BPhil in Philosophy: Thesis

Candidate number: 1036433

Thesis Title: Iteration and Preservation

Wordcount: 29,946

Abstract

According to epistemic iteration principles, epistemic states such as knowledge, belief, or certainty iterate. The KK principle, for example, states that whenever one knows something, one knows that one knows it.

According to epistemic preservation principles, epistemic states such knowledge, belief, or certainty are preserved under certain kinds of revisions. Knowledge preservation, for example, says that if one knows p and does not know that not q, then one knows p after revising one's knowledge with q.

My thesis explores connections between epistemic iteration and preservation principles. The big picture is that iteration and preservation principles are more closely tied together than is generally acknowledged. Denying iteration principles gives us surprising reasons to deny preservation principles. Conversely, accepting preservation principles may give us surprising reasons to accept iteration principles. And some standard arguments for and against iteration principles generalise to arguments for and against preservation principles.

Contents

1 Introduction

1.1 Quick overview

According to epistemic iteration principles, epistemic states such as knowledge, belief, or certainty iterate. The KK principle, for example, states that whenever one knows something, one knows that one knows it. Iteration principles have been discussed in connection to a wide range of issues, including scepticism, peer disagreement, akrasia, internalism vs. externalism, common knowledge, and debates in meta-epistemology.[1]

According to epistemic preservation principles, epistemic states such knowledge, belief, or justification are preserved under certain kinds of revisions. Knowledge preservation, for example, says that if one knows p and does not know that not q, then one knows p after revising one's knowledge with q. Preservation principles have also been discussed in connection to a wide range of topics, including belief revision, the Ramsey test, nonmonotonic consequence, the semantics of conditionals, and knowledge and chance.[2]

1 Cf. Greco (2015a,b) for an overview on epistemic iteration principles. On iteration principles and akrasia see e.g. Horowitz (2014), Lasonen-Aarnio (2010, 2014), Salow (2019), Williamson (2011), on evidence of evidence and peer disagreement see Christensen (2010), Dorst (2020, forthcoming), Williamson (2019), on KK and scepticism see Adler (1981), Stroud (1984), Hall (1976), Williamson (2000), on KK and internalism vs. externalism see Bird & Pettigrew (2019), Dretske (2004), Stalnaker (2015), on KK and common knowledge see Greco (2014a), Lederman (2018), Stalnaker (2002), on level-bridging and meta-epistemology see Greco (2015c, 2019).

2 On preservation, knowledge, and chance cf. Bacon (2014), Dorr et al. (2014), Goodman & Salow (2018), Rothschild & Spectre (2018b); on preservation and belief revision, cf. Alchourrón et al. (1985), Bradley (2012, 2007, 2017b), Chandler (2017), Fermé & Hansson (2018: 44ff.), Lin (2019); on preservation and the Ramsey test, cf. Fuhrmann (1989), Fuhrmann & Levi (1994), Gärdenfors

My thesis explores connections between preservation and iteration principles in epistemology. The papers stand by themselves, and are not written as to form an overarching super-theory. However, if I had to identify a big picture point, it would be that iteration and preservation principles are more closely tied together than is generally acknowledged. Denying iteration principles gives us surprising reasons to deny preservation principles. Conversely, accepting preservation principles may give us new reasons to accept iteration principles. And some standard arguments for and against iteration principles generalise to preservation principles.

Chapter 1 discusses iteration principles for conditional knowledge. I show that given Knowledge Preservation, one can derive KK from a seemingly weak Anti-Moorean constraint on conditional knowledge. One upshot of my result is that given Knowledge Preservation, one needs KK if one wants to align the logic of knowledge with the logic of conditional knowledge. This opens up an interesting abductive argument for KK: One needs KK to preserve harmony between one's logic for knowledge and one's logic for conditional knowledge.

Chapter 2 rebuts the view common in the belief revision literature that preservation is problematic for introspective agents. The view rests on a confusion of an implausible preservation principle for *sentences* with a prima facie plausible preservation principle for *propositions*. Once appropriately understood, preservation turns out unproblematic for *introspective* agents, but problematic for *non-introspective* agents.

Chapters 3 and 4 examine popular arguments for introspection principles from dubious assertions such as "p but I don't know whether I know p." Chapter 3 points out that the KK-based account of dubious conjunctions over-generates, wrongly predicting that "I don't know whether I know p" on its own is unassertable. Chapter 4 shows that the underlying phenomenon is more general, and unrelated

(1986), Gärdenfors (1988), Lindström (1996), Levi (1988), Rabinowicz (1996), Rott (1989, 2017, 2011); on preservation and nonmonotonic consequence, cf. Cross (1990), Koons (2017), Kraus et al. (1990), Lehmann & Magidor (1992), Stalnaker (1994); on preservation and conditionals, cf. Boylan & Schultheis (msb,m), Bradley (2007), Dorst (2019), Holguín (2019), Mandelkern & Khoo (2019), Rothschild & Spectre (2018a).

to knowledge iteration. Whenever q is a defeater for one's knowledge that p, it sounds weird to say "p but it might that q."

Chapter 5 examines a recent argument against iteration principles. If at all, epistemologists usually endorse not KK itself, but weakenings. Liu (2020) and San (2019, ms) challenge this strategy: It seems that any satisfactory weakening of KK must entail KK$^\Diamond$, the claim that knowing entails *possibly* knowing that one knows. But KK$^\Diamond$ can be shown to be just as implausible as KK by Fitch-like arguments from Moorean conjunctions. I generalise the challenge to a wide range of closure, transmission, and preservation principles. This undermines the challenge, as it is implausible to conclude that all these principles are wrong. In fact, well-established weakenings of KK, closure, preservation, and transmission principles are weaker than the modally qualified variants and escape the challenge.

A recurring theme of the thesis is that preservation and iteration principles interact in interesting ways with Anti-Moorean conditions. Preservation principles are often justified by appeal to *informational economy*: Information does not come for free; hence procedures that adjust our beliefs in the light of new information should minimise loss of information (Gärdenfors 1988). On this picture, one should only give up beliefs if information inconsistent with one's beliefs comes up. Stalnaker (2009b: 194) writes that "to fully accept something (to treat it as knowledge) is [...] to continue accepting it unless evidence forces one to give up something." Moorean beliefs give us reason to resist this picture. Some beliefs, such as "p but I do not know that p", are consistent, but still worth avoiding. It may sometimes be better to give up beliefs even if not doing so would lead only to Moorean incoherence, not inconsistency.

There is little need to introduce iteration principles in detail, as they have been discussed prominently in recent epistemology (see the references in fn. 1). However, it may be worth saying a word or two about conditional knowledge. In the remainder of this introduction, I briefly survey work on conditional attitudes, argue that attitude verbs in natural language can express both unconditional and conditional attitudes, and connect conditional attitudes to defeat.

1.2 Conditional Attitudes

Conditional attitudes have been studied in different contexts. The belief revision literature studies operations that take a set of sentences B, thought to represent one's belief state, and a sentence ϕ to a revised belief set $B * \phi$ (Alchourrón et al. 1985). While usually cast in a syntactic setting, equivalent representations in terms of sphere systems, and more generally orderings on worlds are available (Grove 1988, Stalnaker 2009b). There are different interpretations, but the dominant way to think of belief revision frameworks seems to be as capturing *rational belief change* (cf. Lin 2019).

Early discussions of conditional *knowledge* can be found in the context of non-monotonic consequence relations (Kraus et al. 1990, Lehmann & Magidor 1992). These relate closely to belief revision, for a revision operation $*$ for a belief set B corresponds to a non-monotonic consequence relation $\mid\!\sim_B$ through the translation $\phi \in B * \psi$ iff $\psi \mid\!\sim_B \phi$ (Makinson & Gärdenfors 1991). Nonmonotonic consequence relations and belief revision operations are often regarded as two sides of the same coin. The distinction between knowledge and belief is not always clear in this literature; for example talk of conditional knowledge bases is accompanied by talk of belief revision.

More recently, knowledge change has been investigated in dynamic epistemic logic.[3] A difference to belief revision is that dynamic epistemic logic usually studies *dynamic* as opposed to *static* belief change (Ditmarsch 2005). Belief change is static if the objects of belief are assumed not to change, and dynamic if the objects of belief can change. Static belief change is concerned with changing beliefs about a *constant* situation; dynamic belief change with changing beliefs about a *changing* situation (Baltag & Renne 2016).

One can also approach conditional attitudes in a quantitative way. On the Bayesean version, this approach starts with credences, understood as degrees of

3 Cf. Pacuit (2013a,b) for an overview. Cf. van Ditmarsch et al. (2007) for a textbook.

belief, and understands conditional beliefs as conditional credences.[4] Alternatively, one can start with an epistemic probability distribution, such as probability given one's total evidence, and understand conditional knowledge as (epistemic) conditional probability 1. The probabilistic approach can thereby accomodate a number of different conditional attitudes, including knowledge, belief, and being sure.

