φ at (M, w)" means $M, w \models r : \varphi \to \varphi$; and
- "r is veridical at (M, w)" means $M, w \models r : \varphi \to \varphi$ for each formula φ.

In using the above terminology, we may omit mention of (M, w) if it should be clear from context which pointed model is meant. It follows by the semantics that every adequate reason is veridical;[8] however, a reason may be veridical for a proposition without being thereby adequate. Note that being veridical is only a necessary condition for a reason to be adequate. This means we do not take veridicality to provide an explicit definition of adequacy, but only to put a constraint on what it is for a reason to be adequate.

3.3 Weak Belief But Strong Reasons

RBB posits a weak notion of belief. In particular, as reflected in the semantics, beliefs are not necessarily closed under material implication. That is, it is consistent to have

$$B(\varphi \to \psi) \wedge B\varphi \wedge \neg B\psi \ ,$$

which says that the agent believes an implication and the antecedent of the implication but not the consequent.

Reasons, on the other hand, are strong. First, as just seen, they encompass all derivable statements by (RN). Second, they are closed under implication (and hence under (MP)) by (RK). Third, if adequate, then (A) says that they are *veridical*: everything they support is true. Reasons therefore support many assertions (infinitely many, in fact, because each reason supports each of the infinitely many theorems of RBB by (RN)). This puts more requirements on reasons (i.e., they must do more things), which makes them stronger.

Reasons are governed by what is essentially the normal multi-modal logic KT (with one modal operator "$r :$" for each reason r), except that the T axiom $r : \varphi \to \varphi$ (sometimes called "veridicality") is not guaranteed to hold unless, according to (A), we make the additional assumption that r is adequate. This way of having a multi-modal logic like KT but with the "modal operator" r itself a formula (whose truth implies veridicality) is, to our knowledge, new.[9]

We have chosen a theory of weak belief but strong reasons in order to keep things as simple as possible but still address some of the major trends in the epistemological study of knowledge as justified true belief ("JTB"). In all of the examples from epistemology we consider in this paper,

[8] Proof: suppose r is adequate at (M, w). This means $M, w \models r$, which, by the definition of satisfaction, means that $w \in r(w)$. Now take an arbitrary φ. To show that $M, w \models r : \varphi \to \varphi$, assume $M, w \models r : \varphi$. By the definition of satisfaction, this assumption means $r(w) \subseteq [\![\varphi]\!]_M$. Since $w \in r(w)$, it follows that $w \in [\![\varphi]\!]_M$; that is, $M, w \models \varphi$. So r is veridical at (M, w).

[9] The usual way of writing our formula $r : \varphi$ would be $\Box_r \varphi$. So RBB may be viewed as a multi-modal logic with an extra formula r for each $r \in R$, a K-modality "$r :$" (our variant of $\Box_r$) for each $r \in R$ that respects the reflexivity scheme T if r holds (as per (A)) and interacts according to (AS) with an ED-modality B that is governed by (BRK) and (BA). Note that (AS) provides some interesting interaction between the various modal operators. According to our intended semantics (§3.2), we interpret modal operators using a possible worlds semantics (with a neighborhood function for B), and the intended interpretation of the formula r is that the binary accessibility relation corresponding to the modal operator "$r :$" is reflexive. Certain hybrid logics [6] can express reflexivity of modal operators: the formula $\downarrow x. \neg \Box_r \neg x$ says that the accessibility relation corresponding to $\Box_r$ is reflexive. However, hybrid logics generally include additional features permitting greater expressivity than we need.

we need some way to track an agent's logical inferences and some way to link these inferences to what the agent believes. Our theory RBB is a rather minimalistic way of doing just this: reasons are used to handle the relevant inferencing, the agent can "accept" a reason (or not) by believing it to be adequate (or not), the agent comes to believe things supported by reasons she accepts, and the agent's beliefs are always consistent. This way we can encode inferencing using a reason, indicate whether the agent accepts this inferencing, and thereby infer whether she believes some statement based on a reason. We also allow the possibility that she believes something without a reason, by which we mean that the set

$$\{B\varphi\} \cup \{B(r:\varphi) \to \neg Br \mid r \in R\}$$

is consistent with our theory. This assumption is metaphysically disputable, however, and we may provide for every belief to be accompanied by a reason if we wish.

3.4 Consistency of Reasons

One interesting theorem of RBB is the principle

$$\vdash (Br \wedge Bs) \to (B(r:\varphi) \to \neg B(s:\neg\varphi)) \tag{RC}$$

of *reason consistency*. This principle says that if an agent believes reasons r and s are adequate, then she cannot believe that r and s support contradictory assertions. Intuitively, the derivability of (RC) follows because (BA) requires that an agent who accepts a reason and believes that reason to support a formula must also believe that formula, and (D) requires that an agent not have contradictory beliefs. Notice that if we take $r = s$ in (RC), then we obtain a statement provably equivalent to

$$\vdash Br \to (B(r:\varphi) \to \neg B(r:\neg\varphi)) \ , \tag{IC}$$

which says that a reason believed to be adequate is believed to be internally consistent.

If the agent does not believe r is a adequate reason, then she can believe that r is internally inconsistent. Put another way, the formula

$$\neg Br \wedge B(r:\varphi) \wedge B(r:\neg\varphi)$$

is consistent with our theory. Similarly, if the agent believes r is an adequate reason but does not believe s is an adequate reason, then our theory does not rule out the possibility that she believes r and s are inconsistent. That is,

$$Br \wedge \neg Bs \wedge B(r:\varphi) \wedge B(s:\neg\varphi)$$

is also consistent with our theory.

Finally, since the theory RBB is consistent (and hence does not prove both φ and $\neg\varphi$ for some formula φ),[10] any adequate reason is internally consistent. That is,

$$\vdash r:\varphi \to (r \to \neg(r:\neg\varphi)) \ , \tag{AIC}$$

which says that a reason r that supports φ and is adequate cannot also support $\neg\varphi$. It follows from (AIC) that any internally inconsistent reason is not adequate.

[10] Consistency of RBB follows by soundness (Theorem A.1).

3.5 Logical Closure and Combination of Reasons

If ψ is a logical consequence of φ, meaning $\vdash \varphi \to \psi$, then our theory says that r is a reason to believe the consequent ψ whenever r supports the antecedent φ. That is,

$$\vdash \varphi \to \psi \quad \text{implies} \quad \vdash r : \varphi \to r : \psi \ . \tag{RCLC}$$

The proof: from $\varphi \to \psi$, we obtain $r : (\varphi \to \psi)$ by (RN). This is the antecedent of an instance of (RK), so the consequent $r : \varphi \to r : \psi$ of this instance is derivable by an application of (MP). This completes the proof. In examining this proof, we see that (RCLC) is a consequence of the stronger logical principle encompassed by our axiom

$$r : (\varphi \to \psi) \to (r : \varphi \to r : \psi) \ , \tag{RK}$$

which says that reasons are closed under material implication.

The principle (RCLC) says that reasons are closed under logical consequence. It is unclear whether this is a desirable principle. For example, it may make more sense to say that if ψ is a logical consequence of φ and r supports φ, then it is not r itself that supports the consequence ψ. Instead, what is wanted is some more complicated reason r' that not only references r but also provides some reason s as to why ψ obtains from φ. That is, we might seek a principle like this:

$$s : (\varphi \to \psi) \to (r : \varphi \to (s \cdot r) : \psi) \ . \tag{App}$$

This is the principle of *Application* from Justification Logic [3]. It says: if s supports the implication $\varphi \to \psi$ and r supports the antecedent φ, then a new object $s \cdot r$ obtained by combining s and r supports the consequent ψ. In essence, the more complex reason $s \cdot r$ keeps track of everything we would need to check to see that ψ indeed obtains: the initial reason r for the antecedent φ and a reason s for the implication $\varphi \to \psi$. Further, the syntactic structure of $s \cdot r$, with s to the left and r to the right, tells us what kind of a reason we have: based on the form of (App), it is suggested that s is some implication, r is the antecedent of that implication, and $s \cdot r$ is a reason for the consequent. In essence, we are describing specific witnesses for an instance of the rule (MP) of Modus Ponens:

$$\frac{\varphi \to \psi \quad \varphi}{\psi}\,(\text{MP}) \quad \text{is encoded by} \quad \frac{s \quad r}{s \cdot r} \ .$$

(App) is a more nuanced version of (RK): if we have $r : (\varphi \to \psi)$ and $r : \varphi$, then we do not obtain $r : \psi$ straightaway using (App). Instead, we must construct the reason $r \cdot r$ in support of ψ. The single instance of "$\cdot$" in the syntactic structure of the latter reason reflects our use of one derivational step (i.e., one instance of (MP)) to obtain ψ.

To do away with (RCLC), we must do away with (RK) and modify (RN). In particular, let R_0 be a nonempty set of "basic" reasons, define R to be the smallest extension of R_0 satisfying the property that $s, t \in R$ implies $s \cdot t \in R$, replace scheme (RK) by scheme (App), and restrict (RN) by requiring that $r \in R_0$. (It is assumed that all other schemes and rules can use reasons coming from the full set R.) Call the resulting theory RBB+(App). In RBB+(App), it is consistent to have

$$r : (\varphi \to \psi) \land r : \varphi \land \neg(r : \psi) \ , \tag{1}$$

which says that r is not closed under implication. As a result, it can be shown that (RCLC) no longer obtains. But note that (RK) and (RCLC) do not fail in RBB+(App) because logical or materially implied consequences of assertions are no longer "accessible" by some reason. Indeed, in the situation (1) under the theory RBB+(App), the logical consequence ψ of φ is still "accessible" by the reason $r \cdot r$. However, this reason $r \cdot r$ is more "complex" than the original reason r (in terms of the number of instances of the (MP)-encoding Application operator "$\cdot$" that appear inside it). In general, distant consequences that would require many repetitions of (App) are still accessible; it is just that the reasons that access these consequences may be very "complex."

Justification Logic (JL) [3] is the study of logics of reason-based belief (with reasons thought of as "justifications"). Defining JL_0 to be the fragment of RBB+(App) obtained by omitting all belief formulas and belief axioms from the theory, Justification Logics may be thought of as extensions of JL_0.[11] Many JLs permit other kinds of combinations of reasons than what we saw with (App). Our logic RBB is closely related to the JL tradition, though we conspicuously omit (App), retain (RK) (and thereby endorse (RCLC)), leave (RN) without the additional restriction, and maintain a set R of primitive reasons that cannot be combined to form more complex reasons. In so doing, we lose the ability to have the syntactic structure of reasons reflect the structure of derivations in the theory, and thereby forgo a more nuanced tracking of the interaction between logical consequence and the complexity of reasons. We accept these consequences in the interest of developing a system that is simple and yet still of use to the formal epistemologist. Nevertheless, we recognize that a reader may be interested to see a thorough study of more sophisticated extensions of our theory that allow for the combination of reasons along the lines of (App) (and perhaps for other features as well). We advise such a reader to consult the JL literature directly [3].

Since our theory does not allow the agent to combine reasons in the sense of (App) and beliefs are not closed under implication, it is consistent for us to have a situation wherein the agent has the requisite information to draw a belief but simply does not draw it. For example, assuming $s \neq r$, it is consistent to have

$$Bs \wedge Br \wedge B(s:(\varphi \to \psi)) \wedge B(r:\varphi) \wedge \neg B\psi \ , \tag{2}$$

which says that the agent believes s is adequate and believes s supports an implication (and hence believes the implication by (BA)), believes r is adequate and believes r supports the antecedent of the implication (and hence believes the antecedent by (BA)), and yet the agent does not believe the consequent (even though the believes the implication and its antecedent). The problem is that her beliefs are not closed under (MP). This is so despite the fact that, by (RCLC) and (RB), beliefs coming from *the same reason* are closed under (MP):

$$\vdash (Bt \wedge B(t:(\varphi \to \psi)) \wedge B(t:\varphi)) \to B\psi \ .$$

So long as the implication and its antecedent come from *separate* beliefs as in (2), the agent need not believe the consequent.

The consistency of (2) is a consequence of our design: we use reason operators to encode inferencing, and we use belief operators to encode the particular inferencing and the individual assertions that the agent accepts (in terms of her affirmed beliefs). As such, the situation (2) is one in which the agent has not yet performed sufficient inferencing to accept the conclusion ψ, even though she is very close (after all, she has done enough to accept the implication $\varphi \to \psi$ and its antecedent φ). In essence, this lack of closure allows us to place one kind of constraint on the

[11] Actually, Justification Logics are extensions of a fragment of RBB+(App) that places further restrictions on the rule (RN), but we set aside further discussion of this issue here.

agent's inferencing powers. If desired, one can place even more severe constraints as in [5]; however, this seems to require more syntax and additional axioms. One can also go the other way and lift these constraints entirely; in §3.6 we will suggest one natural way to do this in our setting. But for now we retain what we hope is a "happy medium" in the form our theory RBB.

3.6 Implication-Closed and Purely Reason-Based Beliefs

We saw in (2) that reason-based beliefs in RBB need not be closed under implication if the source reasons are different. If we would like to ensure that reason-based beliefs are always closed under implication, even if the beliefs come from separate reasons, then a simple solution is to introduce a "master reason" σ that encodes the combined information of all reasons the agent accepts. This requires us to expand our reason set R to include a new symbol σ not already present and then add the following additional schemes to RBB:

$$\begin{array}{ll} \textbf{(MA)} & \sigma \to (Br \to r) \\ \textbf{(MB)} & B\sigma \\ \textbf{(MR)} & B(r : \varphi) \to (Br \to B(\sigma : \varphi)) \end{array}$$

(MA) says that every accepted reason is adequate if the master reason is adequate, (MB) says that the agent always accepts the master reason, and (MR) says that anything the agent believes is supported by an accepted reason the agent will also believe to be supported by the master reason. Adding these principles makes it so that σ is the sum of the agent's evidence. Calling RBB_σ the theory obtained from RBB by expanding the set R of reasons to include a new symbol σ and adding (MA), (MB), and (MR), it follows that

$$\mathsf{RBB}_\sigma \vdash (Bs \wedge Br \wedge B(s : (\varphi \to \psi)) \wedge B(r : \varphi)) \to B\psi \ . \tag{RCL2}$$

Indeed, if the agent accepts s, believes s supports an implication, accepts r, and believes r supports the antecedent of the implication, then it follows by (MR) that the agent believes σ supports the implication and its antecedent. Applying (BRK), it follows that the agent believes σ supports the consequent. But the agent also accepts σ by (MB), so it follows by an application of (BA) that the agent believes the consequent. It is in this sense that the agent "combines" the information conveyed by reasons the agent accepts into the master reason σ.

