Ramsey's Thesis and Introspection Principles for Knowledge and Belief

Ben Holguín

January 2019

Abstract

Say that a propositional attitude R is positively introspective just in case Rp entails RRp, and is negatively introspective just in case $\neg Rp$ entails $R\neg Rp$. This paper is about an under-explored tension between two theses: the thesis that introspection principles for propositional attitudes like knowledge and belief are false; and the thesis, often referred to as "Ramsey's thesis", that the probability of an indicative conditional is equal to the conditional probability of its consequent given its antecedent. It will argue that given conservative assumptions about the semantics of the indicative conditional, the most plausible way of making good on Ramsey's thesis is to embrace at least one introspection principle for the relevant propositional attitudes. It will then explore the question of how those committed to both Ramsey's thesis and the falsity of introspection principles ought to react to this result, suggesting that it is worse news for the former than the latter. This is because there is good reason to believe that the dialectical dependence between Ramsey's thesis and introspection principles is asymmetric: though it is easy to find non-question-begging arguments from the falsity of introspection principles to the falsity of Ramsey's thesis, it is difficult to find non-question-begging arguments from the truth of Ramsey's thesis to the truth of introspection principles.

1 Introduction

Many epistemologists think introspection principles fail for both knowledge and belief. That is to say, many epistemologists think it is possible to know a proposition without being in a position to know that one knows it, or to believe a proposition without being in a position to believe that one believes it; and likewise that it is possible to fail to know a proposition without being in a position to know that one fails to know it, or to fail to believe a proposition without being in a position to believe that one fails to believe it.

Many philosophers of language think "Ramsey's thesis" holds for the indicative conditional (at least in some shape or form).[1] That is to say, many philosophers of language think the

[1] More on this qualification in §3.

probability of an indicative conditional is equal to the probability of the conditional's consequent conditional on its antecedent.

This paper will argue that these two popular views are in tension. If we reject introspection principles for knowledge and belief, we better reject Ramsey's thesis. Equivalently, if we accept Ramsey's thesis, we better accept at least one introspection principle for knowledge or belief. What is widely regarded as a simple platitude (at least among philosophers of language) turns out to presuppose a set of highly controversial epistemological assumptions.[2]

How ought we react to this conclusion? By letting the epistemologists figure out whether introspection principles for knowledge and belief are true, and then accepting or rejecting Ramsey's thesis accordingly. This is because there is a principle in the vicinity of Ramsey's thesis that is capable of accounting for the various intuitive considerations cited in favor of the original, but that does *not* presuppose introspection principles for knowledge and belief. Put roughly, it is the principle that the probability of the conditional is equal to the *known* (or *believed*) conditional probability. The availability of this surrogate principle suggests the dialectical dependence between Ramsey's thesis and introspection principles is asymmetric: though it will be easy to find non-question-begging arguments from the falsity of introspection principles to the falsity of Ramsey's thesis, it will be difficult to find non-question-begging arguments from the truth of Ramsey's thesis to the truth of introspection principles. We should thus settle the question of whether introspection principles are true before we address the question of whether Ramsey's thesis is.

Here is the plan for the paper. §2 will present and discuss introspection principles for knowledge and belief, §3 Ramsey's thesis. §4 will outline a "qualitative" version of Ramsey's thesis, and then argue that anyone who accepts Ramsey's thesis ought to accept the qualitative thesis as well. §5 will then argue that on some conservative assumptions about the semantics of the indicative conditional, the qualitative analog of Ramsey's thesis can be shown to presuppose at least one introspection principle for knowledge and belief. §6 will do two things. First, present the surrogate principle alluded to in the previous paragraph. And second, argue that the surrogate principle's only defect is an instance of a more general problem facing those who reject introspection principles. Finally, §7 will outline a strategy for answering the more general problem, concluding that the surrogate principle is fit to play the roles traditionally associated with Ramsey's thesis.

2 Introspection principles

We'll start with the introspection principles for knowledge and belief. Letting K stand for 'The relevant agent or group is in a position to know' and letting B stand for 'The relevant agent or

[2] To my knowledge, the only author whose work has touched on these sorts of connections between Ramsey's thesis and introspection principles is Dorst (2019a), who argues for positive introspection for knowledge partly through Ramsey's thesis-like considerations. See footnote 35 for further discussion.

group is in a position to rationally believe',[3] we can state the principles as follows:[4]

K Positive Introspection (KPI)

If Kp, then KKp.

K Negative Introspection (KNI)

If $\neg Kp$, then $K\neg Kp$.

B Positive Introspection (BPI)

If Bp, then BBp.

B Negative Introspection (BNI)

If $\neg Bp$, then $B\neg Bp$.

Among KPI–BNI, the only one whose status is close to a matter of consensus is KNI—and the (near) consensus is that it is untenable. The reason why is that for any false proposition p, $\neg Kp$. But it is not plausible that for any false proposition p, $K\neg Kp$. Why? Consider skeptical scenarios. A (recently envatted) brain-in-a-vat does not have hands, and is thus not in a position to know whether it has hands. But, modulo skepticism about knowledge, it should not follow that the brain-in-a-vat is in a position to *know* that it is not in a position to know whether it has hands. Handless brains-in-vats *believe* they have hands, after all. So, unless skepticism about knowledge is true—and for the purposes of this paper we will take for granted that that is not the case—KNI is false.

The remaining three principles are significantly more controversial.[5] And though the question of which of them is true is of considerable philosophical interest, answering it is not an ambition of this paper. Instead, the ambitions of this paper are: (i) to show that Ramsey's thesis should be accepted only by those who accept introspection principles (§§4–5); and (ii) to show

[3] A terminological note about 'belief'. Throughout I will work under the pretense that the attitude expressed by 'believes' is *strong*—i.e., that it is the same kind of attitude we talk about with expressions like 'outright believes', 'is sure', 'is certain', etc. Whether the English word 'believes' actually denotes such a relation is a matter of controversy: see, e.g., Hawthorne et al. (2016) and Dorst (2019c) for reasons to think not. Thankfully, we may set aside these issues in what follows. Those who are attracted to belief-theoretic accounts of the indicative conditional (and assertion and the like) tend to work implicitly with a strong notion of belief. And the interest is in the relationship between introspection principles *for the attitude that plays those roles* and Ramsey's thesis, regardless of whether that attitude is the one denoted by the English word 'believes'.

[4] Note: the 'If...then...' that appears in the statement of these principles should be interpreted as a necessitated material conditional. In other words, KPI is equivalent to the claim that:

Necessarily: $Kp \supset KKp$.

Likewise for the other principles that will be introduced over the course of the discussion.

[5] For work that ranges from modestly to fully sympathetic with (at least some) of the principles see, e.g.: Hintikka (1962); Dokic and Égré (2009); Stalnaker (2009); Almotahari and Glick (2010); Christensen (2010); Mchugh (2010); Cresto (2012); Smithies (2012); Cohen and Comesaña (2013); Elga (2013); Fernández (2013); Greco (2014a,b, 2015b,c); Titelbaum (2015); Das and Salow (2018); Dorst (2019a,b); Goodman and Salow (2018). And for work that ranges from modestly to fully unsympathetic with (at least some) of the principles, see, e.g.: Williamson (2000, 2011, 2014); Hawthorne and Magidor (2009, 2010); Lasonen-Aarnio (2010, 2014, 2019); Coates (2012); Hazlett (2012); Dorr et al. (2014).

that the question of whether the introspection principles are true ought to be settled *prior* to the question of whether Ramsey's thesis is (§§6–7). And neither (i) nor (ii) requires us to say anything about whether the introspection principles are actually true.

However, we will give special attention to a particular anti-introspection argument: Williamson's (2000, ch. 5) influential critique of KPI. As we will see momentarily, if one accepts the assumptions about K that lead Williamson to reject KPI, it's a short step to accepting assumptions about B that cast similar kinds of doubts on BPI and BNI. So the reasoning behind the argument has the benefit of representing a general (and quite popular) "anti-introspective" epistemological perspective, one that will help frame the significance of this paper's main results.

Williamson's argument begins with a case like the following:

Mr. Magoo

> Mr. Magoo is an adult with normal perceptual capacities judging the heights of trees at a distance. Given the limitations of his powers of discrimination, Mr. Magoo is such that whenever he judges the height of a tree to be x feet tall (from this distance), the actual height of the tree (in feet) could easily be anywhere between $x-10$ and $x+10$. But so long as Mr. Magoo is in normal conditions—which indeed he is—it is guaranteed to fall within that range. The tree Mr. Magoo is currently looking at appears to him to be about 100 feet tall. And in fact it *is* 100 feet tall. But given the limitations of his perceptual capacities, the strongest proposition Mr. Magoo knows about the tree's height is that it falls somewhere between 90 and 110 feet (inclusive).

To transform **Mr. Magoo** into a counterexample to KPI we will make use of two assumptions. The first is that what one is in a position to know is closed under material implication:

K Closure

> If $K(p \supset q)$, then $Kp \supset Kq$.

And where h is the actual height of the tree Mr. Magoo is looking at (in feet) and i is a relevant natural number, the second assumption is that Mr. Magoo knows the following principle:

Margin For Error

> If $h = i$, then: $\neg K(h < i+1)$.