Broadening our view, the literature on indicative conditionals contains interesting discussions of conditional assertions.[5] Although centred around assertion, a wide range of 'conditional' mental states and speech acts are being discussed: "[a]s well as conditional beliefs, there are conditional desires, hopes, fears, etc.. As well as conditional statements, there are conditional commands, questions, offers, promises, bets, etc." (Edgington 2014). On Edgington's view, to assert (believe, desire, ...) p conditional on q is to full-strength assert p, if in fact q is true, and nothing otherwise. The view has weird consequences. For example, it means one can believe all and only the truths simply by believing for any proposition p both 'If p, then p' and 'If $\neg p$, then $\neg p$'. Unfortunate as it is, there is no such easy guarantee to believing all and only the truths.[6]

Finally, *conditional intentions* have been discussed in law and philosophy of action.[7] Conditional intentions are intentions to do something if something else is the case. Since many criminal offences presuppose an intention, it can make a huge difference wether conditional intentions are intentions (cf. Ferrero 2009). This disucssion was kicked of by the case of R. v. Easom.[8] Easom had picked up a handbag left in a cinema, and checked the contents only to replace the handbag without stealing anything. After being convicted of theft at the first instance, the verdict was quashed on the grounds that conditional intent is insufficient for theft, and Easom only intended to steal if there was something worth stealing. The

4 Cf. Bradley (2017a), Joyce (1999), and work on probabilistic accounts of conditionals (Adams 1965, 1966, 1975).

5 Cf. Carter (forthcoming); Edgington (1995); Goldstein (2019); Mackie (1973: ch.4); Stalnaker (2009a)), and Adams (1965, 1966, 1975).

6 Thanks to [Redacted] for suggesting this worry.

7 Cf. Cartwright (1990), Davidson (1978), Ferrero (2009), Klass (2009), Yaffe (2004).

8 1971, 2 QB 315. My discussion follows the presentation in (Cartwright 1990).

doctrine was applied more widely afterwards, but it was soon recognised to have unacceptable consequences in the case of burglary; burglars *generally* only intend to steal if they find something worth stealing. Courts later decided to water down the earlier decision.

> The legal position now appears to be that a conditional intention will suffice in cases of theft and burglary and attempts to commit these crimes, provided that the indictment is drafted in a particular fashion. (Cartwright 1990: 234)

One interesting moral is that conditional attitudes need not be any less 'real' than unconditional attitudes; both may qualify as full-blown mental states.

This concludes my cursory review of literature on conditional attitudes. In contrast to conditional speech acts and intentions, discussions of conditional belief tend to be technical. This can lead to the impression that conditional belief is a technical concept with little basis in ordinary language. In the next section, I argue that this impression is wrong. Attitude verbs in natural language can express both unconditional and conditional attitudes, depending on context.

1.3 Conditional attitudes in natural language

A well-known phenomenon in natural language is that modals, adverbs, and perhaps other operators can be restricted by if-clauses (Lewis 1975, Kratzer 1991a). One way to bring this out is via seeming failures of modus tollens (Yalcin 2012): Imagine we have no idea what Sara ate for dinner, but I say:

(1) a. It's not the case that Sara must have had fish.
b. If Sara had fish and chips, she must have had fish.
c. ∴ Sara didn't have fish and chips.

The reasoning is fishy. The folklore reaction is to assume that 'must' in (1b) is restricted by the if-clause, i. e. it quantifies only over those epistemic possibilities in which Sara had fish and chips. By contrast, 'must' in (1a) and (1c) quantifies over

all epistemic possibilities, and therefore the inference equivocates. So if-clauses can restrict epistemic modals.[9] Whether restriction is the *only* role of if-clauses is controversial. One strong argument against this is that similar restriction effects with disjunctions:

(2) a. Either Sara is inside, or she must be outside.
b. ≈ Either Sara is inside, or it follows from the relevant information plus the fact that she is not inside that she is outside.

Disjunctions are just one example of many showing that restricted readings are not specific to conditionals, but appears with a wide range of constructions. Various authors have argued that this is evidence that epistemic modals are systematically restricted by locally available information.[10]

What does this all have to do with conditional attitudes? Here is the catch: Not just epistemic modals, but also attitude verbs can be restricted by if-clauses, left disjuncts, or more generally locally available information (Blumberg & Holguín 2019, Dietz ms, Jerzak 2019). Imagine we're thinking about buying a phone, but you suspect that it's broken. I argue:

(3) a. It is not the case that I want to buy the phone.
b. If the phone works, I want to buy it.
c. ∴ The phone doesn't work.

My reasoning sounds phoney. This is best explained by the assumption that 'want' in (3a) is unrestricted, whereas 'want' in (3b) is restricted by the if-clause to express preference *on the assumption that the phone is broken*. Blumberg & Holguín (2019) show how one can implement this idea based on a semantics for 'want' proposed by Heim (1992). Take 'S wants ϕ' to be true at w iff for all worlds compatible

9 Cf. Dorr & Hawthorne (2013), Kratzer (1981, 1986), Lewis (1975), Mandelkern (2019a), Silk (2017), Stojnić (2017), Groenendijk et al. (1996), Yalcin (2007, 2010). These systems differ a lot: some are static, others dynamic; some put restriction in the truth-conditions, others in the presuppositions.

10 Groenendijk et al. (1996), Yalcin (2007), Dorr & Hawthorne (2013), Mandelkern (2019a).

with what S believes at w, S prefers the closest $[\![\phi]\!]$-worlds over the closest $[\![\neg\phi]\!]$-worlds. Restriction of 'want' can then be analysed as restriction of the set of doxastic possibilities that 'want' quantifies over.

To convince the reader that this is a robust phenomenon, let me list a few more examples (from Blumberg & Holguín 2019). Imagine you're not sure if and where Bill went on holiday, but you expect him to say goodbye and travel first class. You could then say:

(4) If Bill is on a plane to Cuba, then I'm surprised he didn't say goodbye.

(5) If Bill is on a plane to Cuba, then I suspect that he is travelling first-class.

But of course, whether or not Bill indeed is on a plane to Cuba, you do not know whether he is, and so you can't (unconditionally) be surprised he didn't say goodbye, or suspect he is travelling first class.

Here is another pair, again from Blumberg & Holguín (2019), involving restriction in right disjuncts: You're at a party where you expected to find a lot of people, but to your disappointment there's hardly anyone there. You haven't looked outside yet.

(6) Either a lot of people are outside, or I regret that I didn't bring more friends.

(7) Either a lot of people are outside, or I think I should have stayed home.

Again, since you do not yet know if there are lots of people outside, you may neither (unconditionally) regret not bringing more friends, nor (unconditionally) think that you shouldn't have come.

One can find parallel examples with "know". For example, playing Sudoku I might be unsure whether a 7 or a 9 belongs in a certain square, while knowing that whatever it is, the remaining number belongs in a different square:

(8) If this is a 7, then we know that that is a 9.

(9) Either this is a 7, or we know it is a 9.

But of course we don't, in the unconditional sense, know that a 9 belongs in either of the squares. The most straightforward explanation is that "know", like other attitude verbs, allows restricted readings.

Of course, one can try to explain away the data. One option would be to treat the attitude verbs as wide-scope. For parallel reasons as for modals, wide-scoping will not work in general. For example, it creates non-sense with knowledge-wh or knowing someone. Imagine you're trying to find out a password, and you've reduced the options to 'Language', 'Epistemology', and 'Logic':

(10) If the password starts with an 'E', then I know what it is.

(11) Either the password starts with an 'L', or I know the password.

Wide-scoping yields nonsense here:

(12) *I know that if the password starts with an 'E', what it is.

(13) *I know that either the password starts with an 'L', or the password.

I rehearse further arguments against wide-scoping in a footnote.[11] Drucker (2019) defends a different strategy, arguing that the (unrestricted) attitude ascriptions are true, although the agents in question are not aware that they are. On his view, one can be surprised that Bill didn't say goodbye without even knowing that Bill didn't say goodbye. I think most will be reluctant to endorse such a radical conclusion.[12]

Upshot: Natural language allows one to express belief, surprise, preference, or knowledge *given a certain assumption* by embedding attitude verbs in conditionals

11 (i) Drucker (2019): Wide-scoping fails badly for anaphoric examples: "If Bill is on a plane to Cuba, then that surprises me." (ii) Blumberg & Holguín (2019): Wide-scoping is impossible if the attitude verb appears inside a 'scope island' like a relative clause: "If Bill is on a plane to Cuba, then the person who I think he's travelling with is Mary." (iii) Mandelkern (2018b): Modals embedded under quantifiers in the consequent of conditionals have to be interpreted in situ. The same issue arises for attitude verbs: "If a suit was filed, then most plaintiffs either won or I suspect they got a settlement." Widescoping is impossible because there are plaintiffs only if the suit is filed.

12 Cf. Blumberg & Holguín (2019) for criticism.

and disjunctions. Although 'conditional attitude' and 'conditional knowledge' are technical terms, they describe ordinary mental states.