While we have shown that the RBB_σ-agent may combine the information from two reasons to derive her beliefs, there is no need to restrict the number of reasons to two. Indeed, according to (MB), the agent implicitly combines into σ the information from *every* reason she believes to be adequate, no matter how many of these there may be.

In RBB_σ, the master reason σ serves as a witness for an existential quantifier over the believed support of accepted reasons.[12] In particular, (MR) tells us that if there exists an accepted reason r that the agent believes supports φ, then the agent believes that σ supports φ. Hence by (MB), if there exists such an r, then the agent believes σ supports φ and is itself accepted. If we were to add quantifiers to the language (something we do later in §3.7), we could express this as:

$$(\exists r)(Br \wedge B(r : \varphi)) \to (B\sigma \wedge B(\sigma : \varphi)) \ .$$

[12] Said precisely: σ is a Skolem constant for the existential quantifier over the believed support of accepted reasons. See [31] or any book on mathematical logic for details.

That is, if there exists an accepted reason that the agent believes supports φ, then σ is a witness to the existential quantifier.

Both RBB_σ and the basic theory RBB allow for the possibility that the agent believes a formula φ without any supporting reasons (i.e., she does not believe to be adequate any reason that she believes supports φ). This is the same as saying that the set

$$\{B\varphi\} \cup \{B(r : \varphi) \to \neg Br \mid r \in R\}$$

is consistent with both RBB_σ and RBB. If this situation is undesirable, then a simple remedy is to extend the theory RBB_σ by adding a principle that says all beliefs are believed to be supported by the master reason:

$$\textbf{(MT)} \quad B\varphi \to B(\sigma : \varphi)$$

With (MT) in place, the agent believes only those things she believes are supported by σ and hence, by (MR), she believes only those things she believes are supported by some reason. In short, every belief is "reason-based." Defining RBB^+_σ to be the theory obtained from RBB_σ by adding (MT), it follows by (MT), (RB), and (MB) that

$$\mathsf{RBB}^+_\sigma \vdash B\varphi \leftrightarrow B(\sigma : \varphi) \ ,$$

which says that the agent believes something just in case it is supported by the master reason. But then belief can be conflated with the master reason. As a result, we have by (RCLC) that the beliefs of the RBB^+_σ-agent are always closed under (MP). Belief in RBB^+_σ is therefore governed by the normal modal logic KD.

Semantics for RBB_σ and for RBB^+_σ may be found in §A.1. It is shown in Theorem A.2 that each of RBB_σ and RBB^+_σ is sound and complete for its semantics.

3.7 Quantification Over Reasons

We have observed that the theory RBB does not require that every belief arise from a reason: it is consistent with RBB for the agent to believe φ and yet have no accepted reason r she believes supports φ. If we were to introduce quantifiers over reasons into our language, then we could express this situation by saying that the following formula is consistent:

$$B\varphi \wedge (\forall r)(B(r : \varphi) \to \neg Br) \ .$$

Another example: we might like to say that r is the unique accepted reason the agent believes supports φ:

$$B(r : \varphi) \wedge Br \wedge (\forall s)((s \neq r \wedge B(s : \varphi)) \to \neg Bs) \ .$$

To allow such expressions as formulas, we extend our set of formulas F to the larger set $F^\forall$ consisting of all formulas φ that may be formed by the following grammar:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (r : \varphi) \mid r \mid B\varphi \mid r = r \mid (\forall r)\varphi \qquad {\scriptstyle p \in P,\ r \in R}$$

We adopt usual Boolean connective abbreviations along with the following:

$$\begin{aligned}
r \neq s &:= \neg(r = s) \\
(\exists r)\varphi &:= \neg(\forall r)\neg\varphi \\
(\forall r \neq s)\varphi &:= (\forall r)(r \neq s \to \varphi) \\
(\exists r \neq s)\varphi &:= (\exists r)(r \neq s \wedge \varphi)
\end{aligned}$$

Note that in this language, an element $r \in R$ can act both as a reason (as in the formula $r : p$) and as a quantifier variable (as in the formula $(\forall r)(r : p)$). Therefore, reasons may appear either bound or free in formulas, with the notion of *bound* and *free* defined in the usual way. For reasons s and r and a formula φ, we say that *s is free for r in φ* to mean that r has no free occurrence in φ within the scope of a quantifier $(\forall s)$. Put another way, if s is free for r in φ, then in the formula

$$\varphi[s/r]$$

obtained by substituting all free occurrences of r in φ by s, no newly replaced occurrence is bound. Examples: s is free for r in $(\forall t)(t \neq r)$ but not in $(\forall s)(s \neq r)$.

The theory QRBB of *Quantified Reason-Based Belief* is defined by adding to the axiomatization of RBB the axioms and rules in Table 2. (UD), (UI), and (Gen) are standard principles of first-order quantification. (EP) and (EN) say that two reasons are considered to be the same if and only if they are syntactically identical.

Additional Axiom Schemes

(UD)	$(\forall r)(\varphi \to \psi) \to (\varphi \to (\forall r)\psi)$, where r is not free in φ
(UI)	$(\forall r)\varphi \to \varphi[s/r]$, where s is free for r in φ
(EP)	$r = r$
(EN)	$\neg(r = s)$, where r and s are syntactically different

Additional Rule

$$\frac{\varphi}{(\forall r)\varphi}\ \text{(Gen)}$$

Table 2. The theory QRBB consists of RBB augmented with the above axioms and rule

It is shown in Corollary A.11 that for each $\varphi \in F$ not containing quantifiers, we have $\mathsf{QRBB} \vdash \varphi$ if and only if $\mathsf{RBB} \vdash \varphi$. It is therefore unproblematic for us to simply write $\vdash \varphi$ to say that φ is provable.

We can extend QRBB to the theory QRBB_σ obtained by extending R to include a new master reason σ and adding the schemes (MA), (MB), and (MR) for σ. We can further extend QRBB_σ to the theory QRBB^+_σ obtained by adding the additional scheme (MT) to guarantee all beliefs are reason-based.

Semantics for QRBB, for QRBB_σ, and for QRBB^+_σ may be found in §A.2. It is shown in Theorems A.5 and A.12 that each of QRBB, QRBB_σ, and QRBB^+_σ is sound and complete for its semantics. However, for the completeness results, there is one caveat: our proofs require that the set R of reasons be at least countably infinite.

4 Justified True Belief and Knowledge

We use our logical framework to tease apart two notions of justified true belief (henceforth "JTB"). The first is an internalist notion, which Gettier showed was insufficient for knowledge [16]. The second is an externalist notion that we argue is immune to Gettier scenarios. More generally, we show that our framework can distinguish three "types" of reasons: those that are non-veridical (and hence inadequate), those that are veridical for the proposition they support, but inadequate, and those that are adequate (and hence veridical). Gettier's second case has reasons of the first type: non-veridical. The "fake barn county" case has reasons of the second type: veridical for the proposition they support, but inadequate. Other cases (such as "normal barn county", or "good cases" more broadly, see [33]) have reasons of the third type: adequate.

4.1 Two Notions of Justification

In our theory, there are (at least) two natural ways to define JTB:

- $\mathsf{JTB}^e_r(\varphi) := B(r : \varphi) \wedge Br \wedge r$, and
- $\mathsf{JTB}^i_r(\varphi) := B(r : \varphi) \wedge Br \wedge \varphi$.

Both imply that the agent has a true belief that φ. However, $\mathsf{JTB}^e_r(\varphi)$ suggests that the agent has a true belief justified by an adequate reason, whereas $\mathsf{JTB}^i_r(\varphi)$ suggests that the agent only has a true belief justified by a *prima facie* reason (that may not be adequate). JTB^e_r is thus externalist, while JTB^i_r is internalist.

Gettier's achievement was to deny that $\mathsf{JTB}^i_r(\varphi)$ is the same as knowledge of φ. Thus, if we assume that

$$B(r : p) \wedge Br \wedge B(r : (p \to p \vee q)) \wedge (\neg p \wedge q) \ , \tag{G2}$$

then we have Gettier's second case. This is the case where the agent named Smith has a reason to believe p ("Jones owns a Ford") and "realizes" that $p \vee q$ ("Jones owns a Ford or Brown is in Barcelona") follows from p on that basis; however, unknown to Smith, p is false and q is true, hence $p \vee q$ is premised on a false lemma. Let us call r the reason Smith has to believe p. We can safely assume that r does indeed support p in that case $(r : p)$ (i.e., the past evidence adduced by the agent does correspond to a real experience of his) and we leave this premise implicit in (G2) above, since delusion is not the problem in this case. Since p is false, and $r : p$ by assumption, r cannot be adequate, by (A). However, since r supports p, we have by (RCLC) that r supports $p \vee q$. Smith is said by Gettier to realize that p entails $p \vee q$ on the basis of his reason, and by (BRK) it follows that $B(r : (p \vee q))$. Moreover, Smith has no reason supporting q that she believes is adequate (indeed, in the scenario, "Brown is in Barcelona" is consciously picked at random by Smith). So we are in a situation where Smith has an internally justified true belief that p ($\mathsf{JTB}^i_r p$), and also an internally justified true belief that $p \vee q$ ($\mathsf{JTB}^i_r(p \vee q)$), but he fails to have an externally justified true belief of either proposition ($\neg \mathsf{JTB}^e_r p$, and $\neg \mathsf{JTB}^e_r(p \vee q)$).

In contrast, if we assume that $\mathsf{JTB}^e_r(p)$, which is

$$B(r : p) \wedge Br \wedge r \ , \tag{G2$'$}$$

this time r is adequate. Since r supports p, it also supports $p \vee q$ by (RCLC). By (A) both p and $p \vee q$ are true. Since the agent believes r is an adequate reason supporting p (and therefore also supporting $p \vee q$), she believes both p and $p \vee q$, and in this case her belief is based on an adequate (and hence veridical) reason.

In general, it is easy to see that $\mathsf{JTB}^e_r(\varphi)$ satisfies:

- $\vdash \mathsf{JTB}^e_r(\varphi \to \psi) \to (\mathsf{JTB}^e_r(\varphi) \to \mathsf{JTB}^e_r(\psi))$,

  which says that external JTB based on a reason r is closed under implication;

- $\vdash \mathsf{JTB}^e_r(\varphi) \to \varphi$,

  which says that external JTB is veridical; and

- $\vdash \mathsf{JTB}^e_r(\varphi) \to (r : \psi \to \psi)$,

  which says that if an agent has an external JTB based on reason r, then r cannot support any false assertions (so-called "false lemmas").

To compare with internal JTB, one can show that $\mathsf{JTB}^i_r(\varphi)$ satisfies:

- $\vdash \mathsf{JTB}^i_r(\varphi \to \psi) \to (\mathsf{JTB}^i_r(\varphi) \to \mathsf{JTB}^i_r(\psi))$,

  which says that internal JTB based on a reason r is closed under implication;

- $\vdash \mathsf{JTB}^i_r(\varphi) \to \varphi$,

  which says that internal JTB is veridical; and

- $\nvdash \mathsf{JTB}^i_r(\varphi) \to (r : \psi \to \psi)$,

  which says that if an agent has an internal JTB based on reason r, then r might support false assertions (so-called "false lemmas").

The differences between JTB^e_r and JTB^i_r are in the last property. So we see that the main difference between external and internal JTB is in the reliability of the reason on which the JTB is based.

Using our quantified language, we adopt the following abbreviations:

$$\mathsf{JTB}^e(\varphi) := (\exists r)\mathsf{JTB}^e_r(\varphi) \quad \text{and} \quad \mathsf{JTB}^i(\varphi) := (\exists r)\mathsf{JTB}^i_r(\varphi) \ .$$

$\mathsf{JTB}^e(\varphi)$ says that the agent has an external JTB for φ (based on some reason), and $\mathsf{JTB}^i(\varphi)$ says the same but for internal JTB.

4.2 Is Knowledge JTB^e?

JTB^i falls prey to Gettier's examples because the supporting reason need not be veridical (i.e., it admits "false lemmas"). JTB^e, however, requires an adequate supporting reason, and hence this reason is necessarily veridical (i.e., it admits "no false lemmas"). This suggests we examine the equation

$$K\varphi := \mathsf{JTB}^e(\varphi) \ , \tag{KJTBe}$$

which defines knowledge as external JTB. What should we think of this equation?

Consider the "fake barn county" situation [17]: the agent is in a county that has numerous fake barns that look exactly like real barns. Not knowing she is in this county, she sees what she thinks is a barn and concludes that it is indeed a barn. It turns out she is correct because, by chance, she happens to be looking at the only real barn in the entire county. Obviously, she has an internal JTB that she sees a barn, though most philosophers argue that she does not know she sees a barn.[13] But does she have an external JTB in this case? One reason to answer affirmatively: the agent's reason is veridical, unlike in Gettier's original examples.

However, veridicality does not imply adequacy. Take r and p so that

r is read, "the agent sees what looks to her like a barn," and

p is read, "what the agent sees is a barn."

Our agent is in the situation:

$$B(r:p) \wedge Br \wedge p \ . \tag{Barn}$$

That is, the agent believes that her seeing what looks like a barn supports the assertion that what she sees is a barn, she believes r is adequate to guarantee the truth of what it supports, and the agent does actually see a barn. But is r in fact adequate? If we say it is, then we run into the following problem: had the agent picked a different barn-looking structure that turned out to be a fake, we would have

$$B(r:p) \wedge Br \wedge \neg p \ , \tag{Barn$'$}$$

from which it would follow by (AS) and the assumed adequacy of r that p holds (i.e., we have $(r:p) \wedge p$), but this contradicts the assumed adequacy of r (since in fact $\neg p$). This suggests to us that r is not necessarily adequate; that is, each of (Barn) $\wedge\, r$ and (Barn) $\wedge \neg r$ is consistent with our intuitions about the "fake barn county" example. Conclusion: the agent need not have external JTB in this case.