MARGIN FOR ERROR says that if the actual height of the tree is i feet, then no one (in Mr. Magoo's circumstances) is in a position to know that the height of the tree is less than $i+1$ feet. Equivalently: if in fact $h = i$, then for all anyone (in Mr. Magoo's circumstances) knows, $h = i+1$.

The argument against KPI then runs as follows.[6] Suppose for reductio that KPI is true. By stipulation Mr. Magoo knows $90 \leq h \leq 110$. So he knows $h < 111$. And it follows from KPI that

[6] Here we will ignore the distinction between knowing and merely being in a position to know. I'll also note that this is a simplified version of Williamson's argument—interested readers should consult the original.

he knows that he knows this. MARGIN FOR ERROR (contraposed) says that $K(h < 111) \supset h \neq 110$. But since Mr. Magoo knows MARGIN FOR ERROR and knows that he knows the tree is less than 111 feet tall, it follows from K CLOSURE that Mr. Magoo is in a position to know that the tree is less than 110 feet tall (i.e., $K(h < 110)$). This contradicts the stipulations of the case. Worse, the reasoning involved here can be extended indefinitely, allowing Mr. Magoo to deduce the absurd conclusion that, for any natural number i, the tree is less than i feet tall. Thus, given MARGIN FOR ERROR and K CLOSURE, KPI must be false.

With respect to BPI and BNI, Hawthorne and Magidor (2009, 2010) argue that these sorts of margin-for-error considerations tell as much against these principles as they do KPI. I will not recapitulate the details of their arguments here. But the rough idea is that if one could know MARGIN FOR ERROR without believing that one obeys its belief-theoretic counterpart, then one's beliefs could knowably fail to aim at knowledge. Since rational (full) belief aims at knowledge, it follows from MARGIN FOR ERROR that rational agents believe they obey its belief-theoretic counterpart.And from there it is easy to construct an argument against BPI analogous to Williamson's argument against KPI. As for the connection between margin-for-error-style arguments against positive introspection principles (on the one hand) and negative introspection principles like BNI (on the other), the idea is simple. If we are used to thinking that the boundaries of what we can rationally believe are not transparent in the way they would need to be for BPI to be true, we should get used to thinking the same about the boundaries of what we *cannot* rationally believe. When one is just barely in a position to rationally believe p, one will not be in a position to rationally believe that one is in a position to rationally believe p; likewise, when one just barely fails to be in a position to rationally believe p, one will not be in a position to rationally believe that one fails to be in a position to rationally believe p.

That covers our brief sketch of the introspection principles, as well as a popular argument against them from MARGIN FOR ERROR. I will emphasize again that nothing in what follows presupposes the soundness of this argument. What matters is that we have it on the table.

Before getting to the other main principle of interest—Ramsey's thesis—it will be helpful to make a quick point about the generality of the paper's arguments, as well as some of its notation. The central claims of the paper are designed to be equally plausible whether about knowledge or belief. In fact, they are designed to be plausible for just about *any* propositional attitude one might want to use for the semantics of the indicative conditional and probabilistic vocabulary, be it *presupposition*, *certainty*, *evidence*, etc. Consequently, we will be using '$\Box$' as a placeholder propositional attitude throughout most of the discussion (while '$\Diamond$' will be shorthand for '$\neg\Box\neg$'). We will take for granted that $\Box$ is either K or B, but this is mostly just for the sake of concreteness. Readers should feel free to interpret '$\Box$' as they wish, provided that they do so uniformly. When the difference between factive and non-factive interpretations of $\Box$ might matter, we will be careful to treat the cases separately.

With this in mind, the official introspection principles will be as follows:

Positive Introspection (PI)

If $\Box p$, then $\Box\Box p$.

Negative Introspection (NI)

If $\neg\Box p$, then $\Box\neg\Box p$.

Finally, throughout the paper we will make use of the following two assumptions:

Closure

If $\Box(p \supset q)$, then $\Box p \supset \Box q$.

Rationality

If p is a priori and necessary, then $\Box p$; if $\neg p$ is a priori and necessary, then $\neg\Box p$.

closure is the $\Box$-ed analog of k closure. It is a standard assumption in the literature on Ramsey's thesis, and so should be common ground to all interested parties. rationality is there to capture the fact that principles like pi, closure, margin for error etc., are meant to be "conceptual" truths about the relevant notions—i.e., the kinds of claims that, if true, are both necessary and knowable a priori. Given that $\Box$ is stated in terms of *being in a position to* know/believe, rationality should also be common ground to all interested parties.

3 Ramsey's thesis

So much for the introspection principles. Our second principle of interest goes by many names: Ramsey's thesis, Adam's thesis, Stalnaker's thesis, the thesis, the test, the equation, Ramsey's test, Stalnaker's equation, and probably others. We'll continue to run with "Ramsey's thesis". In slogan form it says that the probability of an indicative conditional is the conditional probability of its consequent given its antecedent. The idea was first presented by Ramsey (1931), and was popularized by Adams (1965) and Stalnaker (1968, 1970) (hence the other names).[7]

By way of precisifying our slogan, I'll note that for the purposes of this paper we lose nothing if we anchor ourselves into a single epistemic context for the entirety of the discussion. So letting '$\rightarrow$' denote the indicative conditional, and 'P' a probability function supplied by context (say, the rational credences according to the contextually salient body of evidence), we will work the following object-language statement of the thesis:[8]

RT

If $P(p) > 0$, then: $P(p \rightarrow q) = P(q|p)$.

[7] See Hájek and Hall (1994); Edgington (1995); Douven and Verbrugge (2013) for helpful surveys of the issues surrounding Ramsey's thesis.

[8] Conditional probabilities are understood in the standard way: $P(q|p) =_{\text{def}} \frac{P(p \wedge q)}{P(p)}$, if $P(p) > 0$; else is undefined.

In words: so long as the contextually salient probability function gives p non-zero odds, it will give the indicative conditional $p \to q$ whatever odds it gives to: the conjunction of q and p divided by the odds it gives to p.

With an important caveat to be discussed in a moment, RT has considerable support among theorists of the indicative conditional. See, e.g., Williams (2009, p. 154) for a representative statement:[9]

> The link between something like a probability of a simple conditional and the corresponding conditional probability is a centerpiece of many accounts of the indicative conditional; clearly many philosophers have found it compelling enough to build theories around it (or some surrogate).

What about RT do theorists find so compelling? For starters, it seems like a platitude: how *could* the probability of $p \to q$ be anything other than the probability of q conditional on p?[10] With respect to more substantive arguments for RT, typically you won't see much more than quick judgments about thought experiments, like the following from Bacon (2015, p. 132):

> Suppose that a card has been picked at random from a standard 52 card deck and placed face down in front of you. Assuming that you are not more confident that some card will be selected over any other, how confident should you be about asserting the following sentences?
>
> (i) The selected card is an ace if it's red.
>
> (ii) It's spades if it's black.
>
> (iii) It's diamonds if it's an eight.
>
> ...The obvious answers to these questions are, in order: $\frac{1}{13}$, $\frac{1}{2}$, and $\frac{1}{4}$. For example, to calculate (ii) I just determine what proportion of black cards are spades. Since one in two black cards are spades the answer is $\frac{1}{2}$.

Here Bacon's idea is that the procedure for assessing the probability of indicative conditionals like (i)–(iii) seems to be exactly that of the procedure for assessing the conditional probability of their consequents given their antecedents. And presumably the thought is that without a principle like RT, it would be difficult to explain why this is so.

A similar style of argument can be constructed on judgments about natural language sentences like the following:

(1) a. ✓ There's a 25% chance the coin first lands heads and then lands tails, and a 50% chance the coin first lands heads. So there's a 50% chance the coin lands tails if it first lands heads.

[9] Cf. Bennett (2003, §12) and Willer (2010).

[10] Though see, e.g,. McGee (2000); Kaufmann (2004); Rothschild (2013); Khoo (2016) for some putative counterexamples.

b. ?? There's a 25% chance the coin first lands heads and then lands tails, and a 50% chance the coin first lands heads. But I'm not sure what the chance is that the coin lands tails if it first lands heads.

(2) a. ✓ Although I'm not certain whether Jim or Jane will toss the coin, I'm certain that the coin will land either heads or tails. So I'm certain that if Jim flips the coin, it will land either heads or tails.

b. ?? Although I'm not certain whether Jim or Jane will toss the coin, I'm certain that the coin will land either heads or tails. But I can't say I'm certain that if Jim flips the coin, it will land either heads or tails.

If RT is true, it is easy to explain why the (a) sentences seem good and the (b) sentences seem bad. Without it, it's less clear.