1.4 Preservation and Defeat

One reason why I am so interested in conditional knowledge is that it allows making issues about *defeat* more precise. Defeat cases are examples where an agent knows, or is justified in believing something, and then subsequently gains counteracting evidence. On defeat-friendly views, sufficiently strong counteracting evidence leads to loss of knowledge or justification.[13] On defeat-scpetical views, one can sometimes retain knowledge or justification even in the face of extremely strong counteracting evidence.[14]

However, it often remains somewhat unclear what exactly counts as a defeater, and when defeat has taken place.[15] To be fair, precise characterisations have been attempted. For example, Pollock & Cruz (1999: 37) famously suggest that if p is a reason for one to believe r, then q is a defeater for this reason iff p and q are jointly consistent, and p is a reason to believe r but $p \wedge q$ is not a reason to believe r. Some undercutting defeaters don't seem to be captured by this definition: q may destroy p's support for r but at the same time q itself is independently a reason for r (Chandler 2013). For example, let q be the proposition that the probability of $\neg p \wedge r$ is high. Of course, this is not the end of the story; one can refine definitions of defeaters in terms of reasons.[16]

A general reason to be avoid accounts cast in terms of reasons is that talk of reasons is not particularly precise. For example, it is often assumed that defeaters divide into *undercutting* and *rebutting* defeaters, where (roughly) rebutting de-

13 Well-known defeat-friendly views include Russell (1912: ch. 13), Chisholm (1982), Nozick (1981), Goldman (1986), Pollock (1986), Bergmann (2006), and Williamson (2000: sect. 9.7 and ch. 10). For overview articles see Grundmann (2011), Moretti & Piazza (2018).

14 Defeat-sceptics include Lasonen-Aarnio (2010, 2014), Baker-Hytch & Benton (2015).

15 A common strategy, pursued by Lasonen-Aarnio (2010, 2014), Baker-Hytch & Benton (2015), is to leave the notion of a defeater imprecise, and simply argue that no knowledge loss need take place in paradigmatic 'defeat cases'.

16 Cf. Casullo (2018), Chandler (2013), Sturgeon (2014) for discussion.

featers for knowledge/justified belief that something is the case are reasons to believe the negation, while undercutting defeaters are merely reasons undermining one's reasons for belief. As Kotzen (2019) points out, this distinction seems to miss that p and q may each individually be reasons to believe r, but their conjunction $p \wedge q$ a reason to believe $\neg r$. Let p the proposition that Sara tells you she couldn't come to your party because of her auntie's funeral, and q be the proposition that Sara tells you that she couldn't come to your party because of her sister's wedding. Individually, p and q are both evidence that Sara would have liked to come (r), but their conjunction $p \wedge q$ is evidence that she didn't want to come ($\neg r$). q seems to be neither a rebutting defeater (because it is not in itself evidence that $\neg r$), nor a merely undercutting defeater (because it does more than just undermining your reasons for believing r). So perhaps the distinction is not exhaustive. In any case, this illustrates how thinking about defeat in terms of reasons can obfuscate important details. Theorising about defeat in terms of conditional credences, as Kotzen (2019) does, or more generally in terms of conditional attitudes is less prone to such oversights.

A simple definition of defeaters in terms of conditional knowledge could be that if one knows p and q is consistent with one's knowledge, then q counts as a defeater for one's knowledge that p iff one does not know p conditional on q. This may not be exactly what people in the defeat literature intended to talk about, but it has the advantage of simplicity, and is backed up by formal models of conditional attitudes. Note that on this definition, defeaters need not be true.

The debate between defeat-friendly and defeat-sceptical views could then be reconstructed as a debate about *Preservation*, the principle that if one knows p and does not know that $\neg q$, one knows p conditional on q. In the following chapters, we will see all kinds of arguments for and against preservation, and interesting connections between preservation and iteration principles. While I will not frame those discussions in terms of defeat, they reveal unnoticed connections between issues about knowledge iteration and defeat.

Whether there is defeat has interesting consequences for the epistemic evaluation of ignorance. If preservation holds, then learning new things never forces one to know less (or at least not on epistemic grounds), and ignorance is thus never epistemically conducive. In the words of Rabelais, "ignorance is the mother of all evil".[17] If defeat is real, on other hand, ignorance is sometimes an epistemic bliss. As Hugo De Groot put it, "ignorance of certain subjects is a great part of wisdom."

17 "L'ignorance est mère de tous les maux." (*Cinquième Livre*).

2 Level-bridging for conditional attitudes

Abstract According to KK, knowing entails knowing that one knows ($K\phi \supset KK\phi$). According to CNI, if one conditionally knows that one doesn't conditionally know, one doesn't conditionally know ($K_\psi \neg K_\psi \phi \supset \neg K_\psi \phi$). Here, I show that the seemingly plausible CNI entails the controversial KK. The result has interesting implications for the logic of knowledge, and related notions such as belief, justification, and being sure. For example, it shows that aligning the logics of knowledge and conditional knowledge requires KK.

2.1 Introduction

Epistemic states can be iterated, yielding different 'epistemic levels'.[1] Just as we know things, we know that we know things, and know that we know that we know things, and so on. Ditto for other epistemic notions such as belief, justification, certainty, and evidence. Epistemic level-bridging principles concern the connections between different epistemic levels. The KK principle, for example, states that whenever one knows something, one knows that one knows it. Level-bridging principles are relevant to a wide range of issues, including scepticism, peer disagreement, akrasia, and internalism vs. externalism.[2]

1 Alston (1980) coined the term 'epistemic levels'.

2 Cf. Greco (2015a,b) for an overview on level-bridging. On level-splitting and akrasia see Horowitz (2014), Lasonen-Aarnio (2010, 2014), Salow (2019), Williamson (2011), on evidence of evidence and peer disagreement see Christensen (2010), Dorst (2020, forthcoming), Williamson (2019), on KK and scepticism see Adler (1981), Stroud (1984), Hall (1976), Williamson (2000), on KK and internalism vs. externalism see Bird & Pettigrew (2019), Dretske (2004), Stalnaker (2015).

A surprising gap in the enormous literature on level-bridging is the study of conditional attitudes.[3] Sometimes one knows, believes, or is sure of one thing only given another thing. To a first approximation, say that one knows ϕ conditional on ψ iff one knows ϕ upon revising one's knowledge with ψ, and likewise for other attitudes. Just as for unconditional attitudes, one can examine level-bridging principles for conditional attitudes. For example, generalising KK to conditional knowledge yields CKK, the principle that conditionally knowing implies conditionally knowing that one conditionally knows. This paper studies level-bridging for conditional attitudes.

Level-bridging principles play out very differently for conditional and unconditional attitudes. In the unconditional case, the negative infallibility thesis NI ($K\neg K\phi \supset \neg K\phi$) is widely accepted, while KK is typically rejected. This paper shows that CNI ($K_\psi\neg K_\psi\phi \supset \neg K_\psi\phi$), the generalisation of NI to conditional knowledge, entails KK. One upshot of my result is that one needs KK if one wants to align the logic of knowledge with the logic of conditional knowledge. Parallel results hold for non-factive notions such as belief, being sure, and justification. Studying level-bridging for conditional attitudes thus promises insights into level-bridging for unconditional attitudes.

Here the plan for the paper: §2.2 sets out the result that CNI entails KK. §2.3 relates syntactic constraints such as KK and CNI to properties of accessibility relations. §2.4 shows that models of conditional attitudes in terms of conditional probability, attitudes to conditionals, or revision operations satisfy my background assumptions. §2.5 discusses the implications of the result, §2.6 concludes.

2.2 The main result

We use a propositional language, extended with a one-place operator K ('one knows that') and a two-place operator K_ψ ('one knows conditional on ψ that').

3 On conditional attitudes in epistemology, cf. Alchourrón et al. (1985), Bradley (2017a: ch.6), Gärdenfors (1988), Joyce (1999: ch.6), Stalnaker (1984: ch.6).

The duals $\neg K \neg$ and $\neg K_\psi \neg$ are abbreviated as M ('one leaves open that') and M_ψ ('one leaves open conditional on ψ that').

A modal logic in the sense of this paper is a set of sentences containing all classical truth-functional tautologies (PC) and closed under modus ponens (MP) and uniform substitution (US). For a given modal logic L, we call ϕ a theorem of L ($\vdash_{\mathrm{L}} \phi$) iff $\phi \in \mathrm{L}$, and we write $\phi_1, ..., \phi_n \vdash_{\mathrm{L}} \psi$ iff $\vdash_{\mathrm{L}} (\phi_1 \wedge ... \wedge \phi_n) \supset \psi$. Where context disambiguates, we use $\vdash$. As our base logic, we assume the smallest modal logic that is *normal*, i.e. closed under the necessitation rules RN_K ($\phi / K\phi$) and RN_{K_ψ} ($\phi / K_\psi \phi$), and containing all instances of K_K ($K(\phi \supset \psi) \supset (K\phi \supset K\psi)$) and K_{K_ψ} ($K_\psi(\phi \supset \chi) \supset (K_\psi \phi \supset K_\psi \chi)$).[4] We call this logic L, and write 'L + X_1 + ... + X_n' for the smallest extension of L containing the axioms $\mathrm{X}_1 + ... + \mathrm{X}_n$. I sloppily speak of principles or axioms to talk about the corresponding *schema*.