We take it that the "fake barn county" example seeks to challenge the agent's acumen in determining when it is safe to reason according to the principle

$$(\text{what I see looks like an } X) \to (\text{what I see is an } X) \ , \tag{WSWG}$$

which has the colloquial reading "what I see is what I get."[14] Since adequacy implies veridicality, one could use our notion of adequacy to indicate that the agent's use of (WSWG) is licensed. In particular, if we assume that (Barn) $\wedge\, r$, then we might construe this as a case in which the agent is in "normal barn county" (where there are no fake barns) and so her use of (WSWG) is licensed: r is an externally valid reason for the agent to infer that she sees a barn, so the agent knows that she sees a barn. In contrast, if we assume that (Barn) $\wedge \neg r$, then we might construe this as a case in which the agent is back in "fake barn county" and not licensed to draw the conclusion: r is not an externally valid reason for her to infer that she sees a barn, so she does not know that she sees a barn.

So let us assume that our definition (KJTBe) of knowledge as external JTB is correct. Is it then possible to define knowledge (i.e., external JTB) in terms of internal JTB plus some other condition? Indeed it is:

$$\vdash \mathsf{JTB}^e(\varphi) \leftrightarrow (\exists r)(r \wedge \mathsf{JTB}^i_r(\varphi))$$

In words: to have external justification it is necessary and sufficient to have an adequate justification that serves as the basis for an internal JTB. Zagzebski's criticism of a JTB-based analysis of

[13] Intuitions about knowledge ascription in fake barn cases are notoriously less stable among philosophers than they are in the original Gettier cases (see [21, 29, 13]). Here we are considering a situation in which the belief seems simply "too lucky" to count as knowledge.

[14] This diagnosis was suggested to the third author by Alexandru Baltag (private communication).

knowledge [34] might apply here: we either must sever the link between truth and justification (thereby going so far as to concede that there is knowledge in Gettier cases) or else assert that "there is no degree of independence at all between truth and justification" (in order to avoid Gettier problems). Zagzebski's position is that neither horn of her dilemma is satisfactory, and so the proper way to avoid the dilemma is to reject the possibility of analyzing knowledge in terms of JTB plus some extra component (i.e., reject the "knowledge is JTB $+ x$" approach all together); see also [33]. We accept that our approach is close to endorsing the second horn of Zagzebski's dilemma. However, by distinguishing adequacy from veridicality, we can still maintain a notion of independence between truth and justification. In particular, *pace* Zagzebski, our semantic analysis distinguishes between "adequate belief" (i.e., JTB^e) and "lucky true belief" (i.e., JTB^i).

5 Infallibilism and the Problem of Mixed Reasons

Our analysis commits us to an infallibilist view of knowledge. Dutant in [11, 12] defines *method infallibilism* as for an agent to know that p if the agent believes that p on the basis of a method that could only yield true beliefs. Our notion of knowledge in terms of external JTB achieves the same result: for an agent to know p is to believe p on the basis of an adequate reason r, hence to believe p on the basis of a reason that could only support true propositions.[15] In this section we propose to discuss two specific issues concerning our definition of knowledge in terms of JTB^e. Both issues raise the problem of whether a definition of knowledge as JTB^e might be either too weak, or too strong, depending on the case.

5.1 Knowledge from mixed reasons?

Let us start with the worry that our account might be too weak. The JTB^e analysis of knowledge raises the issue of the force of the quantifier on the right side of the equivalence. To appreciate the problem, it is worth reminding ourselves of one of the first responses to Gettier's examples: the so-called "no false lemmas" (hereafter "NFL") requirement (see [9, 19, 26]). The NFL requirement is meant to rule out situations like Gettier's second case, wherein the agent starts from a mistaken belief that p to obtain a correct belief that $p \vee q$. Thinking of the reasoning sequence of beliefs $\langle p, p \vee q\rangle$ as a "proof," the initial "lemma" (i.e., assumption) p is false, but then a perfectly legitimate inference step to a logical consequence $p \vee q$ ends up on a formula that just so happens to be true.

In our framework, the obvious counterpart to the NFL requirement is the "no inadequate lemmas" (henceforth "NIL") requirement:

$$\mathsf{NIL}(\varphi) := (\forall s)(\mathsf{JTB}^i_s(\varphi) \to s) \ .$$

This says that every reason that supports an internal JTB of φ is adequate. Since adequate reasons support only true formulas (by axiom scheme (A)), the NIL requirement guarantees that no false "lemma" (i.e., formula) intrudes on a reason justifying a potential internal JTB of φ. This gives rise to the following notion of JTB with no inadequate lemmas:

$$\mathsf{JTB{+}NIL}(\varphi) := \mathsf{JTB}^i(\varphi) \wedge \mathsf{NIL}(\varphi) \ .$$

[15] See Neta's [24], and Dutant's [11, 12] for an in-depth discussion of infallibilism. We leave a comparison of our approach with Neta's and Dutant's respective approaches for another occasion.

This notion of JTB is logically stronger than external JTB: $\mathsf{JTB}^e(\varphi)$ only requires that there be one adequate reason supporting an internal JTB of φ, whereas $\mathsf{JTB}{+}\mathsf{NIL}$ requires adequacy of *every* reason supporting an internal JTB of φ. Thus

$$\vdash \mathsf{JTB}{+}\mathsf{NIL}(\varphi) \to \mathsf{JTB}^e(\varphi)$$

but not the other way around.

These considerations raise a potential worry for the JTB^e analysis of knowledge: what happens when the agent rests her beliefs in a proposition (such as $p \vee q$) on multiple sources? For example,[16] suppose our agent, who has excellent eyesight, sees someone in the distance but cannot quite make out who it is. Nevertheless, based on what she can see (represented by reason s), she correctly believes that the person in the distance is either Tweedle Dee or Tweedle Dum (represented respectively by $p \vee q$). Further, she has another reason r to believe that the person is Tweedle Dee (i.e., p). For example, a friend might have told her that Tweedle Dum is on vacation in some faraway country. Now, unknown to our agent, the person in the distance is actually Tweedle Dum. Put formally:

$$\begin{aligned} B(s:(p \vee q)) \wedge \neg B(s:p) \wedge \neg B(s:q) \wedge B(r:p) \ \wedge& \\ Bs \wedge Br \wedge s \wedge \neg r \wedge (\neg p \wedge q)\ .& \end{aligned} \tag{TDTD}$$

That is, the agent believes s supports the disjunction (that the person is Dee or Dum) but no disjunct, she believes r supports the claim it is Dee, she believes s and r to be adequate, s is adequate (by hypothesis, because the agent's eyesight is excellent, and it could not possibly be someone other than Dee or Dum), r is inadequate (since the friend's information is not reliable), and the person is actually not Dee but Dum. Now suppose we add to (TDTD) the assumption

$$(\forall t)((t \neq s \wedge t \neq r) \to \neg Bt) \tag{NoR}$$

that the agent believes no other reasons to be adequate. It can be shown that

$$\begin{aligned} &\vdash [(\text{TDTD}) \wedge (\text{NoR})] \to \mathsf{JTB}^e(p \vee q) \qquad \text{but} \\ &\nvdash [(\text{TDTD}) \wedge (\text{NoR})] \to \mathsf{JTB}{+}\mathsf{NIL}(p \vee q)\ . \end{aligned}$$

In words: the agent has external JTB that $p \vee q$ (because the adequate reason s supports the disjunction); however, she does not have JTB with NIL of $p \vee q$ (because the inadequate reason r supports the disjunction). But is it a mistake to equate knowledge with JTB^e instead of with $\mathsf{JTB}{+}\mathsf{NIL}$?

One reaction is to deny that there is knowledge when the universal condition is not satisfied. For an example supporting this reaction, suppose the agent proves that a certain Mersenne number $m = 2^n - 1$ is prime. Later, she bolsters her belief in the primality of m by coming to believe (incorrectly) that all Mersenne numbers are prime (i.e., all numbers of the form $2^k - 1$ are prime, which is false). Can the agent still be said to know that m is prime? On one account, it seems not. Such situations of *mixed reasons*, where an agent has both adequate and inadequate reasons supporting the same proposition, arguably occur often in everyday life.

[16] The example was suggested to the first author by Timothy Williamson (private communication). See [13].

We are inclined to the opposite view: in a situation of mixed reasons, the agent can still have knowledge. Returning to the primality example, if the agent learns that not all Mersenne numbers are prime, then she will still believe that m is prime on the basis of her adequate "backup" reason (that she proved m is prime). So she could still be said to know that m is prime.[17] The Dee/Dum case is arguably similar: if the agent were to learn that r is unreliable, then she would still have an external JTB of the disjunction based on the "backup" reason s.

Perhaps the most difficult challenge to the claim that (KJTBe) is correct even in the case of mixed reasons comes when the quantity of inadequate reasons vastly exceeds the quantity of adequate reasons. For example, suppose our agent has an adequate reason s (based on an assertion in some recent official document) that one of the 20 members of the faculty of department D is a logician; further, suppose she has inadequate reasons $r_1, \ldots, r_{19}$ (based on a mistaken understanding of which specialties imply competence in logic) that the first 19 names listed on the department D faculty roster are logicians. We might be hard-pressed to say that our agent knows that department D has a logician on staff.

Perhaps this suggests that the agent in a case of mixed reasons can only be said to know the proposition if she also knows that her reasons are adequate. We resist this move, basically because we accept that an agent may have an adequate reason without necessarily knowing that that reason is adequate (more on this below in the conclusion). Therefore, if we assume for the sake of argument that our agent values all reasons equally, then a tiny island of adequacy within an ocean of inadequacy is sufficient for the defender of mixed-reason knowledge. This grants that the agent's reasons are in some sense confused or that an agent who has only adequate reasons (and hence satisfies $\mathsf{JTB}{+}\mathsf{NIL}$) seems to "know better" than the agent with mixed reasons. But if one agent "knows better," it does not follow that the other does not know at all.

In our view, an account of knowledge that would not allow for mixed reasons would in fact be too demanding. We consider it a virtue of our account that it allows for mixed reasons, precisely because we think it gives us a more realistic picture of the process of acquiring and managing reasons. We say that an agent knows a proposition if he believes that proposition based on at least one adequate reason. But knowing is more than passively believing propositions on the basis of reasons. It obviously also involves comparing and relating reasons. Consider an agent who believes the true proposition p on the basis of r and r', but who eventually realizes that r' supports a false proposition. The agent ought to revise her belief in r', and also to reconsider her reasons for p. Hence, while our account of knowledge commits us to what Dutant calls method infallibilism, it does make room for errors and revisions in how reasons are acquired. It contains, in that sense, a measure of fallibilism.

5.2 Knowledge from inadequate reasons?

Although our account of knowledge allows for mixed reasons, a converse objection is that it may turn out to be too strong relative to ordinary knowledge ascriptions. The problem in this case is even more radical, and concerns whether one can have knowledge on the basis of a single, inadequate reason.

Here is an example:[18] based on his seeing Jones driving a Ford in the past (let us call that reason r), Smith comes to wrongly believe that Jones owns a Ford. Let us modify the scenario and

[17] These ideas are related to the defeasibility theory of knowledge [22, 23].

[18] We are indebted to Elia Zardini for raising the objection, and for the first example.

suppose that Jones does in fact own a car (say, a Mazda). Based on his seeing Jones drive a Ford in the past, Smith also comes to believe that Jones owns a car. Could it not happen, intuitively, that although Smith fails to have knowledge on the basis of r that Jones owns a Ford (p), he nevertheless has knowledge on the basis of r that Smith owns a car (q)? In our system, this is not possible, for by (A), if r is adequate for q, then r must be adequate for any other proposition that r supports, hence for p as well. Dretske's treatment of conclusive reasons would be able to address this problem: a reason r can be conclusive for q without being conclusive for p, even when p entails q. Our approach does not have this flexibility.[19]

The question, more generally, is whether the same reason can adequately justify one to believe q without adequately justifying one to believe a stronger proposition p. Similar cases have been discussed by Warfield [32], Fitelson [14], and Sorensen [25]. Warfield argues that I may know that I am not late for the meeting if I believe that it is currently 2:58pm, when in fact it is 2:56pm, assuming the meeting is at 7pm. On our account, my reason to believe that it is currently less than 7pm is inadequate, simply because it also supports the false proposition that it is 2:58pm. This is a case in which I have JTB^i that it is less than 7pm, without having JTB^e that it is less than 7pm. For anyone whose intuition is that I do hold knowledge that it is is less than 7pm on the basis of my observing "2:58pm" on the watch, our equation between knowledge and JTB^e is too strong in this case.

One option in the face of such examples is to bite the bullet and to resist the intuition that I know I am not late for the meeting, or that I know that Jones owns a car. But we think this is not the right response. My evidence "2:58pm" is obviously wrong regarding the actual time, but still close enough to the actual time to be relevantly used. The case would be different, it seems to us, if the agent's watch indicated 6pm when it is 2:56pm, or even 9am. For the latter cases, our intuition is that I merely have a luckily true belief. More generally, we think the problem concerns how much approximation is tolerated in forming beliefs based on one's evidence. If, when I see "2.58pm" (r) on my watch, I form the belief "it is around 2:58pm" (p), and from that proposition I infer "it is less than 7pm" (q), then my reason r now is veridical for both p and q. A way out, therefore, might be to relativize the adequacy of a reason to the selection of an appropriate domain of propositions supported by that reason.

This nevertheless puts pressure on us to clarify the relation of support between a reason and a proposition. In our statement of the axiom (A), we include no restriction on the support relation. We think it is better to be normative, and not to include any such restriction in the definition of knowledge in terms of JTB^e. On the other hand, we are ready to accept that in actual ascriptions of knowledge, the definition of adequate evidence is relativized to various domains of propositions. Consider the Warfield example again: this is not a perfect case of knowledge. This still counts as evidence that comes close to adequate, though not perfectly adequate. But it is adequate given the relevant domain of propositions. Our account, therefore, does not rule out the familiar mechanisms of contextualization at play in ordinary knowledge ascriptions, despite being fundamentally more normative.

[19] One option we do not explore here: change adequacy from a unary property on reasons to a binary property on reasons and propositions. Thus instead of having "r" for adequacy of reason r with respect to all propositions it supports, we would have "$A(r,p)$" for adequacy of r with respect to proposition p. We would have to adjust other axioms: (A) would be $r:\varphi \to (A(r,\varphi) \to \varphi)$, (BA) would be $B(r:\varphi) \to (B(A(r,\varphi)) \to B\varphi)$, and (AS) would be $B(r:\varphi) \to (A(r,\varphi) \to (r:\varphi))$. This would also require a change to the semantics that would make us much more in-line with the syntactic dependencies connecting reasons with specific formulas, which is familiar from Justification Logic [3]. This approach would add flexibility at the cost of simplicity.