Those are the standard arguments for RT. Now for the caveat about its standing in the literature. A series of results has convinced many theorists that, on pains of trivializing the semantics of the indicative conditional, RT cannot be true in full generality.[11] Interestingly, the existence of these results hasn't undermined support for the thesis *all* that much—at least not in the sense relevant to (e.g.) the quotation from Williams above—though it has forced theorists to clarify what exactly their commitment to RT amounts to. For example, some react to the results by denying that conditionals express propositions. They interpret the '*P*' that appears in the statement of RT not as meaning *probability of truth*, but instead as meaning something along the lines of *degree of assertability* or *degree of believability*.[12] Others leave the interpretation of RT basically as is, but making indicative conditionals highly context-sensitive and substantially weakening their logic.[13]

The dialectic surrounding the triviality results for Ramsey's thesis is far too complicated to be covered in any real detail here. But readers should take comfort knowing that how one lands on the triviality results makes no difference to the central results of this paper. If we reject introspection principles for knowledge and belief, then *regardless* of our stance on whether conditionals express propositions, or whether they are context-sensitive (and if so, in what ways), or which logical principles for the indicative are correct, or whether by 'probability' we mean *degree of assertability/believability*, etc.—regardless of any of that—we should reject the idea that the probability of the conditional is the conditional probability of its consequent given its antecedent. All that matters is that it is possible to talk meaningfully about the probability of a conditional, or of a person's believing or knowing a conditional. So long as one accepts that much—and I know of no one who takes the triviality results to show otherwise—then there is a theoretical question about the status of the inference from ⌜S knows/believes that the probability of q conditional on p is x⌝ to ⌜S knows/believes that the probability that if p, q is x⌝. And

[11] The first of these triviality results was proved by Lewis (1976). Since then there have been numerous extensions. See Hájek and Hall (1994), Edgington (1995), and Bennett (2003) for helpful surveys of the results.
[12] Cf. Gibbard (1981); Edgington (1995).
[13] Cf. Bacon (2015).

it is of some interest to know that, given modest assumptions, the status of the inference hangs on the status of introspection principles.

With these points in mind, the rest of the paper will ignore the triviality results and work on the assumption that conditionals express non-trivial propositions. The paper's arguments do not require this assumption, but they are significantly easier to comprehend if it is granted.

4 Qualitative Ramsey's thesis

Having covered introspection principles for knowledge/belief and Ramsey's thesis (both repeated here)

PI If $\Box p$, then $\Box\Box p$.

NI If $\neg\Box p$, then $\Box\neg\Box p$.

RT If $P(p) > 0$, then: $P(p \rightarrow q) = P(q|p)$.

we may now turn to arguing for the first main conclusion of this paper: that those who accept RT must accept at least one of PI and NI.

At this point it may seem less than entirely clear how RT, a principle about indicative conditionals and probabilities, could have anything to do with PI or NI, principles about the introspectiveness of knowledge/belief. We're going to get there by invoking two auxiliary principles, the first of which is the topic of this section:

Qualitative RT (QRT)
If $\Diamond p \wedge \Box(p \supset q)$, then: $\Box(p \rightarrow q)$.

What QRT says, in words, is that if you know/believe that $\neg p$ or q, and for all you know/believe p, then you know/believe that if p, q. This section defends the view that anyone who accepts RT should accept QRT. The next section defends the view that anyone who accepts QRT should accept at least one of PI or NI. Taken together, we get the connection between Ramsey's thesis and introspection principles for knowledge and belief.

But first things first: why should those who accept RT also accept QRT? Here are two reasons.

First, RT entails QRT given the principle that a proposition has probability 1 if and only if it is known/believed to be true, and has probability greater than 0 if and only if its negation is not known/believed to be true—i.e.:[14,15]

Box Probability
$P(p) = 1$ iff $\Box p$; $P(p) > 0$ iff $\Diamond p$.

[14] *Proof:* By BOX PROBABILITY, if $\Diamond p \wedge \Box(p \supset q)$, then $P(p) > 0$ and $P(q|p) = 1$. So by RT, $P(p \rightarrow q) = 1$. Thus, by BOX PROBABILITY again, $\Box(p \rightarrow q)$.

[15] Cf. Williamson (2000, ch. 10), Hawthorne (2012), Clarke (2013), and Greco (2015a).

And I contend that proponents of RT should find BOX PROBABILITY plausible on the relevant interpretation of 'P'. Admittedly, there are a number of theoretical contexts in which the interpretation of 'P' should make BOX PROBABILITY seem highly contentious.[16] But when restricted to the kinds relevant to semantic theorizing, BOX PROBABILITY ought to be the presumptive view. Consider the natural interpretations of the probabilistic vocabulary in sentences like:

(3) a. ✓ We know it's raining, so there's no chance it's clear and sunny.

b. ?? We know it's raining, but there's some chance it's clear and sunny.

(4) a. ✓ There is some probability I will win the lottery. So no, I'm not sure I will lose.

b. ?? There is some probability I will win the lottery. But I'm sure I'll lose.

Ordinary intuitions about these sentences strongly suggest that there are natural interpretations of 'P' on which the inference from '$P(p) = 1$' to '$\Box p$' (as well as '$P(p) > 0$' to '$\Diamond p$') is valid. Moreover, these intuitions appear to be of a kind with the standard intuitions in favor of RT ((3b)–(4b) are no less abominable than (e.g.) (1b)–(2b)). Those who accept RT on the basis of natural language data and the like should thus be uncomfortable denying a principle like BOX PROBABILITY, and so by extension should be uncomfortable denying QRT.

Second (and less contentiously), all the considerations that can be raised in favor of RT can just as well be raised in favor of QRT. It is the limiting case of the slogan "The degree to which one should believe a conditional equals the degree to which one should believe its consequent conditional on its antecedent", so to the extent that RT seems like a platitude, QRT ought to as well. Moreover, QRT does just as well on the standard thought experiments. If you know that Jane has either the ace of spades or the ace of hearts (but you don't know which), and I ask you whether you know whether Jane has the ace of spades if she doesn't have the ace of hearts, the obvious answer is "Yes"—just as QRT predicts. Lastly, the natural language data in support of QRT is, as far as I can tell, exactly alike that supporting RT. Witness:

(5) a. ✓ I don't know whether Jim is in the office or not, but I know Jane is. So of course I know whether Jane is in the office if Jim is.

b. ?? I don't know whether Jim is in the office or not, but I know Jane is. But I can't say I know whether Jane is in the office if Jim is.

(6) a. ✓ I'm sure that Lexie will either come to the party or stay home. So I'm sure that she'll stay home if she doesn't come to the party.

b. ?? I'm sure that Lexie will either come to the party or stay home, but I'm not sure whether she'll stay home if she doesn't come to the party.

[16] See, e.g., Dougherty and Rysiew (2009), Fantl and McGrath (2009), Bacon (2014), and Worsnip (2015) for arguments that show that on some natural ways of using and thinking about P, it is problematic to assume that $P(p) = 1$ iff $\Box p$.

If our commitment to the falsity of QRT is enough to convince us that our intuitive judgments about (5)–(6) are mistaken, then it is unclear on what grounds one should say any differently about the judgments that have been take to motivate RT.

I conclude that anyone who wants their theory of the conditional to deliver RT better have it deliver QRT as well. So from here on out we will be treating RT/QRT as a package view: if QRT requires introspection principles for $\Box$, then so does RT.

5 From qualitative Ramsey's thesis to introspection principles

QRT tells us that if we non-trivially know/believe a material conditional, then we know/believe the corresponding indicative conditional. How substantive is this thesis? Well, if the indicative conditional were the material conditional (i.e., $p \to q$ iff $p \supset q$), then QRT (repeated here) would express a tautology.

QUALITATIVE RT

If $\Diamond p \wedge \Box(p \supset q)$, then: $\Box(p \to q)$.

Since we want to eventually derive one of PI or NI from QRT, and since neither PI nor NI is a tautology, we'll need to say something about the indicative that distinguishes it from the material conditional.[17]

As it happens, the only principle needed for this is extremely weak. It is a restricted version of the principle of conditional non-contradiction:

WEAK CNC

If $\Diamond p$, then: $\neg(p \to q \wedge p \to \neg q)$.

The principle of conditional non-contradiction (CNC) says that a conditional ($p \to q$) and its contradictory ($p \to \neg q$) cannot both be true; WEAK CNC says that unless the conditional's antecedent is known/believed to be false, then that conditional and its contradictory cannot both be true. CNC entails WEAK CNC, so anyone who accepts the former should accept the latter. And those who think conditionals with contradictory antecedents present counterexamples to CNC can embrace WEAK CNC without reservation. Indeed, with the exception of the material analysis,

[17] I'll note here that if we were to analyze the indicative conditional as a "strict" conditional:

STRICT

$p \to q$ iff $\Box(p \supset q)$. (Cf. Kratzer (1986), Gillies (2010), Rothschild (2013).)

we could straightforwardly prove that QRT entails PI (see the next paragraph). However, since the argument we are about to make applies to an extremely wide range of views on the indicative, we will not pursue this issue further.

Proof: Suppose (as is standard) that $\Box$ is a serial accessibility relation—i.e., for any proposition p, $\Box p \supset \Diamond p$. Also suppose $\Box q$. Let $\top$ be an arbitrary tautology. It follows from RATIONALITY that $\Box\top \wedge \Box q$, which (given seriality) entails $\Diamond\top \wedge \Box(\top \supset q)$. So by QRT, $\Box(\top \to q)$. By STRICT, $\Box(\Box(\top \supset q))$, which by CLOSURE gets you $\Box\Box\top \supset \Box\Box q$. By RATIONALITY, $\Box\Box\top$. Therefore, $\Box\Box q$.