My main result concerns the interplay between three principles:

KK. $K\phi \supset KK\phi$

CNI. $K_\psi \neg K_\psi \phi \supset \neg K_\psi \phi$

QT. $M\psi \supset (K_\psi \phi \equiv K(\psi \supset \phi))$

KK is the controversial principle that whenever one knows something, one knows that one knows it.[5] According to CNI, whenever one conditionally knows that one does not conditionally know something, one does not conditionally know it. CNI extends the negative infallibility thesis NI ($K\neg K\phi \supset \neg K\phi$) to conditional knowledge. NI is widely accepted for knowledge, belief, and justification.[6] Last but not least, the qualitative thesis QT says that if one leaves open ψ, knowing

4 While normality is a standard assumption in epistemic logic, its status is somewhat controversial. Some think normality holds in full generality (Stalnaker 1999, 2006), others take it to be a normative (Colyvan 2013) or descriptive idealisation (Yap 2014). Whatever your favourite take on normality in the unconditional case, adopt the same attitude towards normality of conditional knowledge.

5 KK-fans include Das & Salow (2018), Goodman & Salow (2018), Greco (2014a,b, 2017), Stalnaker (1999, 2006, 2009c). Recent rejections of KK include Carter (2019), Dorr et al. (2014), Hawthorne & Magidor (2009, 2010), Williamson (2000, 2011). Cf. Greco (2015a,b) for a helpful overview.

6 Interpreted in terms of knowledge, NI follows directly from the fact that what's known is true ($K\phi \supset \phi$). For defences of NI for belief, cf. Aucher (2014), Lenzen (1979), Rieger (2015), Stalnaker (2006). For defences of NI for justification, cf. Rosenkranz (2018), Smithies (2012: 327).

conditional on ψ that ϕ coincides with knowing the material conditional $\psi \supset \phi$. QT is supposed to capture the intuition that knowing *p or q* is necessary and sufficient for knowing that *if p, q*, provided on leaves open *p*.[7]

We can prove the following surprising result:

Fact 2.1. L + QT + CNI contains KK.

This result is surprising because NI is universally accepted, and CNI seems to be an innocuous extension of NI to conditional knowledge. But given the plausible background condition QT, CNI entails the widely rejected KK principle.

For readability, proofs of fact 2.1 and further facts are relegated to appendix 2.7. Nevertheless, let me informally explain the basic point. To this end, it is important to understand that QT entails that if KK fails, one knows p conditional on $\neg Kp$. To see this, suppose KK fails, so you know p without knowing that you know it ($Kp \wedge \neg KKp$). Then you leave open that you don't know p, so by QT knowing p conditional on $\neg Kp$ coincides with knowing the material conditional $\neg Kp \supset p$. But you *do* know p, and so you also know the material conditional $\neg Kp \supset p$. It follows that you know p conditional on $\neg Kp$.

The second important point is that CNI and QT entail that one does not know p conditional on $\neg Kp$. There are two cases to consider here:

- Case 1: You don't know p.
  Unconditionally, you can't exclude $\neg p$. Supposing $\neg Kp$ does nothing to rule out $\neg p$, for if $\neg p$ then $\neg Kp$. So you don't know p conditional on $\neg Kp$.
- Case 2: You know p.
  Supposing $\neg Kp$ amounts to supposing you are in case 1. You know that if you are in case 1, you you don't know p conditional on $\neg Kp$ (by the reasoning from case 1). So conditional on $\neg Kp$ you know that conditional on $\neg Kp$ you don't know p. By CNI, you don't know p conditional on $\neg Kp$.

7 For more on 'or-to-if', cf. Bennett (2003), Boylan & Schultheis (msb), Cariani (forthcoming), Holguín (2019), Rothschild & Spectre (2018b), Stalnaker (1975).

Upshot: Whether or not you know p, you do not know p conditional on $\neg Kp$. We already established that if KK fails, you know p conditional on $\neg Kp$. So if CNI and QT hold, KK can't fail.

The argument just given is fairly informal, and it as it stands it is not exactly clear what assumptions are needed to make it work. The proof in appendix 2.7 reveals that little is needed: All one needs to establish that CNI entails KK are normality and QT. In particular, although the informal argument above appealed to the factivity of knowledge, it turns out this assumption is not required.

Before we move on, it is worth noting some extensions of my result. First, we do not require full-strength CNI, but only a restricted version RCNI:

Fact 2.2. $\mathrm{L} + \mathrm{QT} + \mathrm{RCNI}$ contains KK.

RCNI. $\neg K_\psi \bot \supset (K_\psi \neg K_\psi \phi \supset \neg K_\psi \phi)$

RCNI is significantly weaker than CNI because it does not preclude the possibility of trivial conditional knowledge. Suppose one can conditionally know a contradiction, e.g. given a contradiction ($K_\bot \bot$). In normal modal logics, one then conditionally knows everything. This is because contradictions classically entail everything, and in normal modal logics conditional knowledge is closed under entailment. But if one conditionally knows everything, this includes ϕ and $\neg K_\bot \phi$, violating CNI ($K_\bot \phi \wedge K_\bot \neg K_\bot \phi$). RCNI avoids this worry, as it is explicitly restricted to non-trivial conditional knowledge. Fact 2.2 tells us that failures of KK lead to failures of RCNI, i.e. non-trivial CNI-failures.

Another interesting extension concerns knowledge given multiple suppositions. To this end, we add an operator $K_{\psi_1,\ldots,\psi_n}$ ('one knows conditional on $\psi_1, \ldots, \psi_n$ that') to our language. You can think of $K_{\psi_1,\ldots,\psi_n}$ as representing what one knows after first supposing ψ_1, then ψ_2, and so on until ψ_n.[8] We assume the smallest normal modal logic over this extended language, and call it L^+. In L^+, extensions of CNI, RCNI and QT to multiply conditional knowledge entail CKK:

8 We leave open whether supposing $\psi_1, \ldots, \psi_n$ one after another is the same as supposing their conjunction $\psi_1 \wedge \ldots \wedge \psi_n$.

Fact 2.3. $\mathrm{L}^{+}+\mathrm{QT}^{+}$ contains RCNI^{+} iff it contains CKK^{+}.

CKK^{+}. $K_{\psi_1,\ldots,\psi_n}\phi \supset K_{\psi_1,\ldots\psi_n}K_{\psi_1,\ldots\psi_n}\phi$ $(n \geq 1)$

RCNI^{+}. $\neg K_{\psi_1,\ldots\psi_n}\bot \supset (K_{\psi_1,\ldots\psi_n}\neg K_{\psi_1,\ldots\psi_n}\phi \supset \neg K_{\psi_1,\ldots\psi_n}\phi)$ $(n \geq 1)$

QT^{+}. $M_{\psi_1,\ldots\psi_n}\phi \supset (K_{\psi_1,\ldots\psi_n,\phi}\chi \equiv K_{\psi_1,\ldots\psi_n}(\phi \supset \chi))$ $(n \geq 1)$

Corollary 2.1. $\mathrm{L}^{+}+\mathrm{QT}^{+}+\mathrm{CNI}^{+}$ contains CKK^{+}.

CNI^{+}. $K_{\psi_1,\ldots\psi_n}\neg K_{\psi_1,\ldots\psi_n}\phi \supset \neg K_{\psi_1,\ldots\psi_n}\phi$ $(n \geq 1)$

If one accepts QT and CNI or RCNI, it is natural to generalise them to multiply conditional knowledge. What the present extension tells us is that this will mean accepting not just KK, but also CKK^{+} — an extension of KK to conditional knowledge given arbitrarily many suppositions.

So far, my discussion has been entirely syntactic in nature, allowing us to get by without an explicit model-theoretic construction of conditional attitudes. However, a purely syntactic perspective misses out on well-known correspondences between level-bridging principles and properties of accessibility relations in Kripke models. The next section adopts a semantic perspective on the results from this section.

2.3 Semantics

This section provides semantic characterisations of CNI, KK, and QT. Standard correspondences between level-bridging principles and constraints on accessibility relations (Lemmon 1977: 52, 67) allow us to prove semantic analogues of facts 2.1, 2.2, and 2.3.

Extended Kripke frames are triples $\langle W, R, \restriction \rangle$ where W is a set of points, informally called worlds, $R \subseteq W \times W$ is an accessibility relation, and $\restriction$ is a function from accessibility relations and sets of worlds to accessibility relations.[9] Our models

9 For frames like this, cf. Boylan & Schultheis (msb) and dynamic epistemic logic (van Ditmarsch et al. 2007).

$\langle W, R, \restriction, V\rangle$ are frames extended with a valuation function $V : At \to \mathscr{P}(W)$ from atoms to subsets of W. Validity ($\vDash$) is truth at all worlds in all models.