6 Conclusion

Is knowledge the same thing as JTB? We wrote this paper based on a persistent feeling that both answers are defensible. For the negative: Gettier's examples show that plausible reasons may be inadequate. For the positive: a JTB based on an adequate reason seems to rule out the possibility of Gettier cases and can arguably be construed as a form of knowledge.

We have shown that our framework is sufficient to address reason-based belief and that it can be applied to important notions in epistemology. However, we have neglected to provide a further analysis of "adequacy of a reason" into more primitive concepts. While this notion was used as a primitive in this paper, an in-depth study of this notion may be required in a full philosophical analysis of the concept of knowledge. Regardless, we think that our three-part hierarchy of reasons (non-veridical, veridical for a proposition but inadequate, and adequate) is itself sufficiently fruitful to legitimate our approach. Our account, as we have explained, fundamentally commits us to a form of infallibilism in the definition of knowledge. But our treatment of mixed reasons also makes room for the possibility of errors, since inadequate reasons typically coexist with adequate reasons. We can therefore distinguish two levels in our account of knowledge: the level of atomic reasons (and of their support to various propositions), and the level of the network of reasons that an agent needs to compare and manage. A discussion of that second level lies beyond the scope of this paper, but it deserves to be considered, because our externalist account of the notion of adequacy remains compatible with a more internalist perspective on knowledge.

A related question on which we propose to end is the following: how does an agent know whether a reason is adequate? According to (KJTBe), the agent knows p if and only if there exists an adequate reason r that the agent believes is adequate and supports p. Therefore, the agent knows r is adequate if and only if there exists a reason s that she believes is adequate and supports r (i.e., $s \wedge Bs \wedge B(s:r)$).[20] Our framework therefore admits the possibility that the agent may know p based on an adequate reason r without knowing that r is itself adequate. In this, our framework supports the main contention of an externalist account of knowledge: one may know p without knowing that one knows p (see [10, 33]).[21] We think this is right for the externalist, though we emphasize that our theory is in principle neutral regarding the existence of reasons justifying the adequacy of other reasons.

Acknowledgments

The authors thank two anonymous reviewers of the RSL for detailed and helpful comments, as well as Wes Holliday and Sean Walsh for their editorial work. They also thank the following people for comments and discussion: Amir Anvari, Igor Douven, Julien Dutant, Rohan French, Chenwei Shi, Meghdad Ghari, Andreas Herzig, Clayton Littlejohn, Anne Meylan, Jennifer Nagel, Ram Neta, Jean-Baptiste Rauzy, Roy Sorensen, Daniele Sgaravatti, Benjamin Spector, Giuseppe Spolaore, Sylvia Wenmackers, Timothy Williamson, and Elia Zardini.

[20] Note that we may have $s = r$. In particular, it is consistent with our theory for reasons to be self-supporting (i.e., $r : r$). It is also consistent with our theory for reasons to be non–self-supporting (i.e., $\neg(r : r)$). Since our theory permits either option, it is up to the user of our theory to choose which way to go as per her preference. We also note that it is consistent for r to be self-rejecting (i.e., $r:\neg r$), and it is consistent for r to be non–self-rejecting (i.e., $\neg(r:\neg r)$).

[21] The internalist who objects to this need not despair: though we do not do so here, it is possible to extend our framework so that knowledge is internalizable; see [3] (and the "proof checker" operator) for details.

We also thank Joelle Proust, Igor Douven, Fabien Mikol, and Giuliano Torrengo who organized the respective events where our paper was presented (the IJN seminar on epistemic norms, the SND epistemology colloquium, the Paris IV workshop on the Limits of Knowing, and the Latin Meeting in Analytic Philosophy in Milan). We also thank the various audience members present at these events for their questions.

A Technical Results

A.1 Semantics for RBB_σ and RBB_σ^+

The models for RBB can be construed as models for RBB_σ if we require the following additional properties:

(ma) $w \in \sigma(w)$ and $r^\circ \in N(w)$ together imply that $w \in r(w)$.

which says that if σ is reflexive, then each reason r believed to be reflexive is in fact reflexive;

(mb) $\sigma^\circ \in N(w)$,

which says that the agent believes σ is reflexive; and

(mr) $[\![r : \varphi]\!] \in N(w)$ and $r^\circ \in N(w)$ together imply that $[\![\sigma : \varphi]\!] \in N(w)$,

which says that the agent believes σ supports φ whenever she believes r supports φ and she believes r is reflexive.

We write the satisfaction relation $\models_\sigma$ to indicate that we restrict to models satisfying (ma), (mb), and (mr). By Theorem A.2, RBB_σ is sound and complete for the class of models satisfying (ma), (mb), and (mr). Models for the theory RBB_σ^+ must satisfy (ma), (mb), (mr), and the following property:

(mt) $[\![\varphi]\!] \in N(w)$ implies $[\![\sigma : \varphi]\!] \in N(w)$,

which says that the agent believes σ supports φ whenever she believes φ.

We write $\models_\sigma^+$ to indicate that we restrict to models satisfying (ma), (mb), (mr), and (mt). By Theorem A.2, RBB_σ^+ is sound and complete for the class of models satisfying (ma), (mb), (mr), and (mt).

A.2 Semantics for QRBB, QRBB_σ, and QRBB_σ^+

The models for RBB can be used as models for QRBB as well. All that we must do is add the following satisfaction principles:

- $M, w \models r = s$ means that $r = s$.
- $M, w \models (\forall r)\varphi$ means that $M, w \models \varphi[s/r]$ for each s free for r in φ.

It is shown in Theorem A.5 that QRBB is sound and complete for this semantics: for each theory we have $\mathsf{QRBB} \vdash \varphi$ if and only if $\models \varphi$. However, there is one caveat: our proof of the completeness result requires that the set R of reasons be at least countably infinite.

Additional semantic restrictions must be imposed to ensure that QRBB models also work for QRBB_σ or for QRBB_σ^+; see §A.1 for details. Soundness and completeness for QRBB_σ and for QRBB_σ^+ follows by Theorem A.12, with the same caveat for completeness as for QRBB.

A.3 RBB Soundness and Completeness

We now prove the following theorem.

Theorem A.1 (RBB Soundness and Completeness)**.** For each $\varphi \in F$:

$$\mathsf{RBB} \vdash \varphi \quad \text{iff} \quad \models \varphi \ .$$

Soundness is by induction on the length of derivation. The arguments for (CL), (MP), (RK), (D), (RN), and (E) are straighforward. We consider the remaining cases.

- Validity of (A): $\models r\!:\!\varphi \to (r \to \varphi)$.

  $M, w \models r\!:\!\varphi$ and $M, w \models r$ together imply that $r(w) \subseteq [\![\varphi]\!]_M$ and $w \in r(w)$. But then $M, w \models \varphi$.

- Validity of (BRK): $\models B(r\!:\!(\varphi \to \psi)) \to (B(r\!:\!\varphi) \to B(r\!:\!\psi))$.

  Suppose $M, w \models B(r\!:\!(\varphi \to \psi))$ and $M, w \models B(r\!:\!\varphi)$. This means

  $$[\![r\!:\!(\varphi \to \psi)]\!]_M \in N(w) \text{ and } [\![r\!:\!\varphi]\!]_M \in N(w) \ .$$

  Applying (brk), it follows that $[\![r\!:\!\psi]\!]_M \in N(w)$, which is what it means to have $M, w \models r\!:\!\psi$.

- Validity of (BA): $\models B(r\!:\!\varphi) \to (Br \to B\varphi)$.

  Assume $M, w \models B(r\!:\!\varphi)$ and $M, w \models Br$. It follows that $[\![r\!:\!\varphi]\!]_M \in N(w)$ and $r^\circ \in N(w)$. Applying (ba), we obtain $[\![\varphi]\!]_M \in N(w)$. But this is what it means to have $M, w \models B\varphi$.

- Validity of (AS): $\models B(r\!:\!\varphi) \to (r \to (r\!:\!\varphi))$.

  Assume $M, w \models B(r\!:\!\varphi)$ and $w \in r(w)$. Hence $[\![r\!:\!\varphi]\!]_M \in N(w)$ and $w \in r(w)$. Applying (as), we obtain $r(w) \subseteq [\![\varphi]\!]_M$. That is, $M, w \models r\!:\!\varphi$.

So RBB is sound.

For completeness, we prove that $\mathsf{RBB} \nvdash \theta$ implies there exists a pointed model (M_c, Γ_1^θ) satisfying $M_c, \Gamma_1^\theta \not\models \theta$. We use a canonical model construction to build the model $M_c = (W, [\cdot], N, V)$ as follows. First, to say that a set S of formulas is *consistent* means that for no finite $S' \subseteq S$ do we have $\mathsf{RBB} \vdash (\bigwedge S') \to \bot$, where $\bot$ is a fixed contradiction such as $p \wedge \neg p$. To say a set of formulas is *maximal consistent* means that it is consistent and adding any formula not already present will result in a set that is inconsistent (i.e., not consistent). Let M bet the set of all maximal consistent sets of formulas. By a standard Lindenbaum construction, it follows that $\{\neg\theta\}$ can be extended to some $\Gamma^\theta \in M$ and therefore M is not empty. We define $W := M \times \{1, 2\}$ and will write $(\Gamma, i) \in W$ in the abbreviated form Γ_i. Since M is nonempty, W is nonempty. For each reason $r \in R$ and $\Gamma \in M$, define the set

$$\Gamma^r := \{\varphi \in F \mid r\!:\!\varphi \in \Gamma\}$$

of r-supported formulas in Γ. We then define $[r]$ by setting

$$r(\Gamma_i) := \{\Delta_j \in W \mid \Gamma^r \subseteq \Delta \;\&\; (\neg r \in \Gamma \Rightarrow \Delta_j \neq \Gamma_i)\} \ .$$

This way, a world Δ_j is r-accessible from Γ_i iff Δ_j contains all formulas φ that are r-supported at Γ_i (as per membership of $r\!:\!\varphi$ in Γ), unless of course $\Delta_j = \Gamma_i$ and reflexivity is forbidden by $\neg r \in \Gamma$. For each formula $\varphi \in F$, define

$$W(\varphi) := \{\Gamma_i \in W \mid \varphi \in \Gamma\}$$

to be the set of worlds defined by the formula φ. Then let

$$N^+ := \{X \subseteq W \mid \forall \varphi \in F : X \neq W(\varphi)\}$$

be the set of worlds not definable by any formula. For each $\Gamma_i \in W$, we define

$$N^+(\Gamma_i) := \{X \in N^+ \mid \exists Br \in \Gamma : r(\Gamma_i) \subseteq X \;\&\; \forall \theta \in \Gamma^r (B(r\!:\!\theta) \in \Gamma)\} \ .$$

Intuitively, this is the set of non–formula-definable neighborhoods X for which there is a reason the agent accepts that supports X and the agent believes this reason supports all the formulas it actually does support. The neighborhood function N is then defined by

$$N(\Gamma_i) := \{X \subseteq W \mid \exists B\varphi \in \Gamma : X = W(\varphi)\} \cup N^+(\Gamma_i) \ .$$

Therefore, an agent believes a neighborhood X iff the agent believes some formula φ that defines X or, if X is non–formula-definable, there is an accepted reason supporting X and the agent believes the reason supports all the formulas it actually does support. Finally, we define the valuation by

$$V(\Gamma_i) := \{p \in P \mid p \in \Gamma\} \ .$$

This defines M_c. To check that M_c is indeed a pre-model, we must verify that M_c satisfies the property (pr). We prove this now.

- (pr): if $x \in P \cap R$, then $x \in V(\Gamma_i)$ if and only if $\Gamma_i \in x(\Gamma_i)$.

  So assume $x \in V(\Gamma_i)$. This means $x \in \Gamma$. But then we have $\neg x \notin \Gamma$ by the consistency of Γ. Further, since $x \in \Gamma$, it follows by (A) and the maximal consistency of Γ that $\Gamma^x \subseteq \Gamma$. But $\neg x \notin \Gamma$ and $\Gamma^x \subseteq \Gamma$ together imply $\Gamma_i \in x(\Gamma_i)$, which completes the argument for this direction.

  For the converse direction, if $\Gamma_i \in x(\Gamma_i)$, then it follows by the definition of $x(\Gamma_i)$ that $\neg x \notin \Gamma$. So $x \in \Gamma$ by the maximal consistency of Γ. But then we have $x \in V(\Gamma_i)$ by the definition of V.

So M_c is indeed a pre-model.

We prove the *Truth Lemma*: for each formula $\varphi \in F$ and world $\Gamma_i \in W$, we have $\varphi \in \Gamma$ iff $M_c, \Gamma_i \models \varphi$. The proof is by induction on the construction of formulas, and the arguments for the base and Boolean inductive step cases are standard, so we only consider the remaining non-Boolean inductive step cases.

- Inductive step: $r \in \Gamma$ iff $M_c, \Gamma_i \models r$.

  If $r \in \Gamma$, then it follows by (A) and maximal consistency that $\Gamma^r \subseteq \Gamma$ and therefore that $\Gamma_i \in r(\Gamma_i)$. But this is what it means to have $M_c, \Gamma_i \models r$.

  Conversely, if $M_c, \Gamma_i \models r$, then we have $\Gamma_i \in r(\Gamma_i)$. By the definition of $N(\Gamma_i)$, we have $\neg r \notin \Gamma$ and therefore $r \in \Gamma$ by maximal consistency.

- Inductive step: $r:\varphi \in \Gamma$ iff $M_c, \Gamma_i \models r:\varphi$.

  If $r:\varphi \in \Gamma$, then we have $r(\Gamma_i) \subseteq W(\varphi)$. By the induction hypothesis, $r(\Gamma_i) \subseteq [\![\varphi]\!]_{M_c}$. But this is what it means to have $M_c, w \models r:\varphi$.