I know of no truth-conditional theory of the indicative conditional that invalidates WEAK CNC.[18] I will thus assume it without further argument.

With WEAK CNC in place we are now in a position to argue from QRT to the disjunction of PI and NI. The argument has two premises: (i) that if a certain kind of epistemic/doxastic state is possible, then WEAK CNC entails the negation of QRT; and (ii) that if introspection principles for $\Box$ are false, then the relevant epistemic/doxastic state is indeed possible. From (i) and (ii) it follows that that QRT is true only if at least one of PI or NI is.

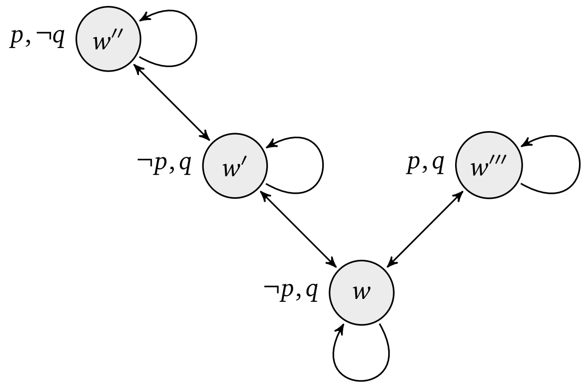
Starting with (i), the kind of epistemic/doxastic state we are looking for is one that contains a pair of worlds w and w' with a certain special property. In particular, we want w to be a world at which some material conditional is non-trivially known (i.e.: at w, for some p and q, $\Diamond p \wedge \Box(p \supset q)$), and we want w' to be a world at which the "reverse" material conditional is non-trivially known (i.e.: at w', $\Diamond p \wedge \Box(p \supset \neg q)$). Then so long as w and w' see each other and themselves—i.e.: for all that is $\Box$-ed at w, w' is actual, and vice-versa—WEAK CNC will entail the falsity of QRT.

Here is a quick proof. Suppose w and w' are two worlds meeting the just outlined stipulations (see w and w' in the figure below for an example). At w we have $\Diamond p$ and $\Box(p \supset q)$, so by QRT we also have $\Box(p \to q)$. Since w sees w' and itself, it follows that at both w and w', $p \to q$. But at w' we have $\Diamond p$ and $\Box(p \supset \neg q)$, so by QRT it follows that at w', $\Box(p \to \neg q)$. And since w' sees w and itself, it follows that at both w and w', $p \to \neg q$. Thus at both w and w': $\Diamond p \wedge (p \to q \wedge p \to \neg q)$. Hence the counterexample to WEAK CNC.

We have premise (i): if a certain kind of epistemic/doxastic state is possible, then WEAK CNC entails the negation of QRT. By way of getting premise (ii)—if introspection principles for $\Box$ are false, then epistemic/doxastic states like this are possible—it will help to get an intuitive feel for what these states look like. So consider the following model of one:[19]

[18] If the indicative were the material conditional, then cases in which p is false but not known/believed to be false would present counterexamples to WEAK CNC.

[19] Arrows track our "seeing" relation. An arrow from one world to the other indicates that, by the lights of what is $\Box$-ed at the first world, the second world is actual. So, for example, since w sees w''' it follows that at w: $\Diamond p$, $\Diamond q$, $\Diamond(p \wedge q)$, etc.



Anti-Ramsey

Here is why **Anti-Ramsey** is a counterexample to QRT of the kind just described. w sees a single p-world (namely w'''), which happens to be a q-world. Since w sees a p-world and every p-world it sees is a q-world, by QRT every world w sees is a $p \to q$ world. That includes itself and w'. However, w' sees a different p-world (namely w''), one that happens to be a $\neg q$-world. Since w' sees a p-world and every p-world it sees is a $\neg q$-world, by QRT every world w' sees is a $p \to q$ world. That includes itself and w. Thus, both w and w' are $p \to q \wedge p \to \neg q$ worlds. But both are also $\Diamond p$-worlds, contradicting WEAK CNC twice over.

Now to make a not so subtle feature of **Anti-Ramsey** explicit: it is a counter-model to both PI and NI. w sees w' and w' sees w'', but w does not see w''. And this is why, contra PI, $\Box(p \supset q)$ can be true at w without being true at every world w sees. Likewise, w sees w' and w sees w''', but w' and w''' do not see each other. And this is why, contra NI, $\neg\Box(p \supset \neg q)$ can be true at w without being true at every world w sees.

That **Anti-Ramsey** is incompatible with both QRT and introspection principles for knowledge/belief is no coincidence. As it turns out, it is a natural model of the kinds of cases that have been countenanced by epistemologists to undermine introspection principles for knowledge (and rational belief). We saw an example of such a case back in §2: Williamson's (2000, ch. 5) **Mr. Magoo**.[20] Here it is again for reference:

Mr. Magoo

Mr. Magoo is an adult with normal perceptual capacities judging the heights of trees at a distance. Given the limitations of his powers of discrimination, Mr. Magoo is such that whenever he judges the height of a tree to be x feet tall (from this distance), the actual height of the tree (in feet) could easily be anywhere between $x - 10$ and $x + 10$. But so long as Mr. Magoo is in normal conditions—which indeed he is—it is guaranteed to fall within

[20] The puzzles concerning unobserved tosses of coins discussed by (e.g.) Bacon (2014); Dorr et al. (2014); Rothschild and Spectre (2016) provide other plausible instances of **Anti-Ramsey**. So do Williamson's (2011, 2014) unmarked clock cases.

that range. The tree Mr. Magoo is currently looking at appears to him to be about 100 feet tall. And in fact it *is* 100 feet tall. But given the limitations of his perceptual capacities, the strongest proposition Mr. Magoo knows about the tree's height is that it falls somewhere between 90 and 110 feet (inclusive).

Letting

$p =$ The tree is either 90 or 111 feet tall.

$q =$ The tree isn't 111 feet tall.

$\Box = K$.

Anti-Ramsey becomes a natural model of Mr. Magoo's epistemic state. Mr. Magoo knows that the tree is between 90 and 110 feet tall, and thus knows that the tree isn't 111 feet tall. But because the proposition that the tree isn't 111 feet tall is at the margins of what he knows, he doesn't know that he knows it. That is to say: for all Mr. Magoo knows, it's compatible what what he knows that the tree is 111 feet tall (hence ¬PI). Likewise, Mr. Magoo doesn't know that the tree is greater than 90 feet tall. But because the proposition that he doesn't know that the tree is greater than 90 feet tall is at the margins of what he *doesn't* know, he doesn't know that he doesn't know it. That is to say: for all Mr. Magoo knows, he knows that the tree is greater than 90 feet tall (hence ¬NI). And this is why Mr. Magoo non-trivially knows the material conditional *If the tree is either 90 or 111 feet tall, it's not 111 feet tall*, while also being such that, for all he knows he non-trivially knows the material conditional *If the tree is either 90 or 111 feet tall, it's 111 feet tall*. Thus: **Mr. Magoo**, one of the standard putative counterexamples to introspection principles for knowledge/belief, is also a putative counterexample to QRT.

Here is another way of seeing the point. Suppose like Mr. Magoo you're looking at a tree that happens to be around 100 feet tall. You know you're in normal conditions, but otherwise have no special information about the tree's height. I ask you "What's the strongest proposition you know about the tree's height"? According to this brand of anti-introspective epistemology, no amount of reflection will deliver the answer to that question. There is *some* such proposition, but you're not in a position to know (or be rationally sure) what it is. And that's because you're just not able to tell the difference between being such that the strongest proposition you know is that the tree is between 90 and 110 feet tall and being such that the strongest proposition you know is that it's between 91 and 111 feet tall. The margins are too small. (*Mutatis mutandis* for the question "What's the strongest proposition you're in a position to rationally believe about the tree's height?".) And if the margins make it such that you know the tree is 90 and 110 feet tall, but for all you know you know the tree is between 91 and 111 tall, then you're automatically in the kind of epistemic/doxastic state of which **Anti-Ramsey** is a model. **Mr. Magoo** provides as much of an argument against QRT as it does PI and NI.

So we have *an* argument for premise (ii), that if introspection principles for $\Box$ are false, then epistemic/doxastic states like **Anti-Ramsey** are possible. It is that that theorists who are attracted to the **Mr. Magoo**-style counterexamples to PI and NI should expect epistemic/doxastic states like **Anti-Ramsey** to be commonplace.

Margin-for-error considerations are among the most popular sources for skepticism about introspection principles (at least in contemporary epistemology), so this result is interesting in its own right. But what about those who reject PI and NI for reasons other than **Mr. Magoo**-style counterexamples? Well, we know that if either PI or NI is true, then epistemic/doxastic states like **Anti-Ramsey** are metaphysically impossible. The question is whether there are any other reasonable constraints on $\Box$ that can be invoked to rule them out.