We assume the usual semantic clauses for atoms and the connectives, plus

- $[\![K\phi]\!]^w = 1$ iff $R(w) \subseteq [\![\phi]\!]$
- $[\![K_\psi\phi]\!]^w = 1$ iff $R \restriction_{[\![\psi]\!]} (w) \subseteq [\![\phi]\!]$

Here and later, $[\![\phi]\!] = \{w \in W \mid [\![\phi]\!]^w = 1\}$. This simple model of conditional knowledge in terms of Kripke frames is a natural setting for characterising CNI, RCNI, QT, and KK. As always, normality is ensured by the structure of Kripke frames. The characterisation of QT is immediate (cf. Boylan & Schultheis msb):

Fact 2.4. $\langle W, R, \restriction\rangle$ validates QT iff whenever $R(w) \cap p \neq \varnothing$, $R \restriction_p (w) = R(w) \cap p$.

Principles KK, NI, and RNI correspond to properties of accessibility relations in the usual way.[10] A relation R is *transitive* just in case whenever Rwv and Rvu also Rwu (equivalently, $\forall w \in W \forall v \in R(w) : R(v) \subseteq R(w)$). We call R *mildly transitive* just in case for all w there is some $v \in R(w)$ such that w sees everything that v sees ($\forall w \in W \exists v \in R(w) : R(v) \subseteq R(w)$). Finally, we call R *very mildly transitive* just in case any world w that sees some world sees some v such that everything seen by v is seen by w (formally: $\forall w \in W(R(w) \neq \varnothing \Rightarrow \exists v \in R(w) : \ R(v) \subseteq R(w))$). These three properties characterise KK, NI, and RNI:

Fact 2.5. $\langle W, R, \restriction\rangle$ validates KK iff R is transitive, validates NI iff R is mildly transitive, and validates RNI ($\neg K\bot \supset (K\neg K\phi \supset \neg K\phi)$) iff R is very midly transitive.

Full, mild, and very mild transitivity characterise KK, NI, and RNI. Parallel characterisations apply in the case of conditional knowledge:

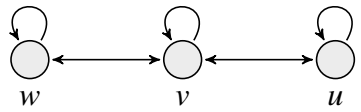
Fact 2.6. $\langle W, R, \restriction\rangle$ validates CKK iff $R \restriction_p$ is transitive for all $p \subseteq W$, validates CNI iff $R \restriction_p$ is mildly transitive for all $p \subseteq W$, and validates RCNI iff $R \restriction_p$ is very mildly transitive for all $p \subseteq W$.

10 See any textbook on modal logic, e. g. Lemmon (1977: 52, 67) or Chellas (1980: 80, 86).

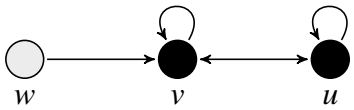
There is a semantic analogue of fact that both RCNI and CNI entail KK in the presence of QT:

Fact 2.7. If $\langle W, R, \restriction \rangle$ validates QT, and $R\restriction_p$ is mildly or at least very mildly transitive for all $p \subseteq W$, then R is transitive.

The proof here is intuitive. If transitivity fails for R, then by choosing the restriction p appropriately, we can always 'zoom in' on the transitivity failure to get violations of mild and very mild transitivity. Suppose R is not transitive, so there are $w, v, u \in W$ with wRv and vRu but not wRu. For example, our frame might look like this:



We choose $p = \{v, u\}$. Since $R(w) \cap p \neq \varnothing$ and $R(v) \cap p \neq \varnothing$, by the validity of QT and fact 2.4 $R\restriction_p(w) = R(w) \cap p$, and $R\restriction_p(v) = R(v) \cap p$. We visualise this in our example by marking p black, and deleting arrows to $\neg p$-worlds:



$R\restriction_p$ is neither mildly nor very mildly transitive, for $R\restriction_p(w) \neq \varnothing$ and for all $x \in R\restriction_p(w)$ there is a $y \in R\restriction_p(x)$ such that $y \notin R\restriction_p(w)$ (the only $x \in R\restriction_p(w)$ is v, and $u \in R\restriction_p(v)$ but $u \notin R\restriction_p(w)$). What looks like a transitivity failure when we look at the whole of R becomes a failure of mild and very mild transitivity when we 'zoom in' on $R\restriction_p$.

A noteworthy class of frames are those where $R\restriction_p = R - \{\langle w, v\rangle \mid v \notin p\}$. One might call these *material conditional frames*, as they are characterised by the principle ID identifying conditional knowledge with knowing the material conditional ($K_\psi \phi \equiv K(\psi \supset \phi)$). One can define a translation τ from our language in the sublanguage without the conditional knowledge operator such that ϕ is true at a world in a material conditional frame just in case its translation $\tau(\phi)$ is true at

the same world in the corresponding Kripke frame.[11] Similarly, one can show that ϕ is a theorem of L + ID + RCNI just in case its translation $\tau(\phi)$ is a theorem of K4. This makes proving a completeness result for the logic L + ID + RCNI fairly straightforward: A sentence is is a theorem of L + ID + RCNI just in case it is valid on all material conditional frames $\langle W, R, \restriction \rangle$ where R_p is very mildly transitive for all $p \subseteq W$.[12] Since we are less interested in technicalities here, I will not discuss questions of completeness further.

To interpret the extended language containing multiply conditional knowledge, we write $R \restriction_{p_1,..,p_n}$ shorthand for $(R \restriction_{p_1}) ... \restriction_{p_n}$:

- $[\![K_{\psi_1,..,\psi_n}\phi]\!]^w = 1$ iff $R \restriction_{[\![\psi_1]\!],...,[\![\psi_n]\!]} (w) \subseteq [\![\phi]\!]$

One can then prove a semantic analogue of fact 2.3 and corollary 2.1:

Fact 2.8. If $\langle W, R, \restriction \rangle$ validates QT^+, then $R \restriction_{p_1,..,p_n}$ is very mildly transitive for all $p_1, .., p_n \subseteq W$ just in case $R \restriction_{p_1,..,p_n}$ is transitive for all $p_1, .., p_n \subseteq W$.

Corollary 2.2. If $\langle W, R, \restriction \rangle$ validates QT^+, then if $R \restriction_{p_1,..,p_n}$ is mildly transitive for all $p_1, .., p_n \subseteq W$ then $R \restriction_{p_1,..,p_n}$ is transitive for all $p_1, .., p_n \subseteq W$.

The models employed in this subsection are impoverished, representing conditional knowledge in terms of accessibility relations. This is convenient because it allows us to use correspondences between level-bridging principles and properties of accessibility relations. But it is unrealistically simple. Plausible models of conditional knowledge build in more structure, such as plausibility orderings,

11 Let τ translate conditional knowledge into knowing the material conditional ($\tau(K_\psi\phi) = K(\tau(\psi) \supset \tau(\phi))$), and leave everything else as it is ($\tau(p) = p$ for atoms p, $\tau(\neg\phi) = \neg\tau(\phi)$, $\tau(\phi \supset \psi) = \tau(\phi) \supset \tau(\psi)$, and $\tau(K\phi) = K\tau(\phi)$).

12 Proof sketch: By the definition of τ, $\vdash_{\mathrm{L+ID+RCNI}} \phi$ just in case $\vdash_{\mathrm{K4}} \tau(\phi)$. By standard completeness results for K4 Lemmon (1977: 54), $\vdash_{\mathrm{K4}} \tau(\phi)$ just in case $\vDash_{\langle W,R\rangle} \tau(\phi)$ for all Kripke frames $\langle W, R\rangle$ where R is transitive. By the definition of τ again, $\vDash_{\langle W,R\rangle} \tau(\phi)$ for all Kripke frames $\langle W, R\rangle$ where R is transitive just in case $\vDash_{\langle W,R,\restriction\rangle} \phi$ for all material conditional frames $\langle W, R, \restriction \rangle$ where R_p is very mildly transitive for all $p \subseteq W$.

probability measures, or selection functions. In the next section, we look at such more interesting models.

2.4 Three models of conditional attitudes

Models can serve epistemologists as a sanity check. If we can construct models incorporating our assumptions, we thereby prove them consistent. Even better, if natural models of a phenomenon incorporate our assumptions, this is evidence that the assumptions are true, or at least benign idealisations.[13] In this spirit, I shall show here that natural models of conditional knowledge and belief incorporate QT and, to some extent, QT^+.

One can think of conditional attitudes in different ways. First, conditional attitudes might encode dispositions to revise one's attitudes: you know conditional on p that q iff you know q upon revising your knowledge with p.[14] Second, conditional attitudes could be attitudes towards a conditional: you know conditional on p that q iff you know that if p, then q. Third, conditional attitudes might be cashed out in terms of conditional probabilities: you know conditional on p that q iff q is certain given p, for some suitable epistemic notion of probability.[15]

How one thinks of conditional knowledge will impact one's choice of model. The first approach, conditional knowledge as revised knowledge, is naturally formalised through AGM revision operations (Alchourrón et al. 1985). The second approach, conditional knowledge as knowing a conditional, can be developed by combining Kripke models with a semantics for the indicative conditional, e.g. in the style of (Lewis 1973, Stalnaker 1968, 1975). The third approach, conditional knowledge as conditional certainty, can be modelled through probability spaces. In this section, I show that standard models of all three kinds incorporate QT.