  Conversely, if $M_c, w \models r:\varphi$, then we have $r(\Gamma_i) \subseteq [\![\varphi]\!]_{M_c}$ and hence $r(\Gamma_i) \subseteq W(\varphi)$ by the induction hypothesis. Toward a contradiction, assume $\neg r:\varphi \in \Gamma$. It follows that $\Gamma^r \cup \{\neg\varphi\}$ is consistent, for otherwise there would exist a finite $S \subseteq \Gamma^r$ such that $\vdash (\bigwedge S) \to \varphi$, hence $\vdash (\bigwedge_{\chi\in S} r:\chi) \to r:\varphi$ by modal reasoning, and therefore $r:\varphi \in \Gamma$, which is impossible because it would follow by the assumption $\neg r:\varphi \in \Gamma$ that Γ is inconsistent. So we may extend the consistent set $\Gamma^r \cup \{\neg\varphi\}$ to some $\Delta \in M$. Taking $j \neq i$, it follows that $\Delta_j \notin W(\varphi)$ and $\Delta_j \in r(\Gamma_i)$, which contradicts $r(\Gamma_i) \subseteq W(\varphi)$. So our assumption that $\neg r:\varphi \in \Gamma$ is incorrect; what we actually have is that $\neg r:\varphi \notin \Gamma$ and therefore that $r:\varphi \in \Gamma$ by maximal consistency.

- Inductive step: $B\varphi \in \Gamma$ iff $M_c, \Gamma_i \models B\varphi$.

  If $B\varphi \in \Gamma$, then it follows by the definition of $N(\Gamma_i)$ that $W(\varphi) \in N(\Gamma_i)$. By the induction hypothesis, $W(\varphi) = [\![\varphi]\!]_{M_c}$, and hence $[\![\varphi]\!]_{M_c} \in N(\Gamma_i)$. But this is what it means to have $M_c, \Gamma_i \models B\varphi$.

  Conversely, if $M_c, w \models B\varphi$, then we have $[\![\varphi]\!]_{M_c} \in N(\Gamma_i)$ and therefore that $W(\varphi) \in N(\Gamma_i)$ by the induction hypothesis. Since $W(\varphi) \notin N^+$, it follows that there exists $B\psi \in \Gamma$ such that $W(\psi) = W(\varphi)$. From this we have that $\vdash \psi \leftrightarrow \varphi$, for otherwise $\{\psi, \neg\varphi\}$ could be extended to $\Delta \in M$ satisfying $\Delta_i \in W(\psi) - W(\varphi)$ or $\{\neg\psi, \varphi\}$ could be extended to $\Omega \in M$ satisfying $\Omega_i \in W(\varphi) - W(\psi)$, but both contradict $W(\varphi) = W(\psi)$. Applying (E), we have $\vdash B\psi \leftrightarrow B\varphi$ and therefore it follows by maximal consistency that $B\varphi \in \Gamma$.

This completes the proof of the Truth Lemma.

We prove the following *Consistency Lemma*: for each $r \in R$ and $\Gamma_i \in W$, if $Br \in \Gamma$ and $\forall\theta \in \Gamma^r(B(r:\theta) \in \Gamma)$, then Γ^r is consistent. Proceeding, assume $Br \in \Gamma$ and $\forall\theta \in \Gamma^r(B(r:\theta) \in \Gamma)$. Since $Br \in \Gamma$, we have $\neg B\neg r \in \Gamma$ by (D) and maximal consistency. Toward a contradiction, suppose Γ^r is not consistent. Then there exists a finite $S \subseteq \Gamma^r$ such that $\vdash (\bigwedge S) \to \bot$. Hence $\vdash (\bigwedge_{\chi\in S} r:\chi) \to r:\bot$ by modal reasoning. Applying maximal consistency and the fact that $S \subseteq \Gamma^r$, we obtain $r:\bot \in \Gamma$. By maximal consistency and the fact that $\vdash r:\bot \to r:\varphi$ for any φ, we obtain $r:\neg r \in \Gamma$ and hence $\neg r \in \Gamma^r$. Applying the assumption $\forall\theta \in \Gamma^r(B(r:\theta) \in \Gamma)$, it follows that $B(r:\neg r) \in \Gamma$. Applying this and the assumption that $Br \in \Gamma$, it follows by (BA) that $B\neg r \in \Gamma$. Since $\neg B\neg r \in \Gamma$, it follows that the maximal consistent set Γ is not consistent, a contradiction. Conclusion: Γ^r is consistent. This completes the proof of the Consistency Lemma.

We now prove that M_c is a model (and not just a pre-model); that is, we prove that M_c satisfies the properties (brk), (ba), (as), and (d).

- M_c satisfies (brk): if $[\![r:(\varphi \to \psi)]\!]_{M_c} \in N(\Gamma_i)$ and $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$, then $[\![r:\psi]\!]_{M_c} \in N(\Gamma_i)$.

  Assume $[\![r:(\varphi \to \psi)]\!]_{M_c} \in N(\Gamma_i)$ and $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$. By the Truth Lemma, it follows that $W(r:(\varphi \to \psi)) \in N(\Gamma_i)$ and $W(r:\varphi) \in N(\Gamma_i)$. Since neither $W(r:(\varphi \to \psi))$ nor $W(r:\varphi)$ is a member of N^+, it follows by the definition of $N(\Gamma_i)$ that there exists $B\theta_1 \in \Gamma$ and $B\theta_2 \in \Gamma$ such that $W(\theta_1) = W(r:(\varphi \to \psi))$ and $W(\theta_2) = W(r:\varphi)$. As in the proof of the last inductive step of the Truth Lemma, it follows using (E) that $\vdash B\theta_1 \leftrightarrow B(r:(\varphi \to \psi))$ and that $\vdash B\theta_2 \leftrightarrow B(r:\varphi)$. Hence we have that $B(r:(\varphi \to \psi)) \in \Gamma$ and $B(r:\varphi) \in \Gamma$ by maximal consistency. Applying (BRK) and maximal consistency, we obtain $B(r:\psi) \in \Gamma$. Applying the definition of $N(\Gamma_i)$, it follows that $W(r:\psi) \in N(\Gamma_i)$. Since we have that $W(r:\psi) = [\![r:\psi]\!]_{M_c}$ by the Truth Lemma, we conclude that $[\![r:\psi]\!]_{M_c} \in N(\Gamma_i)$.

- M_c satisfies (ba): if $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ and $r^\circ \in N(\Gamma_i)$, then $[\![\varphi]\!]_{M_c} \in N(\Gamma_i)$.

  Assume $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ and $r^\circ \in N(\Gamma_i)$. Similar to our argument in the first part of the above proof for (brk), it follows from $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ that $B(r:\varphi) \in \Gamma$. Since we have by the definition of r° that $r^\circ = W(r)$, it follows from $r^\circ \in N(\Gamma_i)$ that $W(r) \in N(\Gamma_i)$. Again by an argument similar to the first part of the above proof for (brk), we obtain from $W(r) \in N(\Gamma_i)$ that $Br \in \Gamma$. But then it follows from $B(r:\varphi) \in \Gamma$ and $Br \in \Gamma$ by (BA) and maximal consistency that $B\varphi \in \Gamma$. Applying the definition of $N(\Gamma_i)$, we obtain $W(\varphi) \in N(\Gamma_i)$. Applying the Truth Lemma, it follows that $[\![\varphi]\!]_{M_c} \in N(\Gamma_i)$.

- M_c satisfies (as): if $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ and $\Gamma_i \in r(\Gamma_i)$, then $r(\Gamma_i) \subseteq [\![\varphi]\!]_{M_c}$.

  Assume $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ and $\Gamma_i \in r(\Gamma_i)$. Similar to our argument in the first part of the above proof for (brk), it follows from $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ that $B(r:\varphi) \in \Gamma$. Now from $\Gamma_i \in r(\Gamma_i)$, it follows by the definition of $N(\Gamma_i)$ that

  $$\neg r \in \Gamma \Rightarrow \Gamma_i \neq \Gamma_i \ .$$

  Hence $\neg r \notin \Gamma$. But then it follows by maximal consistency that $r \in \Gamma$.

  So from $B(r:\varphi) \in \Gamma$ and $r \in \Gamma$ we obtain by (AS) and maximal consistency that $r:\varphi \in \Gamma$. Hence $\varphi \in \Gamma^r$. To complete the argument, we wish to show that $r(\Gamma_i) \subseteq [\![\varphi]\!]_{M_c}$. So let us take an arbitrary $\Delta_j \in r(\Gamma_i)$ and prove that $\Delta_j \in [\![\varphi]\!]_{M_c}$. By the Truth Lemma, it suffices to prove that $\Delta_j \in W(\varphi)$. Proceeding, since $\Delta_j \in r(\Gamma_i)$, it follows by the definition of $r(\Gamma_i)$ that $\Gamma^r \subseteq \Delta$. But then we have by $\varphi \in \Gamma^r$ that $\varphi \in \Delta$ and hence $\Delta_j \in W(\varphi)$.

- M_c satisfies (d): if $X \in N(\Gamma_i)$, then $X' := W - X \notin N(\Gamma_i)$. There are two cases to consider.

  First case for (d): assume $X \in N(\Gamma_i) - N^+(\Gamma_i)$. It follows that there exists $B\varphi \in \Gamma$ such that $X = W(\varphi)$. By (D) and the maximal consistency of Γ, we have $\neg B\neg\varphi \in \Gamma$. Toward a contradiction, assume $X' \in N(\Gamma_i)$. Since $X = W(\varphi)$, we have $X' = W(\neg\varphi)$ by maximal consistency and therefore that $X' \notin N^+$. Hence $X' \in N(\Gamma_i) - N^+(\Gamma_i)$, which means there exists $B\psi \in \Gamma$ such that $X' = W(\psi)$. It follows that $\vdash \psi \leftrightarrow \neg\varphi$, since otherwise $\{\psi, \varphi\}$ could be extended to some $\Delta \in M$ satisfying $\Delta_j \in W(\psi) = X'$ and $\Delta_j \in W(\varphi) = X$ or $\{\neg\psi, \neg\varphi\}$ could be extended to some $\Omega \in M$ satisfying $\Omega_k \in W - W(\psi) = X$ and $\Omega_k \in W - W(\varphi) = X'$, but both situations are impossible because $X' \cap X = \emptyset$. Hence $\vdash \psi \leftrightarrow \neg\varphi$. Applying (E), we obtain $\vdash B\psi \leftrightarrow B\neg\varphi$ and therefore that $B\neg\varphi \in \Gamma$, contradicting the consistency of Γ. Conclusion: $X' \notin N(\Gamma_i)$.

  Second case for (d): assume $X \in N^+(\Gamma_i)$. This means there exists $Br \in \Gamma$ such that $r(\Gamma_i) \subseteq X$ and $\forall\theta \in \Gamma^r(B(r:\theta) \in \Gamma)$. Since $X \in N^+$, it follows that $X' \in N^+$ as well. So, toward a contradiction, assume $X' \in N^+(\Gamma_i)$. Then we have $Br' \in \Gamma$ such that $r'(\Gamma_i) \subseteq X'$ and $\forall\theta \in \Gamma^{r'}(B(r':\theta) \in \Gamma)$, and hence $r'(\Gamma_i) \cap r(\Gamma_i) = \emptyset$. If $\Gamma^r \cup \Gamma^{r'}$ were consistent, then we could extend this set to some $\Delta \in M$. Taking $j \neq i$, it would follow that $\Gamma^r \subseteq \Delta$ and $\Gamma^{r'} \subseteq \Delta$ and that $\Delta_j \neq \Gamma_i$. Hence we would have that $\Delta_j \in r'(\Gamma_i) \cap r(\Gamma_i) = \emptyset$, an impossibility. So $\Gamma^r \cup \Gamma^{r'}$ is not consistent. Applying the Consistency Lemma and the fact we have $\{Br, Br'\} \subseteq \Gamma$ with $\forall\theta \in \Gamma^r(B(r:\theta) \in \Gamma)$ and $\forall\theta \in \Gamma^{r'}(B(r':\theta) \in \Gamma)$, each of Γ^r and $\Gamma^{r'}$ is consistent, so it follows from the inconsistency of $\Gamma^r \cup \Gamma^{r'}$ that there exists a finite nonempty subset S of one of the two sets Γ^r and $\Gamma^{r'}$ such that for some formula φ that is a member of the other of these two sets we have $\vdash (\bigwedge S) \to \neg\varphi$. Let us assume $S \subseteq \Gamma^r$ and $\varphi \in \Gamma^{r'}$; the argument for the other possibility, where $S \subseteq \Gamma^{r'}$ and $\varphi \in \Gamma^r$, is similar. Proceeding, we have $\vdash (\bigwedge_{\chi\in S} r:\chi) \to r:\neg\varphi$ by modal reasoning. Since $S \subseteq \Gamma^r$, it follows by maximal consistency that $r:\neg\varphi \in \Gamma$ and hence $\neg\varphi \in \Gamma^r$. As we have $\forall\theta \in \Gamma^r(B(r:\theta) \in \Gamma)$, it follows that $B(r:\neg\varphi) \in \Gamma$. Since we also have that $Br \in \Gamma$, it follows by (BA) and maximal consistency that $B\neg\varphi \in \Gamma$. But $\varphi \in \Gamma^{r'}$, from which it follows by $\forall\theta \in \Gamma^{r'}(B(r':\theta) \in \Gamma)$ that $B(r':\varphi) \in \Gamma$. Since we also have that $Br' \in \Gamma$, it follows by (BA) and maximal consistency that $B\varphi \in \Gamma$ and therefore that $\neg B\neg\varphi \in \Gamma$ by (D) and maximal consistency. But then we have shown that $\neg B\neg\varphi \in \Gamma$ and $B\neg\varphi \in \Gamma$, which implies Γ is inconsistent, a contradiction. Conclusion: $X' \notin N^+(\Gamma_i)$.

So M_c is indeed a model, and therefore (M_c, Γ_1^θ) is a pointed model. Thus, since $\neg\theta \in \Gamma^\theta$, it follows by the Truth Lemma that $M_c, \Gamma_1^\theta \not\models \theta$. Completeness follows.

A.4 RBB_σ and RBB_σ^+ Soundness and Completeness

Recalling the semantics for RBB_σ and for RBB_σ^+ from §A.1, we prove the following theorem.