Since this question invokes a distinction between reasonable and unreasonable constraints on $\Box$, it is not reasonable to expect a *proof* of a negative answer. But we can still give an argument for a negative answer—and in fact a rather simple one. The argument is that **Anti-Ramsey** is consistent with every standard structural constraint on $\Box$ (other than PI and NI, of course). In particular, its existence is consistent with (but does not imply!) each of:

REFLEXIVITY

If $\Box p$, then p.

SERIALITY

If $\Box p$, then $\Diamond p$.

SYMMETRY

If p, then $\Box\Diamond p$.

So without invoking an introspection principle, or a principle that one would accept only if one were already convinced of the truth of introspection principles, it is difficult to see what natural constraint on $\Box$ could be invoked to rule out **Anti-Ramsey**.[21]

This completes the argument: if introspection principles for $\Box$ are false, then epistemic/doxastic states like **Anti-Ramsey** are possible. And if epistemic/doxastic states like **Anti-Ramsey** are possible, then QRT is incompatible with WEAK CNC. Since WEAK CNC is true, QRT is true only if at least one of PI or NI is. And this means that Ramsey's "platitudinous" observation about the indicative conditional smuggles in at least one highly controversial introspection principle for knowledge/belief. Now we turn to the question of what to make of these principles in light of this conclusion.

[21] The following "disjunctive introspection principle" could do it:

PI OR NI

Either if $\Box p$, then p; or if $\neg\Box p$ then $\Box\neg\Box p$.

What PI OR NI says, intuitively, is that no world can be a counterexample to both PI and NI at once. PI OR NI rules out **Anti-Ramsey**, since w and w' are each in violation of both introspection principles. But I know of no principled philosophical position that would deliver PI OR NI while rejecting both PI and NI. The claim is that similar worries will beset other candidate principles.

6 Quasi-Ramsey's thesis

If we think PI and NI are false, then we must reject QRT and, by extension, RT. But of course that's just another way of saying that if we think RT is true, then we must accept at least one of PI or NI. So given the massive support for RT in the literature on conditionals, one might wonder whether we've discovered a new argument for introspection principles for knowledge and belief.

This section and the next will argue against this line of thinking. Though there is a sense in which the discovery of the incompatibility of RT and the denial of introspection principles automatically constitutes a new argument for introspection principles, dialectically the introspection-denier is not really any worse off than they were before. This is because there is a principle in the vicinity of RT that does *not* presuppose introspection principles for $\Box$, and that is capable of accounting for the various intuitive considerations cited in favor of RT. Thus, those who accept RT should do so only because they're antecedently convinced of introspection principles. To argue in the other direction is problematically question-begging.

6.1 The surrogate principle

By way of introducing the surrogate principle that will be of use to the introspection-denier, it is worth making an observation about the evidence that is typically cited in favor of RT. And it is that there is not a single thought experiment in the literature on Ramsey's thesis designed to provide intuitive support for the thesis in which the underlying probabilities are unknown. To give just one representative example, recall Bacon's (2015, p. 132) argument for RT. He points out that when we assess the probability of indicative conditionals like *If the playing card is black, it's spades*, we do so in exactly the way we would were we to calculate the conditional probability of the conditional's consequent given its antecedent, arriving at the answer $\frac{1}{2}$. And I agree that really does seem to be how we go about doing it. But there is an obvious feature of the example that tends to go unnoticed: namely that we *know* that the odds of getting a spade given that we're getting a black card is $\frac{1}{2}$. (In fact, this is plausibly something we know to many orders.) I know of no example unlike Bacon's in this regard.

That the literature's thought experiments have this feature is no coincidence: it is not at all clear that the verdicts of RT are correct for cases in which one is *not* in a position to know the underlying probabilities. Suppose on Jane's evidence $P(q|p) = .6$, but that for all Jane knows/believes it's anywhere between .5 and .7. What should Jane judge the value of $P(p \to q)$ be? The intuitive answer, as far as I can tell, is 'Somewhere between .5 and .7', not '.6'. *We* might want to answer '.6', but that's because we know something Jane doesn't: namely that the probability on her evidence of q given p is .6. It is far from obvious that Jane ought to follow Ramsey's advice to the letter.

What is the upshot? It is that what the intuitive thought experiments directly support is not the principle that the probability of the conditional is equal to the conditional probability; rather, what they directly support is the principle that the probability of the conditional is equal to the *known/believed* conditional probability. So rather than accept RT/QRT (repeated here):

RT

If $P(p) > 0$, then: $P(p \to q) = P(q|p)$.

QUALITATIVE RT

If $\Diamond p \wedge \Box(p \supset q)$, then: $\Box(p \to q)$.

we might instead accept:

QUASI-RT

If $\Box(P(p) > 0 \wedge P(q|p) = x)$, then $P(p \to q) = x$.[22]

QUASI-QRT

If $\Box(\Diamond p \wedge \Box(p \supset q))$, then $\Box(p \to q)$.

At the very least this is what the introspection-denier should do, because unlike QRT, QUASI-QRT does *not* entail the disjunction of PI and NI. Like QRT, QUASI-QRT delivers $\Box(p \to q)$. But unlike QRT, to satisfy the antecedent of QUASI-QRT, it already has to be the case that $\Box\Diamond p$ and $\Box\Box(p \supset q)$. So the introspective work is done before the principle gets going, rather than because of it.

The question, then, is how much of RT/QRT's explanatory power can be captured by QUASI-RT/QUASI-QRT. If enough of it can—which is what this section and the next will argue—then we can be confident that the the intuitive considerations cited in favor of RT do not constitute independent grounds in favor of introspection principles for $\Box$.

6.2 Contradictions, blindspots, and one-off blindspots

The dialectic is about to get quite complicated, so it will help to get a sense of where we're going before we get into the details. To that end, consider the following three kinds of propositions (and examples of sentences that plausibly express them):

(S_0) $p \wedge \neg p$.
'It's raining and it's not raining.'

[22] We will take for granted that $P(p \to q)$ has a value even when there is no x such that $\Box(P(q|p) = x)$. RT tells us that $P(p \to q) = P(q|p)$. The proponent of QUASI-RT could, in principle, stipulate as much. But it would be considerably less ad hoc to derive the value of $P(p \to q)$ from facts about the underlying epistemic/doxastic state. There are natural ways of doing this, but one's view on which is best may depend on one's view on whether conditionals express propositions. Those who think they do can treat $p \to q$ like any other proposition and assign it a value in accordance with a probability measure over the appropriate epistemic/doxastic state—i.e., in accordance with the answer to the question 'What proportion of worlds compatible with $\Box$ are worlds in which $p \to q$?'. Those who deny that conditionals express propositions will likely have to do something else. A natural idea is to derive the value of $P(p \to q)$ from a weighted sum of all the values x such that $\Diamond(P(q|p) = x)$. So, for example, if $P(P(q|p) = .5)) = .5$ and $P(P(q|p) = 1)) = .5$, then $P(p \to q) = .75$.

(S$_1$) $p \wedge \neg\Box p$.
'It's raining but I don't know/believe it's raining.'

(S$_2$) $p \wedge \neg\Box\Box p$.
'It's raining but I might not know/believe it's raining.'

There seems to be something wrong with all of (S$_0$)–(S$_2$). Explaining our judgments about (S$_0$) is easy: it's a contradiction. Explaining our judgments about (S$_1$), an instance of Moore's (1942) paradox, is also fairly easy—we'll get to the standard account in a moment. But as has been argued in the literature on introspection principles, whether explaining our judgments about (S$_2$) is easy or not seems to depend on whether PI is true.[23] If PI is true, it's easy; if not, it's hard. This is not to say it's impossible if PI is false—see §7—but explaining the badness of (S$_2$) is regarded as a non-trivial challenge for the introspection-denier.

This section will argue that the relevant theoretical difference between RT/QUALITATIVE RT (on the one hand) and QUASI RT/QUASI QUALITATIVE RT (on the other) is *which of (S$_0$)–(S$_2$) they associate certain kinds of sentences with*. In particular, sentences like the following:

(7) a. I don't know whether p, but I do know that q. ?? But I don't know whether if p, q.
b. I'm unsure whether p, yet sure that q. ?? But I'm not sure whether if p, q.

(8) a. I don't know whether p, but I do know that q. ?? But I might not know that if p, q.
b. I'm unsure whether p, yet sure that q. ?? But I'm not sure whether I'm sure that if p, q.

We will argue that if RT/QUALITATIVE RT is true, then (7) is like (S$_0$) while (8) is like (S$_1$); while if QUASI RT/QUASI QUALITATIVE RT is true, then (7) is like (S$_1$) while (8) is like (S$_2$). This will in turn situate the dialectic as follows: if there is an argument from the intuitive appeal of RT to introspection principles for $\Box$, it is an argument to the effect that we must treat (7) like (S$_0$) rather than (S$_1$) and (8) like (S$_1$) rather than (S$_2$). What we will argue over the course of this section (and the next) is that we should be skeptical that any such argument can work.

That is the dialectic in abstract. Now to run through it in detail. We'll start with the explanation of the badness of the Moore-paradoxical (S$_1$) ($p \wedge \neg\Box p$). What makes (S$_1$) problematic is not that it can't be true—after all, there are plenty of true propositions that we don't $\Box$. It is that it is a *blindpsot*: the kind of proposition that—though possibly true—is neither $\Box$-able nor assertable.