13 Cf. Williamson (2017) on the advantages of model-building in epistemology.

14 Stalnaker (1984: 103) writes: "[a]n agent's rational dispositions to change what he accepts may be identified with his conditional beliefs, expressed in conditional sentences. To be disposed to accept B on learning A is to accept B conditionally on A, or to accept that if A, then B."

15 See e.g. Joyce (1999) for such an approach to conditional belief.

You might wonder why I don't work with a model which unifies all three approaches. This is because it is not obvious that there is such a model. The claim that the first two approaches align is (in)famous as the Ramsey test, the idea that one knows 'If p, then q' *iff* one knows q after revising one's knowledge with p.[16] The Ramsey test famously leads to triviality results (e.g. Gärdenfors 1986). The claim that the second and third approach align is an instance of Stalnaker's thesis, the idea that the probability of a conditional is the conditional probability. Stalnaker's thesis also figures in triviality results (Lewis 1987: and followers). The first and third approach to conditional knowledge are largely compatible (Gärdenfors 1988: ch. 5). While one can align all three approaches to conditional knowledge if one adopts a form of contextualism about conditionals,[17] I do not want to incur such controversial commitments here and consider them separately.

2.4.1 Probabilistic models

Probability frames are triples $\langle W, P, \restriction \rangle$ where W is a finite set of points, informally called worlds, P is a function from worlds w to probability measures[18] P_w over the subsets of W, and $\restriction$ is a function from a probability measure Pr and a proposition $p \subseteq W$ to another probability measure $Pr \restriction p$, written as $Pr(\cdot \mid p)$.[19] $P_w(\cdot)$ is intended to represent some kind of epistemic probability, say probability given an agent's total evidence.[20] We constrain $\restriction$ by requiring that when $Pr(q) > 0$, $Pr(p \mid q) = Pr(p \wedge q)/Pr(q)$. Our models $\langle W, P, \restriction, V \rangle$ are probability frames extended with a valuation function $V : At \to \mathscr{P}(W)$ from atoms to subsets of W.

16 The name is due to a famous footnote in Ramsey (1931), cf. Stalnaker (1968). Formulations in terms of belief are more common.

17 Lindström (1996) shows that contextualists can retain a version of the Ramsey test that holds context fixed: At a context c where the relevant knowledge is K_c, the proposition $[\![(p \to q)]\!]^c$ is known by K_c iff $K_c * [\![p]\!]^c$ knows $[\![q]\!]^c$. Bacon (2015) nd Mandelkern & Khoo (2019) explain how contextualists can retain versions of Stalnaker's thesis that hold context fixed: At a context c where the relevant epistemic probabilities are P_c, $P_c([\![(p \to q)]\!]^c) = P_c([\![q]\!]^c \mid [\![p]\!]^c)$.

18 That is $P_w(p) \geq 0$ for $p \subseteq W$, $P_w(W) = 1$, $P_w(p \cup q) = P_w(p) + P_w(q)$ for $p \cap q = \varnothing$.

19 Cf. Joyce (1999) for an approach to conditional belief in terms of functions from probability measures and propositions to probability measures.

20 Cf. Williamson (2000), Dorst (forthcoming, 2020) for discussion of epistemic probability.

We assume the usual semantic clauses for atoms, negation, and the material conditional.[21] As usual, $[\![\phi]\!] = \{w \in W \mid [\![\phi]\!]^w = 1\}$. The semantic clauses for knowledge and conditional knowledge are as expected:

- $[\![K\phi]\!]^w = 1$ iff $P_w([\![\phi]\!]) = 1$
- $[\![K_\psi\phi]\!]^w = 1$ iff $P_w([\![\phi]\!] \mid [\![\psi]\!]) = 1$

Normality is directly built into probability frames, and QT also turns out valid:

Fact 2.9. QT is valid on all probability frames.

Probability models similar to these are familiar from some work in economics (Samet 1998, 1997), epistemology (Dorst 2020, Williamson 2000, 2019), and semantics (Yalcin 2010). A small difference is that we assume a conditionalisation operation $\restriction$, rather than defining conditional probabilities via the ratio formula, to avoid undefinedness when conditioning on zero-probability events. We are in good company here, cf. Fraassen (1976), Hájek (2003), Joyce (1999), Popper (1959), Rényi (1970). If you think conditional probabilities for zero-probability events should be undefined, you can assume that $K_\phi\psi$ is trivially true at w if $P_w([\![\phi]\!]) = 0$, and still retain QT. The fact that QT holds in natural probabilistic models is some evidence that QT is an innocuous assumption (but see §2.5 for discussion).

If we want, we can use the probabilistic models employed here also to interpret our extended language containing multiply conditional knowledge. We adopt the convention of writing $P(\cdot \mid p_1, ..., p_n)$ for the probability measure $(P \restriction p_1) ... \restriction p_n$,[22] and give the obvious semantic clause for multiply conditional knowledge:

- $[\![K_{\psi_1,...,\psi_n}\phi]\!]^w = 1$ iff $P_w([\![\phi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!]) = 1$

Given this semantic clause, QT$^+$ is valid.

Fact 2.10. QT$^+$ is valid on all probability frames.

21 $[\![p]\!]^w = 1$ iff $w \in V(p)$, $[\![\neg\phi]\!]^w = 1$ iff $[\![\phi]\!]^w = 0$, and $[\![\phi \supset \psi]\!]^w = 1$ iff $[\![\phi]\!]^w = 0$ or $[\![\psi]\!]^w = 1$.
22 We do not require that $P(\cdot \mid p_1, ..., p_n) = P(\cdot \mid p_1 \cap ... \cap p_n)$.

Note that we did not impose any constraints on $P_w(\cdot)$ beyond requiring it to be a probability measure, in particular no factivity constraint such as $P_w(\{w\}) > 0$. This means we could just as well think of $P_w(\cdot)$ as representing the agent's actual or rational credences at world w. Given that belief or being sure are identified with actual or rational credence 1, we get natural probabilistic models of these states obeying all the assumptions of my proofs, underlining the generality of my result.

One might identify knowledge not with certainty, but with probability above some threshold (cf. Rothschild & Spectre 2018a). In that case, normality will fail. I happily embrace this limitation of my result — I believe that knowledge corresponds to nothing short of evidential probability 1. (In infinite cases, knowledge may require *more* than evidential probability 1 — see Williamson (2007). I ignore infinitary complications here, requiring W to be finite.) Similarly, if belief does not require certainty,[23] normality will fail. I mean to talk about a kind of belief that requires subjective certainty, perhaps more appropriately called full belief.

2.4.2 Knowing conditionals

Instead of modelling conditional knowledge as conditional certainty, you may prefer to think of it as knowing a conditional. To implement this suggestion, we add an indicative conditional '$\rightarrow$' to our language. In terms of the semantics for the conditional, there is some choice. Working with the material conditional trivially validates QT. Interestingly, however, combining (other) standard approaches to conditionals with the so-called indicative constraint also ensures QT.

We work with conditional Kripke frames $\langle W, R, f\rangle$ where W is a finite set of points, informally called worlds, $R \subseteq W \times W$ is an accessibility relation, and $f : \mathscr{P}(W) \times W \rightarrow \mathscr{P}(W)$ is a function that takes a proposition and a world into a set of worlds. (We allow but don't require $f(p,w)$ to be a singleton.) Our models $\langle W, R, f, V\rangle$ combine a conditional Kripke frames with a valuation $V : At \rightarrow \mathscr{P}(W)$. We assume the usual semantic clauses for atoms and connectives, plus

23 Cf. Hawthorne et al. (2016), Dorst (2017), Rothschild (forthcoming), Holguín (ms).

- $[\![K\phi]\!]^w = 1$ iff $R(w) \subseteq [\![\phi]\!]$
- $[\![K_\psi\phi]\!]^w = 1$ iff $[\![K(\psi \to \phi)]\!]^w = 1$
- $[\![(\phi \to \psi)]\!]^w = 1$ iff $f([\![\phi]\!], w) \subseteq [\![\psi]\!]$.

We impose three constraints on f, Success, Weak Centering, and the Indicative Constraint. *Success* requires $f(p,w) \subseteq p$, ensuring the validity of 'If p, then p.'[24] *Weak Centering* says that $w \in f(p,w)$ when $w \in p$, predicting the validity of modus ponens.[25] The version of the *Indicative Constraint* that I assume here says that if $R(w) \cap p \neq \varnothing$, then for all $v \in R(w)$, $f(p,v) \subseteq R(w)$.[26] Due to Stalnaker (1975), the indicative constraint is motivated by the observation that the 'or'-to-'if' inference sounds valid. For example, if we know that it was either the gardener or the butler, and we leave open that it was not the gardener, then we know that if it was not the gardener, it was the butler. The indicative constraint is the standard way to ensure that knowing a disjunction is sufficient for knowing the indicative conditional (provided one doesn't know the first disjunct).[27] For more motivation and discussion, see von Fintel (2001), Gillies (2009), Bacon (2015), Mandelkern (ms), Boylan & Schultheis (msb).