Theorem A.2 (RBB_σ and RBB_σ^+ Soundness and Completeness)**.** Assume R contains the symbol σ. For each $\varphi \in F$:

$$\begin{aligned} \mathsf{RBB}_\sigma \vdash \varphi \quad &\text{iff} \quad \models_\sigma \varphi \quad \text{and} \\ \mathsf{RBB}_\sigma^+ \vdash \varphi \quad &\text{iff} \quad \models_\sigma^+ \varphi \quad . \end{aligned}$$

Soundness for RBB_σ and for RBB_σ^+ are as for RBB (Theorem A.1) except that we must check soundness of the additional axioms. We consider each in turn.

- Validity of (MA): $\models_\sigma \sigma \to (Br \to r)$ and $\models_\sigma^+ \sigma \to (Br \to r)$.

  Assume $M, w \models_\sigma \sigma$ and $M, w \models_\sigma Br$. This means $w \in \sigma(w)$ and $r^\circ \in N(w)$, from which it follows by (ma) that $w \in r(w)$. But then $M, w \models_\sigma r$. The argument for the satisfaction operator $\models_\sigma^+$ is the same.

- Validity of (MB): $\models_\sigma B\sigma$ and $\models_\sigma^+ B\sigma$.

  Given (M, w), we have $\sigma^\circ \in N(w)$ by (mb). So $M, w \models_\sigma B\sigma$ and $M, w \models_\sigma^+ B\sigma$.

- Validity of (MR): $\models_\sigma B(r:\varphi) \to (Br \to B(\sigma:\varphi))$ and
  $\models_\sigma^+ B(r:\varphi) \to (Br \to B(\sigma:\varphi))$.

  Assume $M, w \models_\sigma B(r:\varphi)$ and $M, w \models_\sigma Br$. Then $[\![r:\varphi]\!]_M \in N(w)$ and $r^\circ \in N(w)$. Applying (mr), it follows that $[\![\sigma:\varphi]\!]_M \in N(w)$. But then $M, w \models_\sigma B(\sigma:\varphi)$. The argument for the satisfaction operator $\models_\sigma^+$ is the same.

- Validity of (MT): $\models_\sigma^+ B\varphi \to B(\sigma:\varphi)$.

  Assume $M, w \models_\sigma^+ B\varphi$. This means $[\![\varphi]\!]_M \in N(w)$. Applying (mt), we have $[\![\sigma:\varphi]\!]_M \in N(w)$. But this is what it means to have $M, w \models_\sigma^+ B(\sigma:\varphi)$.

So RBB_σ and RBB_σ^+ are sound. Completeness for RBB_σ and for RBB_σ^+ is as for RBB (Theorem A.1) except that provability is taken with respect to the theory in question and we must show that M_c satisfies the additional properties required of models of this theory (§A.1).

- M_c satisfies (ma) for RBB_σ and for RBB^+_σ: $\Gamma_i \in \sigma(\Gamma_i)$ and $r^\circ \in N(\Gamma_i)$ together imply that $\Gamma_i \in r(\Gamma_i)$.

  Assume $\Gamma_i \in \sigma(\Gamma_i)$ and $r^\circ \in N(w)$. As in the proof that M_c satisfies (ba) from Theorem A.1, it follows from $r^\circ \in N(\Gamma_i)$ that $Br \in \Gamma$. Applying the definition of $\sigma(\Gamma_i)$ to our assumption $\Gamma_i \in \sigma(\Gamma_i)$, it follows that $\neg\sigma \notin \Gamma$ and therefore $\sigma \in \Gamma$ by maximal consistency. Since $\sigma \in \Gamma$ and $Br \in \Gamma$, we have by (MA) and maximal consistency that $r \in \Gamma$. But then $\Gamma^r \subseteq \Gamma$ by (A) and maximal consistency. Since it follows from $r \in \Gamma$ by maximal consistency that $\neg r \notin \Gamma$ and we have shown that $\Gamma^r \subseteq \Gamma$, it follows by the definition of $r(\Gamma_i)$ that $\Gamma_i \in r(\Gamma_i)$.

- M_c satisfies (mb) for RBB_σ and for RBB^+_σ: $\sigma^\circ \in N(\Gamma_i)$.

  We have $B\sigma \in \Gamma$ by (MB). Hence $\sigma^\circ = W(\sigma) \in N(\Gamma_i)$.

- M_c satisfies (mr) for RBB_σ and for RBB^+_σ: $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ and $r^\circ \in N(\Gamma_i)$ together imply that $[\![\sigma:\varphi]\!]_{M_c} \in N(\Gamma_i)$.

  Assume $[\![r:\varphi]\!]_{M_c} \in N(\Gamma_i)$ and $r^\circ \in N(\Gamma_i)$. As in the proof that M_c satisfies (ba) from Theorem A.1, it follows that $B(r:\varphi) \in \Gamma$ and $Br \in \Gamma$. Applying (MR) and maximal consistency, we obtain $B(\sigma:\varphi) \in \Gamma$. By the definition of $N(\Gamma_i)$, it follows that $W(\sigma:\varphi) \in N(\Gamma_i)$. Applying the Truth Lemma, we obtain $[\![\sigma:\varphi]\!]_{M_c} \in N(\Gamma_i)$.

- M_c satisfies (mt) for RBB^+_σ: $[\![\varphi]\!]_{M_c} \in N(\Gamma_i)$ implies $[\![\sigma:\varphi]\!]_{M_c} \in N(\Gamma_i)$.

  Assume $[\![\varphi]\!]_{M_c} \in N(\Gamma_i)$. Applying the Truth Lemma, we obtain $W(\varphi) \in N(\Gamma_i)$. We therefore have $B\varphi \in \Gamma$ by the definition of $N(\Gamma_i)$. Applying (MT) and maximal consistency, it follows from $B\varphi \in \Gamma$ that $B(\sigma:\varphi) \in \Gamma$. Applying the definition of $N(\Gamma_i)$, we obtain $W(\sigma:\varphi) \in N(\Gamma_i)$. We therefore conclude that $[\![\sigma:\varphi]\!]_{M_c} \in N(\Gamma_i)$ by the Truth Lemma.

A.5 Lemmas for QRBB Completeness

The results of this section will be used in §A.6 to prove completeness for QRBB. All provability in this section is taken with respect to QRBB.

Lemma A.3. QRBB satisfies the following:

- *Distributivity*: $\vdash (\forall r)(\varphi \to \psi) \to ((\forall r)\varphi \to (\forall r)\psi)$;
- the *Distribution Rule*: $\vdash \varphi \to \psi$ implies $\vdash (\forall r)\varphi \to (\forall r)\psi$;
- the *Renaming Rule*: if s has no occurrence in φ, then

$$\vdash (\forall r)\varphi \leftrightarrow (\forall s)\varphi[s/r] \quad \text{and} \quad \vdash (\exists r)\varphi \leftrightarrow (\exists s)\varphi[s/r] \ ;$$

- the *Equivalence Rule*: $\vdash \varphi \leftrightarrow \varphi'$ implies

$$\vdash (\forall r)\varphi \leftrightarrow (\forall r)\varphi' \quad \text{and} \quad \vdash (\exists r)\varphi \leftrightarrow (\exists r)\varphi' \ .$$

Proof. For Distributivity:

1.	$\vdash (\forall r)(\varphi \to \psi) \to (\varphi \to \psi)$	(UI)
2.	$\vdash (\forall r)\varphi \to \varphi$	(UI)
3.	$\vdash ((\forall r)(\varphi \to \psi) \wedge (\forall r)\varphi) \to \psi$	by 1, 2
4.	$\vdash (\forall r)[((\forall r)(\varphi \to \psi) \wedge (\forall r)\varphi) \to \psi]$	by 3, (Gen)
5.	$\vdash ((\forall r)(\varphi \to \psi) \wedge (\forall r)\varphi) \to (\forall r)\psi$	by 4, (UD), (MP)
6.	$\vdash (\forall r)(\varphi \to \psi) \to ((\forall r)\varphi \to (\forall r)\psi)$	by 5

For the Distribution Rule:

1.	$\vdash \varphi \to \psi$	assumption
2.	$\vdash (\forall r)\varphi \to \varphi$	(UI)
3.	$\vdash (\forall r)\varphi \to \psi$	by 1, 2
4.	$\vdash (\forall r)\varphi \to (\forall r)\psi$	by 3, (UD), (MP)

For the Renaming Rule: if s has no occurrence in φ, then

1.	$\vdash (\forall r)\varphi \to \varphi[s/r]$	(UI)
2.	$\vdash (\forall s)((\forall r)\varphi \to \varphi[s/r])$	by 1, (Gen)
3.	$\vdash (\forall r)\varphi \to (\forall s)\varphi[s/r]$	by 2, (UD), (MP)
4.	$\vdash (\forall s)\varphi[s/r] \to \varphi[s/r][r/s]$	(UI)
5.	$\vdash (\forall s)\varphi[s/r] \to \varphi$	by 4, $\varphi[s/r][r/s] = \varphi$
6.	$\vdash (\forall r)((\forall s)\varphi[s/r] \to \varphi)$	by 5, (Gen)
7.	$\vdash (\forall s)\varphi[s/r] \to (\forall r)\varphi$	by 6, (UD), (MP)
8.	$\vdash (\forall r)\varphi \leftrightarrow (\forall s)\varphi[s/r]$	by 3, 7 — our first result
9.	$\vdash (\forall r)\neg\varphi \leftrightarrow (\forall s)(\neg\varphi)[s/r]$	by an argument like 1–8
10.	$\vdash (\forall r)\neg\varphi \leftrightarrow (\forall s)\neg\varphi[s/r]$	by 9, $(\neg\varphi)[s/r] = \neg(\varphi[s/r])$
11.	$\vdash \neg(\forall r)\neg\varphi \leftrightarrow \neg(\forall s)\neg\varphi[s/r]$	by 10
12.	$\vdash (\exists r)\varphi \leftrightarrow (\exists s)\varphi[s/r]$	by 11

For the Equivalence Rule:

1.	$\vdash \varphi \leftrightarrow \varphi'$	assumption
2.	$\vdash \varphi \to \varphi'$	by 1
3.	$\vdash (\forall r)\varphi \to (\forall r)\varphi'$	by 2, Distribution Rule
4.	$\vdash \varphi' \to \varphi$	by 1
5.	$\vdash (\forall r)\varphi' \to (\forall r)\varphi$	by 4, Distribution Rule
6.	$\vdash (\forall r)\varphi \leftrightarrow (\forall r)\varphi'$	by 3, 5 — our first result
7.	$\vdash \neg\varphi \leftrightarrow \neg\varphi'$	by 1
8.	$\vdash (\forall r)\neg\varphi \leftrightarrow (\forall r)\neg\varphi'$	from 7 by an argument like 1–6
9.	$\vdash \neg(\forall r)\neg\varphi \leftrightarrow \neg(\forall r)\neg\varphi'$	by 8
10.	$\vdash (\exists r)\varphi \leftrightarrow (\exists r)\varphi'$	by 9

□

Lemma A.4. If r is not free in ψ, then:

1. $\vdash (\forall r)(\varphi \to \psi) \to ((\exists r)\varphi \to \psi)$; and
2. $\vdash (\exists r)(\psi \to \varphi) \to (\psi \to (\exists r)\varphi)$.

Proof. For Item 1: if r is not free in ψ, then

1.	$\vdash (\forall r)(\varphi \to \psi) \to (\forall r)(\neg\psi \to \neg\varphi)$	Equivalence Rule, Classical Logic
2.	$\vdash (\forall r)(\neg\psi \to \neg\varphi) \to (\neg\psi \to (\forall r)\neg\varphi)$	(UI)
3.	$\vdash (\forall r)(\varphi \to \psi) \to (\neg\psi \to (\forall r)\neg\varphi)$	by 1, 2
4.	$\vdash (\forall r)(\varphi \to \psi) \to (\neg(\forall r)\neg\varphi \to \psi)$	from 3
5.	$\vdash (\forall r)(\varphi \to \psi) \to ((\exists r)\varphi \to \psi)$	from 4

For Item 2: if r is not free in ψ, then

1.	$\vdash \psi \to (\neg\varphi \to (\psi \wedge \neg\varphi))$	(CL)
2.	$\vdash (\forall r)\psi \to (\forall r)(\neg\varphi \to (\psi \wedge \neg\varphi))$	by 1, Distribution Rule
3.	$\vdash (\forall r)\psi \to ((\forall r)\neg\varphi \to (\forall r)(\psi \wedge \neg\varphi))$	by 2, Distributivity
4.	$\vdash ((\forall r)\psi \wedge (\forall r)\neg\varphi) \to (\forall r)(\psi \wedge \neg\varphi)$	by 3
5.	$\vdash \psi \to \psi$	(CL)
6.	$\vdash (\forall r)(\psi \to \psi)$	by 5, (Gen)
7.	$\vdash \psi \to (\forall r)\psi$	by 6, (UD), (MP)
8.	$\vdash (\psi \wedge (\forall r)\neg\varphi) \to (\forall r)(\psi \wedge \neg\varphi)$	by 4, 7
9.	$\vdash \neg(\forall r)(\psi \wedge \neg\varphi) \to (\psi \to \neg(\forall r)\neg\varphi)$	by 8
10.	$\vdash \neg(\forall r)\neg(\psi \to \varphi) \to (\psi \to \neg(\forall r)\neg\varphi)$	by 9, Equivalence Rule
11.	$\vdash (\exists r)(\psi \to \varphi) \to (\psi \to (\exists r)\varphi)$	by 10

□

A.6 QRBB Soundness and Completeness

Recalling the semantics for QRBB from §A.2, we prove the following theorem.

Theorem A.5 (QRBB Soundness and Completeness)**.** We have:

- QRBB is sound: QRBB $\vdash \varphi$ implies $\models \varphi$ for each $\varphi \in F^{\forall}$; and
- if R is at least countably infinite, then QRBB is sound and complete: for each $\varphi \in F^{\forall}$,

$$\mathsf{QRBB} \vdash \varphi \quad \text{iff} \quad \models \varphi \ .$$

Soundness is by induction on the length of derivation. Most cases are addressed in the proof of Theorem A.1. We only address the remaining cases.

- Validity of (UD): $\models (\forall r)(\varphi \to \psi) \to (\varphi \to (\forall r)\psi)$, where r is not free in φ.