Why isn't (S$_1$) $\Box$-able? Because on either a knowledge-theoretic or belief-theoretic conception of $\Box$, NO BLINDSPOTS is true:

NO BLINDSPOTS
$\neg\Box(p \wedge \neg\Box p)$.

[23] Cf. Cohen and Comesaña (2013); Greco (2014a, 2015b); Dorst (2019a).

On a knowledge-theoretic interpretation of $\Box$, NO BLINDSPOTS is a straightforward consequence of REFLEXIVITY, the principle that $\Box p$ entails p.[24] And on a belief-theoretic interpretation of $\Box$, one only has to accept two principles connecting (rational) belief to knowledge:

- **K ENTAILS B**

  If Kp, then Bp.

- **NO CERTAIN IGNORANCE**

  If $B\neg Kp$, then $\neg Bp$.[25]

  (In words: If one is (rationally) sure that one doesn't know that p, then one cannot be (rationally) sure that p.)

NO BLINDSPOTS then falls out straightforwardly.[26]

Why isn't (S_1) assertable? Because there is a $\Box$ norm on assertion:

ASSERTION

If p is assertable, then $\Box p$.

This is not the place to argue for ASSERTION at length.[27] It suffices to observe its usefulness in explaining the unassertability of sentences like (S_1), which is something both introspection-affirmers and introspection-deniers need. If by NO BLINDSPOTS one cannot know/believe propositions of the form $p \wedge \neg\Box p$, then by ASSERTION one cannot assert them either.

So far so good: (S_0) is a contradiction, (S_1) is a blindspot. But what about (S_2) ($p \wedge \neg\Box\Box p$)? It's obviously not a contradiction. Whether it is a blindspot depends on whether PI is true or not. If PI is true, then it is a blindspot.[28] But if PI is false, then it is merely a *one-off blindspot*: the kind of proposition that can be $\Box$-ed (and is thus assertable), but cannot be $\Box\Box$-ed (and thus cannot be $\Box$-ed to be assertable).[29]

We know why blindspots seem bad, and so know why (S_2) seems bad conditional on PI. But if PI is false and (S_2) is merely a one-off blindspot, it is less clear whether we have a compelling account of its badness. It just seems very difficult to accept that sentences like the following could be known, rationally believed, or felicitously asserted:

(9) ?? I'm reasonably sure I don't know it's raining, but it's raining.

(10) ?? I might not be sure Jane's home, but she's home.

[24] *Proof*: Given CLOSURE, $K(p \wedge \neg Kp)$ entails $Kp \wedge K\neg Kp$. And given REFLEXIVITY, $Kp \wedge K\neg Kp$ entails the contradictory $Kp \wedge \neg Kp$.

[25] Cf. Williamson (2000); Adler (2002); Bergmann (2005); Gibbons (2006); Huemer (2011); Horowitz (2014).

[26] *Proof*: Given CLOSURE, $B(p \wedge \neg Bp)$ entails $Bp \wedge B\neg Bp$. By K ENTAILS B (plus RATIONALITY), this gives you $Bp \wedge B\neg Kp$. And from NO CERTAIN IGNORANCE this entails the contradictory $Bp \wedge \neg Bp$.

[27] Cf. Unger (1975); Williamson (2000); DeRose (2002).

[28] *Proof*: Given PI, $\neg\Box\Box p$ entails $\neg\Box p$. So $p \wedge \neg\Box\Box p$ entails $p \wedge \neg\Box p$, which is a blindspot.

[29] The argument for $\neg\Box\Box(p \wedge \neg\Box\Box p)$ is exactly analogous to the argument for NO BLINDSPOTS. If $\Box = K$, it's in violation of the factivity of knowledge; if $\Box = B$, it's in violation of the conjunction of K ENTAILS B and NO CERTAIN IGNORANCE.

Yet this is what the PI-denier appears to be committed to. And unless she has some story about why we sometimes conflate the status of being un-$\Box$-able with the status of being un-$\Box\Box$-able, considerations from sentences like (9)–(10) constitute a strong argument in favor of PI.

6.3 Ramsey versus Quasi-Ramsey

We will get back to what the introspection denier should say about one-off blindspots in the next section. For now, having identified three different statuses—contradiction, blindspot, one-off blindspot—we are ready to tie things back to Ramsey's thesis. In order to simplify matters, we will follow the lead of the previous section and focus on the qualitative principles:

QUALITATIVE RT

If $\Diamond p \wedge \Box(p \supset q)$, then $\Box(p \to q)$.

QUASI-QRT

If $\Box(\Diamond p \wedge \Box(p \supset q))$, then $\Box(p \to q)$.

According to QRT, non-trivially $\Box$-ing a material conditional suffices for $\Box$-ing the indicative. By contrast, QUASI-QRT says that only $\Box$-ing that one non-trivially $\Box$'s the material conditional suffices for $\Box$-ing the indicative. This means the kinds of cases that can distinguish the views are those where, for some natural number n, $\Box^n(\Diamond p \wedge \Box(p \supset q))$, but $\neg\Box^{n+1}(\Diamond p \wedge \Box(p \supset q))$. For given QRT, you'll be such that $\Box^{n+1}(p \to q)$; but given QUASI-QRT, you'll merely be such that $\Box^n(p \to q)$.

With this pint in mind, consider the following two propositions:

(A) $\Diamond p \wedge \Box(p \supset q) \wedge \neg\Box(p \to q)$.

(B) $\Diamond p \wedge \Box(p \supset q) \wedge \neg\Box\Box(p \to q)$.

If QRT is true, then (A) is a contradiction, while (B) is a blindspot.[30] By contrast, if QUASI-QRT is true (and QRT isn't), then (A) is a blindspot, while (B) is a one-off blindspot.[31] So now consider some sentences that express (A)- and (B)-like propositions:

(A) (11) I don't know whether Jim is in the office or not, though I know Jane is. ?? But it's not the case that I know whether Jane is in the office if Jim is.

(12) I'm sure Lexie's either coming to the party or staying home. ?? But I'm not sure whether she's staying home if she's not coming to the party.

[30] Why is (B) a blindspot given QRT? Because then $\Diamond p \wedge \Box(p \supset q) \wedge \neg\Box\Box(p \to q)$ entails $\Box(p \to q) \wedge \neg\Box\Box(p \to q)$—which is a proposition of the form $p \wedge \neg\Box p$.

[31] Why, given QUASI-QRT, is (A) a blindspot and (B) a one-off blindspot? Because given QUASI-QRT, $\neg\Box^n(p \to q)$ entails $\neg\Box^n(\Diamond p \wedge \Box(p \supset q))$. One can thus apply the same reasoning as in the previous footnote to derive the conclusion that (A) is not $\Box$-able and that (B) is not $\Box\Box$-able.

(B) (13) I don't know whether Jim is in the office or not, though I know Jane is. ?? But I don't think I know whether Jane is in the office if Jim is.

(14) I'm sure Lexie's either coming to the party or staying home. ?? But I might not be sure whether she's staying home if she's not coming to the party.

I take it to be a datum that (11)–(14) all seem quite bad. QRT offers a nice explanation of this datum: (11)–(12) express contradictions, while (13)–(14) express blindspots. And we know why both of those would seem bad. But given, QUASI-QRT, (11)–(12) express blindspots, while (13)–(14) express one-off blindspots. And though we know why sentences that express blindspots would seem bad, we still don't know why sentences that express one-off blindspots would.[32]

This, I claim, is the real challenge for the proponent of QUASI-RT/QUASI-QRT. It's not that she cannot explain our intuitions about thought experiments involving cards, bets, etc.—those tend to confirm QUASI-RT/QUASI-QRT as much as they confirm RT/QRT. Nor is it that the thought behind RT/QRT is too platitudinous to be denied—it turns out it is as controversial as introspection principles for $\Box$. It is that without RT/QRT, the propositions expressed by sentences like (13)–(14) are one-off blindspots, and are thus in principle knowable and assertable.

Before we try to answer this challenge on behalf of the introspection-denier, it is worth pausing to explain how it affects the dialectic between Ramsey's thesis and introspection principles for knowledge/belief more broadly. We know from §§4–5 that we should accept RT only if we are willing to accept at least one of PI and NI. What we want to know is whether given this fact, the evidence that has been cited in favor of RT (and QRT) can be used in a non-question-begging way to make a case for the disjunction of PI and NI. The existence of QUASI-RT (and QUASI-QRT) means that the only kind of evidence that could do this is evidence that is predicted by RT but not QUASI-RT. And there the only real candidate is our intuitive judgments about sentences like (13)–(14).

So let us suppose that the proponent of QUASI-RT *doesn't* have a neat explanation of our judgments about (13)–(14). Would we then have a new, RT-based argument for introspection principles for knowledge/belief? Well, probably not. As we saw with sentences like (9)–(10),

(9) ?? I'm reasonably sure I don't know it's raining, but it's raining.