Again, normality of K and K_ψ is directly built into our frames. More interestingly, QT is valid on conditional Kripke frames:

24 Cf. Mandelkern (ms) for arguments for the validity of 'If p, then p'.

25 McGee (1985) famously proposed a counterexample to modus ponens. I think this shouldn't affect QT. MP provides merely the easiest route to the left-to-right direction of QT. (MP gives us $\phi \to \psi \vdash \phi \supset \psi$ by the deduction theorem, and so $K(\phi \to \psi) \vdash K(\phi \supset \psi)$ by RM_K.) But a weakening of MP will also do. In particular, it is generally accepted that modus ponens is informationally valid in the sense if one fully accepts $\phi, \phi \to \psi$, one must fully accept ψ (cf. Over 1987, Santorio 2018, Mandelkern forthcoming). But then plausibly if one knows ϕ and one knows $\phi \to \psi$, one knows ψ. By extension, if one knows $\phi \to \psi$, one knows $\phi \supset \psi$, ensuring the left-to-right direction of QT.

26 There are somewhat different versions of the Indicative constraint, see Mandelkern (2018a), Boylan & Schultheis (msb). Normally, the indicative constraint is cashed out in terms of the context set of a conversation. For simplicity, I equate the context set here with what is known.

27 Strictly speaking, principles like this one must hold context fixed to avoid triviality (cf. Mandelkern & Khoo 2019): If the knowledge K_c associated with context c leaves open $[\![\phi]\!]^c$, then K_c contains $[\![\phi \to \psi]\!]^c$ just in case it contains $[\![\phi \supset \psi]\!]^c$. I ignore this complication for simplicity.

Fact 2.11. QT is valid on all conditional Kripke frames where f fulfils Success, Weak Centering and the Indicative Constraint.

We can also use conditional Kripke frames to interpret our extended language containing attributions of multiply conditional knowledge.

- $[\![K_{\psi_1,\ldots,\psi_n}\phi]\!]^w = 1$ iff $[\![K(\psi_1 \to \ldots(\psi_n \to \phi))]\!]^w = 1$

Interestingly, QT^+ comes out invalid given this semantic clause, so perhaps QT^+ is dubious.[28] Nevertheless, my main result relies only on QT, and QT is valid on the probability frames.

Upshot: If we model conditional knowledge as knowing a conditional, standard constraints on indicative conditionals enforce QT. I take this to be some evidence in favour of QT. Note we have not required that R to be reflexive. Our models thus remain applicable to non-factive notions like belief, justification, or being sure, and support QT also for these non-factive states.

2.4.3 AGM revision

A third way to think about conditional knowledge is in terms of knowledge revision. Roughly: One knows ϕ conditional on ψ just in case one knows ϕ upon revising one's knowledge with ψ. In this section, I show that standard AGM revision operations Alchourrón et al. (1985) satisfy QT, and popular ways to extend them to multiply conditional knowledge satisfy QT^+.

Following Alchourrón et al. (1985), much of the belief revision literature models belief states as sets of sentences K, and revisions operations as taking a set of sentences K and a sentence ϕ into another set of sentences $K * \phi$. Nothing prevents us from following this syntactic approach to revision. Let K be a set of sentences closed under logical consequence, and $*$ a partial meet revision

28 Here is a model where QT^+ fails: $W = \{w_i \mid 1 \leq i \leq 6\}$, $R = \{\langle w_i, w_j\rangle \mid |i-j| < 2\}$. Let $p = \{w_1, w_5, w_6\}$, $q = \{w_1, w_6\}$, $r = \{w_1\}$, and $f(w_1, p) = f(w_1, q) = f(w_2, p) = \{w_1\}$, $f(w_4, p) = \{w_5\}$, and $f(w_5, q) = \{w_6\}$ (the other values of f don't matter). Let $[\![\phi]\!] = p$, $[\![\psi]\!] = q$, and $[\![\chi]\!] = r$. Then at w_3, $\neg K(\phi \to \neg\psi)$, and $K(\phi \to (\psi \supset \chi)$, but $\neg K(\phi \to (\psi \to \chi))$.

(Alchourrón et al. 1985).[29] We can then give semantic clauses for knowledge in terms of K and conditional knowledge in terms of $*$ which validate QT:

- $[\![K\phi]\!] = 1$ iff $\phi \in K$
- $[\![K_\psi\phi]\!] = 1$ iff $\phi \in K * \psi$

Fact 2.12. Any partial meet revision fulfils QT.

How about QT$^+$? The original AGM revision operations only deal with single revisions, so they do not provide the resources to interpret multiply conditional knowledge. However, as a workaround one can interpret multiply conditional knowledge as knowledge conditional on the conjunction:

- $[\![K_{\psi_1,\ldots,\psi_n}\phi]\!] = 1$ iff $\phi \in K * (\psi_1 \wedge \ldots \wedge \psi_n)$

Given this workaround, QT$^+$ comes out valid for *transitively relational* partial meet revisions, i. e. partial meet revisions fulfilling Superexpansion and Subexpansion:

- Superexpansion: $K * (p \wedge q) \subseteq Cl((K * p) \cup \{q\})$
- Subexpansion: $q \notin Cl(K * p) \Rightarrow Cl((K * p) \cup \{q\}) \subseteq K * (p \wedge q)$

Fact 2.13. Any *transitively relational* partial meet revision fulfils QT$^+$.

Partial meet revision operations, and transitively relational partial meet revisions in particular, were the belief revision operations originally introduced by Alchourrón et al. (1985) and remain the most standard belief revision operations (cf. Fermé & Hansson 2018). QT, and perhaps also QT$^+$, thus have some plausibility if we cash out conditional knowledge as revised knowledge. At the very least, challenging them requires meddling with the classic frameworks for belief revision.

29 For our purposes, we can define a partial meet revision as an operation $*$ that fulfils the following postulates: Closure $K * p = Cl(K * p)$, Success $p \in K * p$, Inclusion $K * p \subseteq Cl(K \cup \{p\})$, Vacuity $p \notin K \Rightarrow K * p = Cl(K \cup \{p\})$, Consistency $p \neq \bot \Rightarrow \bot \notin K * p$, Extensionality $\vdash p \equiv q \Rightarrow K * p = K * q$.

However, the approach we just sketched is unfortunate in a number of ways. First, the identification of knowledge conditional on $\psi_1, ..., \psi_n$ with knowledge given the conjunction $\psi_1 \wedge ... \wedge \psi_n$ seems dubious.[30] Second, it contains no explicit representation of knowledge about knowledge (cf. van Ditmarsch et al. 2007: sect. 3.4). Third, one might prefer representing beliefs in terms of sets of possible worlds instead of sets of sentences on philosophical grounds (cf. Stalnaker 2009b). Fortunately, all of these worries can be be met. One can equivalently reformulate the AGM framework in terms of orderings on worlds (Grove 1988), extend the framework to allow for iterated belief revision (Darwiche & Pearl 1997, Lin 2019), and represent iterated knowledge in the way customary for Kripke frames.

Let's start with simple revision frames, i.e. pairs $\langle W, \geq \rangle$ where W is a set of points, informally called worlds, and $\geq$ is a function from worlds w to total preorders $\geq_w$ over W.[31] We assume the limit assumption for $\geq$, i.e. for all worlds $w \in W$ and consistent propositions $p \neq \varnothing$, $\min_{\geq_w}(p) := \{v \in p \mid \nexists u \in p : u >_w v\} \neq \varnothing$.[32] Our models $\langle W, \geq, V \rangle$ are revision frames extended with a valuation function $V : At \to \mathscr{P}(W)$. Apart from the usual semantic clauses for atoms and connectives, we interpret knowledge in terms of $\geq$:

- $[\![K\phi]\!]^w = 1$ iff $\min_{\geq_w}(W) \subseteq [\![\phi]\!]$
- $[\![K_\psi \phi]\!]^w = 1$ iff $\min_{\geq_w}([\![\psi]\!]) \subseteq [\![\phi]\!]$

Fact 2.14. Any simple revision frame $\langle W, \geq \rangle$ validates QT.

To interpret our extended language containing ascriptions of multiply conditional knowledge, we extend our simple revision frames to triples $\langle W, \geq, \restriction \rangle$, where W and $\geq$ are as before, and $\restriction$ is a function taking a total preorder $\geq_w$ on W and a proposition $p \subseteq W$ into another total preorder $\geq_w \restriction p$ on W, written as $\geq_{w,p}$.[33] We

30 The identification is strongly reminiscent of import-export in conditional logic, which is known to cause havoc. Cf. Gibbard (1981), Fitelson (2013, 2015), Mandelkern (ms).

31 That is for any $w \in W$, $\geq_w$ is reflexive ($v \geq_w v$ for all $v \in W$), transitive (for any $v_1, v_2, v_3 \in W$: if $v_1 \geq_w v_2$ and $v_2 \geq_w v_3$, then $v_1 \geq_w v_3$), and total (for all $u, v \in W$: either $v \geq_w u$ or $u \geq_w v$).