  Assume $M, w \models (\forall r)(\varphi \to \psi)$ and $M, w \models \varphi$. From the former, we have $M, w \models (\varphi \to \psi)[s/r]$ for each s free for r in $\varphi \to \psi$. Since r is not free in φ, it follows that $M, w \models \varphi \to \psi[s/r]$ for each s free for r in ψ. By our assumption $M, w \models \varphi$, it follows that $M, w \models \psi[s/r]$ for each s free for r in ψ. That is, $M, w \models (\forall r)\psi$.

- Validity of (UI): $\models (\forall r)\varphi \to \varphi[s/r]$, where s is free for r in φ.

  By the definition of satisfaction.

- Validity of (EP) and (EN): $\models r = r$ and $\models \neg(r = s)$, where r and s are different.

  By the definition of satisfaction.

- (Gen) preserves validity: $\models \varphi$ implies $\models (\forall r)\varphi$.

  If $\not\models (\forall r)\varphi$, then there exists (M, w) and s free for r in φ such that $M, w \not\models \varphi[s/r]$. Given that $M = (W, [\cdot], N, V)$, define the model $M' = (W, [\cdot]', N, V)$ by setting

$$[t]' := \begin{cases} [s] & \text{if } t = r, \\ [t] & \text{otherwise.} \end{cases}$$

  It follows that $r^{M'}(w) = s^M(w)$ and $t^{M'}(w) = t^M(w)$ for all $t \neq r$. By the usual arguments about the preservation of truth of formulas under the renaming of quantified variables and their corresponding bound occurrences, we may assume without loss of generality that every occurrence of r in φ is free. But then it is easy to see that we have $M, w \not\models \varphi[s/r]$ iff $M', w \not\models \varphi$. After all, φ and $\varphi[s/r]$ differ only in the occurrences of r in φ that are replaced by s in $\varphi[s/r]$, and M' interprets all such occurrences of r just as M interprets the corresponding occurrences of s. And all other occurrences of symbols in φ are the same as they are in $\varphi[s/r]$, they are syntactically different than r, and M' and M interpret them in the same way. Conclusion: $\not\models \varphi$.

So QRBB is sound.

For completeness, we adapt the standard Henkin-style construction in [31, §3.1] to the present setting. To begin, our language $F^{\forall}$ depends on two parameters: a nonempty set R of reasons and a nonempty set P of propositional letters. We shall keep P fixed but consider different options for R.

As such, it will be convenient to write $L(R)$ to denote the set of formulas with quantifiers that we can form using $R \neq \emptyset$ as our set of reasons. By convention in this proof, we restrict all derivation to be with respect to QRBB. Also, we shall assume for the remainder of the argument that R is at least countably infinite.

To say that a set $\Gamma \subseteq L(R)$ is *consistent* means that for no finite $\Gamma' \subseteq \Gamma$ is it the case that $\vdash (\bigwedge \Gamma') \to \bot$. To say that $\Gamma \subseteq L(R)$ is *maximal* $L(R)$*-consistent* means that Γ is consistent and adding to Γ any formula of $L(R)$ not already present would produce an inconsistent set.

For the purposes of the present proof, a *theory* in the language $L(R)$ is a set $T \subseteq L(R)$ of formulas in $L(R)$ satisfying the following properties:

- *Closure under theorems*: if $\varphi \in L(R)$ and $\vdash \varphi$, then $\varphi \in T$; and
- *Closure under* (MP): if $\varphi \to \psi \in T$ and $\varphi \in T$, then $\psi \in T$.

Given $\Gamma \subseteq L(R)$, let $\mathbb{T}_R(\Gamma)$ be the set of all theories in $L(R)$ that contain Γ. The intersection of a collection of theories in $L(R)$ is also a theory in $L(R)$. Hence for each $\Gamma \subseteq L(R)$, we may define the theory

$$T_R(\Gamma) := \bigcap \mathbb{T}_R(\Gamma)$$

called the *theory in* $L(R)$ *generated by* Γ.

An $L(R)$*-proof from* Γ is a finite nonempty sequence $\langle \varphi_1, \dots, \varphi_n \rangle$ of formulas in $L(R)$ such that for each φ_i in the sequence, we have one of the following: $\varphi_i \in \Gamma$, $\mathsf{QRBB} \vdash \varphi_i$, or there exist φ_j and φ_k from earlier in the sequence (i.e., $j < i$ and $k < i$) such that φ_i follows by (MP) from φ_j and φ_k (i.e., $\varphi_k = \varphi_j \to \varphi_i$). To say that an $L(R)$-proof from Γ is *of* φ, means that φ is the last formula in the sequence. We write $\Gamma \vdash_R \varphi$ to mean that there exists an $L(R)$-proof of φ from Γ. Notation: in writing the set to the left of the turnstile $\vdash_R$, we may use a comma to denote set-theoretic union, we may identify an individual formula with the singleton set containing the formula in question, and we may omit any set-indicating notation if the set is empty. We state without proof the following results, grouped together under the name *Simple Lemma*:

- $\Gamma \vdash_R \varphi$ iff $\varphi \in T_R(\Gamma)$;
- $\Gamma \vdash_R \varphi$ iff $T_R(\Gamma) \vdash_R \varphi$;
- $\Gamma \vdash_R \varphi$ iff there exists a $\Gamma' \subseteq \Gamma$ such that $\Gamma' \vdash_R \varphi$;
- $\Gamma \vdash_R \varphi$ iff there exists a finite $\Gamma' \subseteq \Gamma$ such that $\Gamma' \vdash_R \varphi$;
- if Γ is finite, then $\Gamma' \vdash_R \varphi$ iff $\bigwedge \Gamma' \vdash_R \varphi$, where $\bigwedge \Gamma' := \bigwedge_{\chi \in \Gamma'} \chi$; and
- $\vdash_R \varphi$ iff $\mathsf{QRBB} \vdash \varphi$.

Generally the Simple Lemma will be used only tacitly.

Given a theory T in $L(R)$ and a theory T' in $L(R')$, to say that T' is an *extension* of T means that $T \subseteq T'$. To say that T' is a *conservative extension* of T means that $T' \cap L(R) = T$.

To say that a theory T in $L(R)$ is *Henkin* means that for each closed formula (i.e., containing no free variables) of the form $\neg(\forall r)\varphi \in L(R)$, there exists a reason $r_\varphi \in R$ called a *witness (or Henkin constant) for* $\neg(\forall r)\varphi$ for which we have

$$(\neg(\forall r)\varphi \to \neg\varphi[r_\varphi/r]) \in T \ .$$

Given a theory T in $L(R)$, let R^* be the set obtained from R by adding for each closed $\neg(\forall r)\varphi \in L(R)$ a new reason r_φ. To be clear: there is a bijection between the set of closed formulas $\neg(\forall r)\varphi \in L(R)$ and the set $R^* - R$. We define the set

$$H(R) := \{\neg(\forall r)\varphi \to \neg\varphi[r_\varphi/r] \mid \neg(\forall r)\varphi \in L(R) \text{ is closed }\}$$

of *Henkin axioms in* $L(R)$ and let $T^* := T_{R^*}(T \cup H(R))$ be the theory in $L(R^*)$ generated by $T \cup H(R)$.

Lemma A.6 (Constants)**.** Assume R is at least countably infinite and $R \subseteq R'$. If $\Gamma \cup \{\varphi\} \subseteq L(R)$, then $\Gamma \vdash_{R'} \varphi$ iff $\Gamma \vdash_R \varphi$.

Proof. The right-to-left direction is immediate (since $R \subseteq R'$), so we prove only the left-to-right direction. Proceeding, assume $\Gamma \cup \{\varphi\} \subseteq L(R)$ and $\Gamma \vdash_{R'} \varphi$; that is, there exists an $L(R')$-proof $\pi' = \langle \psi'_1, \dots, \psi'_n \rangle$ of φ from Γ. Let $r'_1, \dots, r'_m$ be a non-repeating list of all reasons in $R' - R$ that appear in π'. Since R is at least countably infinite and π' is finite, we may choose a non-repeating list $r_1, \dots, r_m$ of reasons in R that do not appear anywhere in π'. Such a list exists because R is at least countably infinite. Form $\pi := \langle \psi_1, \dots, \psi_n \rangle$ by defining ψ_i as the formula obtained from ψ'_i by replacing all occurrences of $r'_1, \dots, r'_m$ by $r_1, \dots, r_m$ (respectively). Since $\Gamma \cup \{\varphi\} \subseteq L(R)$, one may verify that π is an $L(R)$-proof of φ from Γ (i.e., formulas in $\Gamma \subseteq L(R)$ are left unchanged, QRBB-theorems in $L(R')$ are mapped to QRBB-theorems in $L(R)$, formulas in $L(R')$ obtained via (MP) in π' are mapped to formulas in $L(R)$ obtained via (MP) in π, and $\varphi \in L(R)$ is left unchanged). Hence $\Gamma \vdash_R \varphi$. $\square$

Lemma A.7 (Deduction)**.** For each R, we have:

$$\Gamma \cup \{\varphi\} \vdash_R \psi \quad \text{iff} \quad \Gamma \vdash_R \varphi \to \psi \ .$$

Proof. The right-to-left direction is easy, so we only address the left-to-right direction. Proceeding, assume $\Gamma \cup \{\varphi\} \vdash_R \psi$, which implies there exists an $L(R)$-proof $\langle \chi_1, \dots, \chi_n \rangle$ of ψ from $\Gamma \cup \{\varphi\}$. It suffices for us to prove by induction on $i \leq n$ that $\Gamma \vdash_R \varphi \to \chi_i$.

In the base case, $i = 1$ and therefore either $\vdash \chi_i$ or $\chi_i \in \Gamma \cup \{\varphi\}$. If $\vdash \chi_i$, then $\vdash \varphi \to \chi_i$ and therefore $\Gamma \vdash_R \varphi \to \chi_i$. If $\chi_i \in \Gamma \cup \{\varphi\}$, then either $\chi_i \in \Gamma$ or $\chi_i = \varphi$. If $\chi_i = \varphi$, then since $\vdash \varphi \to \varphi$ by (CL), we have $\Gamma \vdash_R \varphi \to \varphi$. So suppose $\chi_i \in \Gamma$. Hence $\Gamma \vdash_R \chi_i$. Since for $\epsilon_i := \chi_i \to (\varphi \to \chi_i)$ we have $\vdash \epsilon_i$ by (CL), we have $\Gamma \vdash_R \epsilon_i$ and therefore $\Gamma \vdash_R \varphi \to \chi_i$.

For the induction step ($i > 1$), we have that $\vdash \chi_i$, that $\chi_i \in \Gamma \cup \{\varphi\}$, or that χ_i follows by (MP) from χ_k and $\chi_k \to \chi_i$ appearing earlier in the $L(R)$-proof. The argument for the first two possibilities is as in the base case. So assume the third possibility obtains. By the induction hypothesis, we have $\Gamma \vdash_R \varphi \to \chi_k$ and $\Gamma \vdash_R \varphi \to (\chi_k \to \chi_i)$ Let θ_i be the classical tautology

$$\theta_i := (\varphi \to (\chi_k \to \chi_i)) \to ((\varphi \to \chi_k) \to (\varphi \to \chi_i)) \ .$$

We have $\vdash \theta_i$ by (CL) and hence $\Gamma \vdash_R \theta_i$. But then $\Gamma \vdash_R \varphi \to \chi_i$ by (MP). $\square$

We remark that the version of Lemma A.7 for the QRBB consequence relation $\vdash$ does not hold in general. For example, we have $r\!:\!\varphi \vdash (\forall r)(r\!:\!\varphi)$ and yet $\nvdash r\!:\!\varphi \to (\forall r)(r\!:\!\varphi)$. As another example, we have $p \vdash r\!:\!p$ and yet $\nvdash p \to r\!:\!p$. Lemma A.7 does not fail in similar ways because the consequence relation given by the R-specific turnstile $\vdash_R$ gives rise to a notion of proof (i.e., the $L(R)$-proof) that forbids the direct use of any QRBB rule other than (MP).

Lemma A.8 (Fresh Variable)**.** If $s \in R$ does not occur in any formula in $\Gamma \cup \{\varphi\} \subseteq L(R)$, then, letting $\Gamma[s/r] := \{\chi[s/r] \mid \chi \in \Gamma\}$, we have:

1. $\vdash \varphi$ iff $\vdash \varphi[s/r]$, and
2. $\Gamma \vdash_R \varphi$ iff $\Gamma[s/r] \vdash_R \varphi[s/r]$.

Proof. 1 follows by induction on the length of QRBB derivations. 2 follows by induction on the length of $L(R)$-proofs and makes use of 1. $\square$

Lemma A.9 (Conservativity)**.** Assume R is at least countably infinite. If T is a theory in $L(R)$, then T^* is a conservative extension of T.