(10) ?? I might not be sure Jane's home, but she's home.

the problems raised by one-off blindspots for the PI-denier are quite general. So to claim that considerations from RT/QRT provide new evidence in favor of introspection principles for knowledge/belief, it better be that there is something the PI-denier can say to explain the badness of

[32] One might worry that *embedded* instances of sentences like (11)–(12)—say in conditionals or modals—are no better than their unembedded counterparts, and thus that QRT's contradiction-theoretic explanation of their badness is needed after all. Two points in response. First, when I try to embed (11) or (12) in the antecedent of a conditional, the sentence I get seems bad because of processing difficulties, not because it sounds like I'm trying to embed a contradiction. Second (and relatedly), (13)–(14) seem just as bad when embedded in these environments, yet not even the proponent of QRT should want to classify them as contradictions.

(9)–(10) that she cannot also say to explain the badness (13)–(14). But it's hard to see how this could be. As far as I can tell, the only interesting difference between (9)–(10) and (13)–(14) is that the latter two contain indicative conditionals. There is thus a sense in which *no matter how we land on the question* "Can we explain the badness of (13)–(14) without recourse to RT/QUALITATIVE RT?", we should regard arguments for introspection principles from Ramsey's thesis dialectically inert.

Still though, it would be interesting if the introspection-denier had something plausible to say about one-off blindspots in general. So we will close the paper by outlining an introspection-neutral treatment of the badness of one-off blindspots like (9)–(10)—one that can easily be extended to cover the problematic seeming (13)–(14).

7 Iterated knowledge/belief norms

The introspection-denier needs to explain why sentences that express one-off blindspots seem abominable given that they are (apparently) both knowable and assertable. To that end, consider the following meta-normative principle:[33]

NORM SAFETY

If there is something wrong with X-ing when $\neg\phi$, then there is something wrong with X-ing when $\neg\Box\phi$.

Here 'X' stands for an action or state, while 'ϕ' stands for a condition. What NORM SAFETY says, in words, is that if there is a ϕ-norm on X-ing, then derivatively there is a *knowing/believing-that-ϕ* norm on X-ing too.

NORM SAFETY is an expression of the platitudinous thought that actions that risk bad outcomes inherit some of the badness of those outcomes. If we think it's bad to serve drinks laced with poison, we should also expect it to be bad to serve drinks when for all one is sure of, they're laced with poison. Likewise, if we think it is bad to try to pronounce a marriage when one lacks the authority to do so, we should also expect it to be bad to try to pronounce a marriage when one doesn't know one has the authority to do so. NORM SAFETY is the general principle underlying these sorts of judgments.

Supposing NORM SAFETY is true, it follows that norms like ASSERTION iterate *by themselves*. ASSERTION says that you ought not assert p unless $\Box p$. So by NORM SAFETY, there's a norm derivative on ASSERTION that says that you ought not assert p unless $\Box\Box p$. But since there's a norm that says that you ought not assert p unless $\Box\Box p$, by NORM SAFETY there's a norm derivative on *that* one that says that you ought not assert p unless $\Box\Box\Box p$. And so on. Thus, even if PI is false, we should still expect that, for any natural number n, there is a norm that says there is something problematic about an assertion that expresses a proposition of the form $\neg\Box^n p$.

[33] The discussion of NORM SAFETY borrows heavily from Williamson (2005, pp. 229–235). Similar ideas can also be found in Williamson (2011, 2013, 2014); Benton (2013).

Similarly, suppose we think there is a knowledge norm on (full, rational) belief along the lines of:

BELIEF NORM

If Bp, then it ought to be that Kp.

It can then be shown that on either a knowledge- or belief-theoretic understanding of $\Box$, it follows from NORM SAFETY (plus BELIEF NORM) that, for any natural number n, there is something problematic about being such that $\Box(p \wedge \neg\Box^n p)$.[34]

So why are one-off blindspots problematic? Because they are not $\Box\Box$-able. They are therefore guaranteed to be in violation of the (NORM SAFETY-generated) second-order norms on ASSERTION and BELIEF NORM. It is thus unsurprising that they seem abominable, even if they are in principle $\Box$-able. By extension, the proponent of QUASI-RT/QUASI-QRT is in a position to answer the challenge raised by (13)–(14):

(13) I don't know whether Jim is in the office or not, though I know Jane is. ?? But I don't think I know whether Jane is in the office if Jim is.

(14) I'm sure Lexie's either coming to the party or staying home. ?? But I might not be sure whether she's staying home if she's not coming to the party.

On her view these express propositions that are knowable/believable, but not knowably knowable or believably believable. There is a certain negative status associated with such propositions—that of being in violation of the second-order norms on ASSERTION and BELIEF—that explains the impression of abominableness.

Two quick objections (with replies) before concluding.

Objection: On many of the standard ways of rejecting PI, just about every contingent proposition p is such that there is some natural number n for which $\neg\Box^n p$. So wouldn't the present view predict that just about *every* assertion ought to sound as bad as an assertion of a one-off blindspot?

Reply: All NORM SAFETY says is if it's wrong to X when $\neg\phi$ then it's wrong to X when $\neg\Box\phi$. It doesn't say that the wrongs must be equally so. And quite plausibly, the badness of risking a certain negative outcome is less than the badness of the outcome itself. We should thus expect the severity of the violations of the norms to decrease as we ascends up the NORM SAFETY-generated hierarchy. If one ought not X without it being such that ϕ, then one ought not X without it being such that $\Box\phi$. But X-ing merely without $\Box\phi$ is better than X-ing without ϕ, and X-ing merely without $\Box\Box\phi$ is even better than that. Although ideally one would X only if

[34] *Proof*: First suppose $\Box = K$. Given K ENTAILS B, if Kp then Bp. And by BELIEF NORM plus NORM SAFETY we have that for any natural number n: if Bp, then it ought to be that $K^n p$. So for any n, one ought not be such that $Kp \wedge \neg K^n p$. Now suppose $\Box = B$. Given BELIEF NORM plus NORM SAFETY we have that for any n: if Bp, then it ought to be that $B^n Kp$. By K ENTAILS B, this is equivalent to the norm: for any n: if Bp, then it ought to be that $B^n p$. So for any n, one ought not be such that $Bp \wedge \neg B^n p$.

$\Box^{\omega}\phi$, one can take comfort in the fact that as n approaches infinity, the severity of one's failure to abide by the relevant nth-order norm will approach a lower bound.

Objection: If the violation of a derivative norm is better than the violation of its primary norm, then haven't we undercut the explanation of the badness of one-off blindspots like (9)–(10) and (13)–(14)?

Reply: Other than that it decreases, we have said nothing about how the force of the NORM SAFETY-generated norms scales as one ascends the hierarchy. So the view's only prediction is that it is better to believe or assert a one-off blindspot than it is to believe or assert a blindspot. That is consistent with both being well past the threshold of abominable. For there to be a real worry here, we'd have to convince ourselves that one-off blindspots are not just abominable, they're *as* abominable as ordinary blindspots. And it is far from clear how we are supposed to go about doing that.

8 Conclusion

We learned in §§4–5 that one should accept RT only if one is prepared to accept at least one of PI or NI. In light of this fact, it is natural to ask whether the intuitive considerations in favor of RT can be leveraged into a new argument for introspection principles for knowledge/belief. §§6–7 presented an extended argument for a negative answer to this question. Most of the intuitive considerations in favor of RT can be captured just as well by QUASI-RT, a principle that does not presuppose the disjunction of PI and NI. And those that cannot obviously be captured just as well (e.g., our intuitive judgments about sentences like (13)–(14)) are instances of a more general problem facing those who deny introspection principles. Moreover, we know from §7 that this general problem admits of principled answer in the form of NORM SAFETY. In short, then, the only reason to accept RT is if one is antecedently convinced of introspection principles for knowledge/belief. Those who are convinced of their truth may continue to use RT in their theories of the indicative conditional, but those who are convinced of their falsity will have to take on QUASI-RT instead. At least they can take comfort knowing that if the switch to QUASI-RT doesn't get them everything they had with RT, it's because they were wrong to even try it in the first place.[35]

[35] The arguments of §§5–7 have interesting implications for Dorst's (2019a) novel argument for KPI. In broad strokes, Dorst's argument is that iff KPI is false are conditionals like the following knowable (and thus assertable):

(15) ?? Even if I don't know it, p.

But, claims Dorst, (15) seems manifestly unknowable, and so KPI must be true.

Unfortunately, I lack the space to give this argument the attention it deserves. So I'll settle for two quick points about how I take the claims of this paper interact with it. First, one of the two main premises in Dorst's argument is:

STABILITY If $P(p) > 0 \wedge P(q) = 1$, then: $P(p \rightarrow q) = 1$.

which is a corollary of RT. But like RT, STABILITY is only as plausible as its qualitative analog:

QUALITATIVE STABILITY If $\Diamond p \wedge \Box q$, then: $\Box(p \rightarrow q)$.

(In fact, the natural language data Dorst cites in favor of STABILITY (pp. 14-15) is really just evidence for QUALITATIVE STABILITY.) But cases like **Anti-Ramsey** raise as many problems for QUALITATIVE STABILITY as they do QUALITATIVE RT. Most importantly, if **Anti-Ramsey** is possible, then QUALITATIVE INSTABILITY is incompatible with the following highly plausible principle about the semantics of the indicative:

LOCALITY If $(\Diamond p \wedge \Box(p \supset q) \wedge \neg p)$, then: $p \rightarrow q$.