Proof. We prove that for each $\varphi \in L(R)$, we have $T^* \vdash_{R^*} \varphi$ iff $T \vdash_R \varphi$. The right-to-left direction is immediate, so we only need prove the left-to-right direction. Proceeding, if $\varphi \in L(R)$, then we have $T^* \vdash_{R^*} \varphi$ iff there exists a finite set $H \subseteq H(R)$ of Henkin axioms satisfying $T, H \vdash_{R^*} \varphi$. So it suffices for us to prove by induction on the finite cardinality of $H \subseteq H(R)$ that $T, H \vdash_{R^*} \varphi$ implies $T \vdash_R \varphi$. The base case (where $H = \emptyset$) follows by Lemma A.6, so we proceed directly to the induction step. That is, we assume that the result holds for $H \subseteq H(R)$ having $|H| = n$ and we prove the result holds for $H \subseteq H(R)$ having $|H| = n + 1$. Proceeding, take $H \subseteq H(R)$ satisfying $|H| = n + 1$, choose a Henkin axiom $h \in H$ with

$$h = \neg(\forall r)\psi \to \neg\psi[r_\psi/r] \ ,$$

and define $H' := H - \{h\}$ so that $H = H' \cup \{h\}$ and $|H'| = n$. Now assume $T, H \vdash_{R^*} \varphi$ with $\varphi \in L(R)$, and hence $T, H', h \vdash_{R^*} \varphi$. It follows that there is a finite $T' \subseteq T$ such that $T', H', h \vdash_{R^*} \varphi$. Let $s \in R$ be a variable not occuring in any formula in the finite set $T' \cup H' \cup \{h, \varphi\}$. Such s exists because R is at least countably infinite. Define

$$h' := \neg(\forall r)\psi \to \neg\psi[s/r] \ .$$

Then, omitting mention of instances of classical reasoning and the use of the Simple Lemma in the last six lines of the derivation, we have:

$T', H', h \vdash_{R^*} \varphi$	(derived above)
$T', H', h' \vdash_{R^*} \varphi$	Lemma A.8
$\vdash_{R^*} \bigwedge(T' \cup H') \to (h' \to \varphi)$	Lemma A.7
$\vdash \bigwedge(T' \cup H') \to (h' \to \varphi)$	Simple Lemma
$\vdash (\forall s)(\bigwedge(T' \cup H') \to (h' \to \varphi))$	(Gen)
$\vdash \bigwedge(T' \cup H') \to (\forall s)(h' \to \varphi)$	(UD), no s in $T' \cup H'$
$\vdash_{R^*} \bigwedge(T' \cup H') \to (\forall s)(h' \to \varphi)$	Simple Lemma
$T', H' \vdash_{R^*} (\forall s)(h' \to \varphi)$	Lemma A.7
$T', H' \vdash_{R^*} (\forall s)((\neg(\forall r)\psi \to \neg\psi[s/r]) \to \varphi)$	write out h'
$T', H' \vdash_{R^*} (\neg(\forall r)\psi \to (\exists s)\neg\psi[s/r]) \to \varphi$	Lem. A.4, no s in φ or $\neg(\forall r)\psi$
$T', H' \vdash_{R^*} (\neg(\forall r)\psi \to \neg(\forall s)\neg\neg\psi[s/r]) \to \varphi$	definition of $\exists$
$T', H' \vdash_{R^*} (\neg(\forall r)\psi \to \neg(\forall s)\psi[s/r]) \to \varphi$	Equivalence (Lemma A.3)
$T', H' \vdash_{R^*} \varphi$	Renaming (Lemma A.3)

Therefore $T, H' \vdash_{R^*} \varphi$. Applying the induction hypothesis, $T \vdash_R \varphi$. $\square$

Now, given a theory T in $L(R)$, define:

$$\begin{array}{lcl@{\qquad}lcl} T_0 & = & T & R_0 & = & R \\ T_{i+1} & = & (T_i)^* \text{ for } i \in \omega & R_{i+1} & = & (R_i)^* \text{ for } i \in \omega \\ T_\omega & = & \bigcup_{i\in\omega} T_i & R_\omega & = & \bigcup_{i\in\omega} R_i \end{array}$$

Lemma A.10 (Henkin)**.** Let R be at least countably infinite and T be a theory in $L(R)$. Then T_ω is a Henkin theory that is a conservative extension of T.

Proof. Take a closed $\neg(\forall r)\varphi \in L(R_\omega)$. Then there exists $i \in \omega$ such that $\neg(\forall r)\varphi \in L(R_i)$. But then there is a witness $r_\varphi \in L(R_{i+1}) \subseteq L(R_\omega)$ such that

$$\neg(\forall r)\varphi \to \neg\varphi[r_\varphi/r] \in T_{i+1} \subseteq T_\omega \ \ .$$

So T_ω is a Henkin theory.

By induction on $i \in \omega$, we prove that T_i is a conservative extension of T. Base case: $T_0 = T$ and the result is immediate. Induction step: T_{i+1} is a conservative extension of T_i by Lemma A.9; that is, $T_{i+1} \cap L(R_i) = T_i$. By the induction hypothesis, $T_i \cap L(R) = T$. But then, since $L(R) \subseteq L(R_j) \subseteq L(R_k)$ if $j < k$, we have

$$T_{i+1} \cap L(R) = (T_{i+1} \cap L(R_i)) \cap L(R) = T_i \cap L(R) = T \ \ .$$

So each T_i is a conservative extension of T. But then

$$T_\omega \cap L(R) = (\textstyle\bigcup_{i\in\omega} T_i) \cap L(R) = \bigcup_{i\in\omega}(T_i \cap L(R)) = T \ \ ,$$

which shows that T_ω is a conservative extension of T. $\square$

By the usual Lindenbaum argument (using Zorn's Lemma) [31, §3.1], for each R, any consistent set in $L(R)$ may be extended to a maximal $L(R)$-consistent set. Hence for a consistent theory T in $L(R)$, the theory T_ω in $L(R_\omega)$ is consistent and may be extended to a maximal $L(R_\omega)$-consistent set T'_ω. This set is a theory in $L(R_\omega)$. Further, this theory is Henkin because $T_\omega \subseteq T'_\omega$, both theories are in the same language, T_ω is Henkin by Lemma A.10, and any extension of a Henkin theory within the same language is still Henkin (because all Henkin axioms are already present).

To prove completeness of QRBB, we take θ such that $\mathsf{QRBB} \nvdash \theta$. We construct a structure $M_c = (W, [\cdot], N, V)$ as in the proof of Theorem A.1 except that our set of worlds W is defined differently. First, let M_0 be the set of all maximal $L(R)$-consistent sets; each such set is a theory in $L(R)$. For each theory $T \in M_0$, define $M_\omega(T)$ to be the set of all maximal $L(R_\omega)$-consistent extensions of T_ω. As we have seen, each member of $M_\omega(T)$ is a maximal $L(R_\omega)$-consistent Henkin theory that is conservative over T (Lemma A.10). Define the set

$$M := \textstyle\bigcup_{T\in M_0} M_\omega(T)$$

whose members are maximal $L(R_\omega)$-consistent extensions of T_ω for each $T \in M_0$. It follows that $\{\neg\theta\}$ can be extended to a $T^\theta \in M_0$ and hence neither $M_\omega(T^\theta)$ nor M is empty. We define $W := M \times \{1, 2\}$ and write $(\Gamma, i) \in W$ in the abbreviated form Γ_i. Since M is nonempty, W is nonempty. The remaining components of M_c are defined as in the proof of Theorem A.1 except that all language-specific aspects of definitions are extended to the larger language $L(R_\omega)$.

We prove the *Truth Lemma*: for each formula $\varphi \in L(R_\omega)$ and world $\Gamma_i \in W$, we have $\varphi \in \Gamma$ iff $M_c, \Gamma_i \models \varphi$. The proof is by induction on the construction of formulas. The arguments for all but two cases are as in the proof of Theorem A.1. All that remains are the equality and quantifier inductive step cases.

- Inductive step: $(s = r) \in \Gamma$ iff $M_c, \Gamma_i \models s = r$.

  By (EP) and (EN), we have $(s = r) \in \Gamma$ iff $s = r$. But the latter holds iff $M_c, \Gamma_i \models s = r$.

- Inductive step: $(\forall r)\varphi \in \Gamma$ iff $M_c, \Gamma_i \models (\forall r)\varphi$.

  If $(\forall r)\varphi \in \Gamma$, then it follows by maximal $L(R_\omega)$-consistency and (UI) that $\varphi[s/r] \in \Gamma$ for each $s \in R_\omega$ that is free for r in φ. By the induction hypothesis, we have $M_c, \Gamma_i \models \varphi[s/r]$ for each such $s \in R$. But this is what it means to have $M_c, \Gamma_i \models (\forall r)\varphi$.

  Conversely, if $M_c, \Gamma_i \models (\forall r)\varphi$, then it follows that $M_c, \Gamma_i \models \varphi[s/r]$ for all $s \in R_\omega$ free for r in φ. By the induction hypothesis, we have $\varphi[s/r] \in \Gamma$ for all such s. Since Γ is a Henkin theory, there is a Henkin constant $r_\varphi \in R_\omega$ for $\neg(\forall r)\varphi$. Hence $\varphi[r_\varphi/r] \in \Gamma$. But Γ contains the Henkin axiom

  $$\neg(\forall r)\varphi \to \neg\varphi[r_\varphi/r] \ ,$$

  and so we have by $L(R_\omega)$-consistency that $(\forall r)\varphi \in \Gamma$.

This completes the proof of the Truth Lemma.

The proof that M_c satisfies (pr), (brk), (ba), (as), and (d) is as in the proof of Theorem A.1. So M_c is indeed a model (and not just a pre-model).

To complete the proof of completeness, we recall that we obtained $T^\theta \in M_0$ as a maximal $L(R)$-consistent extension of $\{\neg\theta\}$. Hence there exists $\Gamma^\theta \in M_\omega(T^\theta) \subseteq M$. But Γ^θ is a maximal $L(R_\omega)$-consistent extension of $(T^\theta)_\omega$, and $(T^\theta)_\omega$ is a conservative extension of T^θ by Lemma A.10. Therefore, since $\theta \notin T^\theta$ by consistency, we have $\theta \notin \Gamma^\theta$. Applying the Truth Lemma, $M_c, \Gamma^\theta_1 \not\models \theta$. Completeness follows.

A.7 Conservativity of QRBB Over RBB

As a corollary of Theorems A.1 and A.5, we have the following.

Corollary A.11. QRBB is a conservative extension of RBB: for each $\varphi \in F$,

$$\mathsf{QRBB} \vdash \varphi \quad \text{iff} \quad \mathsf{RBB} \vdash \varphi \ .$$

Proof. The right-to-left direction is obvious (QRBB contains all the axioms and rules of RBB). The left-to-right direction follows by QRBB soundness (Theorem A.5) and RBB completeness (Theorem A.1). $\square$

A.8 QRBB_σ and QRBB^+_σ Soundness and Completeness

Recalling the semantics for QRBB_σ and for QRBB^+_σ from §A.2, we prove the following theorem.

Theorem A.12 (QRBB_σ and QRBB^+_σ Soundness and Completeness)**.** Assume R contains the symbol σ.

- QRBB_σ is sound: $\mathsf{QRBB}_\sigma \vdash \varphi$ implies $\models_\sigma \varphi$ for each $\varphi \in F^\forall$.

- if R is at least countably infinite, then QRBB_σ is sound and complete: for each $\varphi \in F^\forall$,

$$\mathsf{QRBB}_\sigma \vdash \varphi \quad \text{iff} \quad \models_\sigma \varphi \ .$$

- analogous soundness and completeness results hold for QRBB^+_σ with respect to the satisfaction relation $\models^+_\sigma$.

Soundness is proved as in Theorem A.2. Completeness is proved as in Theorem A.5, except that provability is taken with respect to either QRBB_σ or QRBB^+_σ and one must show (using an argument as in the completeness portion of Theorem A.2) that M_c satisfies the relevant properties.

References

[1] David M. Armstrong. *Belief, truth and knowledge.* CUP Archive, 1973.

[2] Sergei [N.] Artemov. The logic of justification. 1(4):477–513, December 2008.

[3] Sergei [N.] Artemov and Melvin Fitting. Justification logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, 2011.

[4] Alexandru Baltag, Bryan Renne, and Sonja Smets. The logic of justified belief change, soft evidence and defeasible knowledge. In L. Ong and R. de Queiroz, editors, *Logic, Language, Information and Computation: 19th International Workshop, WoLLIC 2012, Buenos Aires, Argentina, September 3–6, 2012. Proceedings*, volume 7456 of *Lecture Notes in Computer Science*, pages 168–190, Buenos Aires, Argentina, 2012. Springer-Verlag Berlin Heidelberg.

[5] Alexandru Baltag, Bryan Renne, and Sonja Smets. The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81, 2014.

[6] Torben Braüner. Hybrid logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, 2011.

[7] Brian F. Chellas. *Modal Logic: An Introduction.* Cambridge University Press, 1980.

[8] Roderick M. Chisholm. *Theory of knowledge.* Prentice Hall, Englewood Cliffs, NJ, 1977. 2nd edition.

[9] Michael Clark. Knowledge and grounds: A comment on Mr. Gettier's paper. *Analysis*, 24(2):46–48, 1963.

[10] Fred Dretske. Conclusive reasons. *Australasian Journal of Philosophy*, 1(49):1–22, 1971.

[11] Julien Dutant. *Knowledge, Methods, and the Impossibility of Error.* 2010. PhD, University of Geneva.

[12] Julien Dutant. Methods models for belief and knowledge. 2015. Manuscript, London.

[13] Paul Égré. Knowledge as *de re* true belief? *Synthese*, 194(5):1517–1529, 2017.

[14] Branden Fitelson. Strengthening the case for knowledge from falsehood. *Analysis*, 70(4):666–669, 2010.

[15] Richard Fumerton. Theories of justification. *The Oxford handbook of epistemology*, pages 204–233, 2002.

[16] Edmund Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963.

[17] Alvin I. Goldman. Discrimination and perceptual knowledge. *The Journal of Philosophy*, 73(20):771–791, 1976.

[18] Alvin I. Goldman. What is justified belief? In G. S. Pappas, editor, *Justification and Knowledge*, pages 1–23. Dordrecht: Reidel, 1979.

[19] Gilbert H. Harman. Knowledge, reasons, and causes. *The Journal of Philosophy*, pages 841–855, 1970.

[20] Daniel Kahneman and Shane Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and biases: The psychology of intuitive judgment*, pages 49–81. Cambridge University Press, 2002.

[21] Angelika Kratzer. Fact: Particulars or information units? *Linguistics and Philosophy*, 25:655–670, 2002.

[22] K. Lehrer and T. Jr Paxson. Knowledge: Undefeated justified true belief. *Journal of Philosophy*, 66:225–237, 1969.

[23] Keith Lehrer. *Theory of Knowledge*. Westview Press, 2nd edition, 2000.

[24] Ram Neta. A refutation of Cartesian fallibilism. *Noûs*, 45(4):658–695, 2011.

[25] Roy Sorensen. Fugu for logicians. *Philosophy and Phenomenological Research*, 2015. Forthcoming.

[26] Ernest Sosa. How do you know? *American Philosophical Quarterly*, pages 113–122, 1974.

[27] Ernest Sosa. Epistemic presupposition. In G. Pappas, editor, *Justification and Knowledge*, pages 79–92. Springer, 1979.

[28] Spinoza. *Ethica*. 1677. Latin and English versions available at `http://ethicadb.org/`.

[29] John Turri. Is knowledge justified true belief? *Synthese*, 184(3):247–259, 2012.

[30] Johan van Benthem and Eric Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1-3):61–92, 2011.

[31] Dirk van Dalen. *Logic and Structure*. Springer-Verlag, 1994.

[32] T. Warfield. Knowledge from falsehood. *Philosophical Perspectives*, pages 405–416, 2005.

[33] Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.

[34] Linda Zagzebski. The inescapability of Gettier problems. *The Philosophical Quarterly*, 44(74):65–73, 1994.