(The proof works in basically the same way as §5's proof that **Anti-Ramsey** makes QUALITATIVE RT and WEAK CNC incompatible.) LOCALITY is so-called because it ties that the truth (or degree of assertability/believability) of a conditional at a world to the facts about what's $\Box$-ed at that world when the conditional's antecedent is false. This is not the place to argue for LOCALITY at length. I'll just observe that it is about as difficult to find intuitive counterexamples to it as it is to find counterexamples to WEAK CNC. Those who reject introspection principles thus have strong reason to reject QUALITATIVE STABILITY. By extension, the KPI-denier will have strong reason to reject a crucial premise of Dorst's argument for more basic reasons than that it contributes to a reductio of her position centered on (15).

Second, even supposing Dorst is right that if KPI is false, then conditionals like (15) express propositions that are knowable, it can be shown that they are not knowably knowable (see Dorst's §2). If the propositions they express are not knowably knowable, then they are one-off blindspots. And we just saw that the KPI-denier can give a principled account of the badness of one-off blindspots while accepting their knowability.

References

Adams, E. (1965). The logic of conditionals. *Inquiry: An Interdisciplinary Journal of Philosophy*, 8(1-4):166–197.

Adler, J. (2002). *Belief's Own Ethics*. MIT Press.

Almotahari, M. and Glick, E. (2010). Context, content, and epistemic transparency. *Mind*, 119(476):1067–1086.

Bacon, A. (2014). Giving your knowledge half a chance. *Philosophical Studies*, (2):1–25.

Bacon, A. (2015). Stalnaker's thesis in context. *Review of Symbolic Logic*, 8(1):131–163.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press.

Benton, M. A. (2013). Dubious objections from iterated conjunctions. *Philosophical Studies*, 162(2):355–358.

Bergmann, M. (2005). Defeaters and higher-level requirements. *Philosophical Quarterly*, 55(220):419–436.

Christensen, D. (2010). Rational reflection. *Philosophical Perspectives*, 24(1):121–140.

Clarke, R. (2013). Belief is credence one (in context). *Philosophers' Imprint*, 13:1–18.

Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly*, 49(2):113–24.

Cohen, S. and Comesaña, J. (2013). Williamson on gettier cases and epistemic logic. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(1):15–29.

Cresto, E. (2012). A defense of temperate epistemic transparency. *Journal of Philosophical Logic*, 41(6):923–955.

Das, N. and Salow, B. (2018). Transparency and the kk principle. *Noûs*, 52(1):3–23.

DeRose, K. (2002). Assertion, knowledge, and context. *Philosophical Review*, 111(2):167–203.

Dokic, J. and Égré, P. (2009). Margin for error and the transparency of knowledge. *Synthese*, 166(1):1–20.

Dorr, C., Goodman, J., and Hawthorne, J. (2014). Knowing against the odds. *Philosophical Studies*, 170(2):277–287.

Dorst, K. (2019a). Abominable kk failures. *Mind*, pages 1–29.

Dorst, K. (2019b). Evidence: A guide for the uncertain. *Philosophy and Phenomenological Research*, pages 1–54.

Dorst, K. (2019c). Lockeans maximize expected accuracy. *Mind*, pages 1–37.

Dougherty, T. and Rysiew, P. (2009). Fallibilism, epistemic possibility, and concessive knowledge attributions. *Philosophy and Phenomenological Research*, 78(1):123–132.

Douven, I. and Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, 37(4):711–730.

Edgington, D. (1995). On conditionals. *Mind*, 104(414):235–329.

Elga, A. (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies*, 164(1):127–139.

Fantl, J. and McGrath, M. (2009). *Knowledge in an Uncertain World*. Oxford University Press.

Fernández, J. (2013). *Transparent Minds: A Study of Self-Knowledge*. Oxford University Press.

Gibbard, A. (1981). Two recent theories of conditionals. In Harper, W., Stalnaker, R. C., and Pearce, G., editors, *Ifs*, pages 211–247. Reidel.

Gibbons, J. (2006). Access externalism. *Mind*, 115(457):19–39.

Gillies, A. (2010). Iffiness. *Semantics and Pragmatics*, 3(4):1–42.

Goodman, J. and Salow, B. (2018). Taking a chance on kk. *Philosophical Studies*, 175(1):183–196.

Greco, D. (2014a). Could kk be ok? *Journal of Philosophy*, 111(4):169–197.

Greco, D. (2014b). Iteration and fragmentation. *Philosophy and Phenomenological Research*, 88(1):656–673.

Greco, D. (2015a). How i learned to stop worrying and love probability 1. *Philosophical Perspectives*, 29(1):179–201.

Greco, D. (2015b). Iteration principles in epistemology i: Arguments for. *Philosophy Compass*, 10(11):754–764.

Greco, D. (2015c). Iteration principles in epistemology ii: Arguments against. *Philosophy Compass*, 10(11):765–771.

Hájek, A. and Hall, N. (1994). The hypothesis of the conditional construal of conditional probability. In Eells, E., Skyrms, B., and Adams, E. W., editors, *Probability and Conditionals: Belief Revision and Rational Decision*, page 75. Cambridge University Press.

Hawthorne, J. (2012). Knowledge and epistemic necessity. *Philosophical Studies*, 158(3):493–501.

Hawthorne, J. and Magidor, O. (2009). Assertion, context, and epistemic accessibility. *Mind*, 118(470):377–397.

Hawthorne, J. and Magidor, O. (2010). Assertion and epistemic opacity. *Mind*, 119(476):1087–1105.

Hawthorne, J., Rothschild, D., and Spectre, L. (2016). Belief is weak. *Philosophical Studies*, 173(5):1393–1404.

Hazlett, A. (2012). Higher-order epistemic attitudes and intellectual humility. *Episteme*, 9(3):205–223.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca: Cornell University Press.

Horowitz, S. (2014). Epistemic akrasia. *Noûs*, 48(4):718–744.

Huemer, M. (2011). The puzzle of metacoherence. *Philosophy and Phenomenological Research*, 82(1):1–21.

Kaufmann, S. (2004). Conditioning against the grain. *Journal of Philosophical Logic*, 33(6):583–606.

Khoo, J. (2016). Probabilities of conditionals in context. *Linguistics and Philosophy*, 39(1):1–43.

Kratzer, A. (1986). Conditionals. *Chicago Linguistics Society*, 22(2):1–15.

Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives*, 24(1):1–21.

Lasonen-Aarnio, M. (2014). Higherorder evidence and the limits of defeat. *Philosophy and Phenomenological Research*, 88(2):314–345.

Lasonen-Aarnio, M. (2019). Enkrasia or evidentialism? learning to love mismatch. *Philosophical Studies*, pages 1–36.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85(3):297–315.

McGee, V. (2000). To tell the truth about conditionals. *Analysis*, 60(1):107–111.

Mchugh, C. (2010). Self-knowledge and the kk principle. *Synthese*, 173(3):231–257.

Moore, G. E. (1942). A reply to my critics. In Schilpp, P. A., editor, *The Philosophy of G. E. Moore*. Open Court.

Ramsey, F. P. (1931). *The Foundations of Mathematics and Other Logical Essays*. Paterson, N.J., Littlefield, Adams.

Rothschild, D. (2013). Do indicative conditionals express propositions? *Noûs*, 47(1):49–68.

Rothschild, D. and Spectre, L. (2016). A puzzle about knowing conditionals. *Noûs*.

Smithies, D. (2012). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research*, 85(2):273–300.

Stalnaker, R. C. (1968). A theory of conditionals. *Americal Philosophical Quarterly*, pages 98–112.

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of Science*, 37(1):64–80.

Stalnaker, R. C. (2009). On hawthorne and magidor on assertion, context, and epistemic accessibility. *Mind*, 118(470):399–409.

Titelbaum, M. G. (2015). Rationality's fixed point (or: In defence of right reason). In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology 5*, pages 253–294. Oxford Scholarship Online.

Unger, P. K. (1975). *Ignorance: A Case for Scepticism*. Number 105. Oxford University Press.

Willer, M. (2010). New surprises for the ramsey test. *Synthese*, 176(2):291–309.

Williams, R. (2009). Vagueness, conditionals and probability. *Erkenntnis*, 70(2):151–171.

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

Williamson, T. (2005). Contextualism, subject-sensitive invariantism and knowledge of knowledge. *The Philosophical Quarterly*, 55(219):213–235.

Williamson, T. (2011). Improbable knowing. In Dougherty, T., editor, *Evidentialism and its Discontents*, pages 147–64. Oxford University Press.

Williamson, T. (2013). Response to cohen, comesaña, goodman, nagel, and weatherson on gettier cases in epistemic logic. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(1):77–96.

Williamson, T. (2014). Very improbable knowing. *Erkenntnis*, 79(5):971–999.

Worsnip, A. (2015). Possibly false knowledge. *Journal of Philosophy*, 112(5):225–246.