32 $u >_w v$ is defined as $u \geq_w v$ and not $v \geq_w u$.

33 For similar approaches to multiple revision, cf. Darwiche & Pearl (1997), Lin (2019).

require $\min_{\geq_{w,p}}(W) \subseteq p$ if $p \neq \varnothing$. Models are quadruples $\langle W, \geq, \restriction, V \rangle$. We write $\geq_{w,p_1,\ldots,p_n}$ for $(\geq_w \restriction p_1) \ldots \restriction p_n$, and amend the clause for conditional knowledge:

- $[\![K_{\psi_1,..,\psi_n}\phi]\!]^w = 1$ iff $\min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!]}(W) \subseteq [\![\phi]\!]$

Generally, one will want to constrain $\restriction$, i.e. the relation between $\geq_w$ and $\geq_{w,p}$. There is some disagreement about what the best postulates are (cf. Fermé & Hansson 2018). For our purposes, the first condition of the standard approach due to Darwiche & Pearl (1997) suffices:[34]

(DP1) For $u, v \in p$, $u \geq_w v$ iff $u \geq_{w,p} v$.

DP1 suffices to ensure QT^+:

Fact 2.15. Any extended revision frame $\langle W, \geq, \restriction \rangle$ where $\restriction$ meets DP1 validates QT^+.

Upshot: The orthodox approach to belief revision ensures that QT holds, and the most common way to extend this approach to iterated belief revision ensures that QT^+ holds. I do not want to overstate my conclusion; there are certainly ways to develop belief revision frameworks that avoid committing to QT and QT^+.[35] However, we should think twice before rejecting QT and QT^+, and only reject them with strong reasons.

Let's sum up this section. I have introduced three formal models of conditional attitudes, all of which brute-force normality and validate QT. Slightly stronger constraints are needed to get QT^+, but even QT^+ has some plausibility at least on probabilistic and AGM revision models. Denying the background assumptions of my proof is thus costly, although of course not impossible.

34 The other postulates of Darwiche & Pearl (1997) are: (DP2) for $u, v \notin p$, $u \geq_w v$ iff $u \geq_{w,p} v$. (DP3) if $u \in p, v \notin p$ and $u >_w v$ then $u >_{w,p} v$. (DP4) if $u \in p, v \notin p$ and $u \geq_w v$ then $u \geq_{w,p} v$.

35 All one needs is partial instead of total orderings, cf. Katsuno & Mendelzon (1991), Lin (2019).

2.5 Discussion

My results leave us in a somewhat uncomfortable spot. NI is widely endorsed for knowledge, belief, and justification, and CNI looks like an innocuous extension. KK is commonly, although by no means universally, rejected. But given QT, we must either reject CNI or accept KK. Rejecting QT comes with costs of its own. This section discusses how we should react to this predicament.

As I have stressed again and again, nothing in my result forces interpreting my language in terms of knowledge. We can also consider interpretations in terms of belief, justified belief, or being sure. Different reactions to my result might be appropriate for different interpretations of the language. For any given interpretation, there are broadly three different takes on my result. First, one can take the result to be a new argument for KK. Second, one can reject KK and take the result to undermine CNI. Third, one can accept CNI and reject KK by blaming QT. I'm not sure which reaction is right. The choice will likely depend on one's prior opinions about KK, CNI, and QT. In the following sections, I develop each of these takes on the result a little more.

2.5.1 Take 1: A new argument for KK

If one views my result as a new argument for KK, one needs to say something to motivate CNI. To motivate CNI, let's look at NI first. For knowledge, NI follows directly from factivity ($K\phi \supset \phi$). But NI is plausible for other attitudes as well, among other things because NI encodes an anti-Moorean requirement:

Anti-Moore: Don't hold S towards $\phi \wedge \neg S\phi$!

Anti-Moore is plausible for lots of attitudes S, including knowledge, belief, justification, and rational sureness. *Anti-Moore* can explain why Moorean conjunctions of the form 'p but I don't know/believe/I'm not sure that/justified in believing p' sound bad. It also explains why attributing Moorean beliefs to other people is weird. More generally, NI is the weakest principle prohibiting beliefs which can be

true but only if they aren't believed.[36] For these and other reasons, NI is widely endorsed for knowledge, belief, and justification.[37]

Here is a natural generalisation of *Anti-Moore*:

General Anti-Moore: Don't hold S_ψ towards $\phi \wedge \neg S_\psi \phi$!

Just like it is irrational to believe 'p but I don't know that p', it seems irrational to believe on supposition q that 'p but I don't know p on the supposition that q.' This generalises to other attitudes. It seems equally weird to be sure given q that 'p but I'm not sure given q that p'. In similar vein, assertions of the form 'If q, then p but I don't know that if q, then p' sound quite odd. These weirdness of these broadly Moorean assertions motivates extending *Anti-Moore* to *General Anti-Moore*. But given minimal assumptions, fulfilling *General Anti-Moore* guarantees CNI.[38] If conditional knowledge encodes dispositions to revise one's knowledge, *General Anti-Moore* seems particularly plausible, for presumably one should not only avoid Moorean attitudes, but also avoid being disposed to hold Moorean attitudes.

A related route to CNI is via interaction principles for conditional knowledge and conditional belief:

(1) $K_\phi \psi \supset B_\phi \psi$

(2) $B_\phi \psi \supset \neg K_\phi \neg K_\phi \psi$

(1) and (2) directly entail CNI, and (1) and (2) themselves extend popular interaction principles for knowledge and belief to conditional knowledge and conditional belief. The requirement that knowledge entails belief ($K\phi \supset B\phi$) is almost universally accepted. (1) is the parallel requirement that conditional knowledge entails conditional belief. The assumption that one can't or shouldn't believe ϕ while

36 To be precise, call ϕ Moorean just in case $\nvdash_{\mathrm{KD}} \phi$ but $\vdash_{\mathrm{KD}} \neg(\phi \wedge K\phi)$. Rieger (2015) shows that an extension L of KD contains NI if it ensures for all Moorean ϕ that $\vdash_{\mathrm{L}} \neg K\phi$.

37 Cf. Aucher (2014), Lenzen (1979), Rieger (2015), Stalnaker (2006). For defences of NI for justification, cf. Rosenkranz (2018), Smithies (2012: 327).

38 All that's required is the agglomeration principle $(S_\psi \phi \wedge S_\psi \chi) \supset S_\psi(\phi \wedge \chi)$.

knowing one doesn't know ϕ ($B\phi \supset \neg K \neg K \phi$) is also widely shared.[39] (2) extends this requirement to conditional states.[40]

A nice feature of the first take on my result is that it can align the logic of K with the logic of K_ψ. If one accepts NI but rejects KK, the only way to correspondingly accept CNI and reject CKK is to deny QT. But if we accept both NI and KK, we can accept CNI and CKK without being forced to deny QT. Accepting all of KK, CKK, NI, and CNI means accepting a KD4 logic for both knowledge and conditional knowledge. (Of course, factivity will disrupt complete harmony, i.e. one will want to accept T_K ($K\phi \supset \phi$) but reject T_{K_ψ} ($K_\psi \phi \supset \phi$). Still, conditional knowledge plausibly implies conditional truth, i.e. one will accept $K_\psi \phi \supset (\psi \supset \phi)$, or even $K_\psi \phi \supset (\psi \rightarrow \phi)$ for some stronger conditional '$\rightarrow$'. So although knowledge is factive and conditional knowledge isn't, this is not a real disanalogy as conditional knowledge is 'conditionally factive'.)

This opens up the possibility for an abductive argument for KK, and more generally for positive introspection for belief, justification, and sureness: Positive introspection is required to preserve harmony between our logics for unconditional and conditional attitudes. Such harmony seems desirable. For one thing, it simplifies our logic. For another, it is hard to see why descriptive or normative constraints on knowledge, belief, or being sure should diverge from constraints on their conditional counterparts.[41] In practice, such harmony has generally been assumed.[42] Sometimes, the motivations for a constraint in the unconditional case even generalise directly to the conditional case (as for *Anti-Moore*).

39 For defences of $B\phi \supset \neg K \neg K \phi$, cf. Stalnaker (2006), Lenzen (1979), Aucher (2014), and Chalki et al. (2018). If you like the knowledge norm for belief, you should accept $B\phi \supset \neg K \neg K \phi$ for agents who don't knowingly violate the knowledge norm. See also Smullyan (1987: 84).

40 If you think belief is weak (Hawthorne et al. 2016, Dorst 2017, Rothschild forthcoming), please interpret (1) and (2) as about *outright belief* or *being sure*.

41 Cf. Bradley (2017a: 92): "But since conditional belief is a form of belief, and conditional desire a form of desire, it is reasonable to expect them at least to satisfy the usual rationality constraints on these types of attitude." See also Dorst (2019: 1237) for the thought that principles like KK should hold for *contracted* knowledge states just like for unconditional knowledge.

42 E.g., cf. Joyce (1999) on parallels between conditional and unconditional belief.

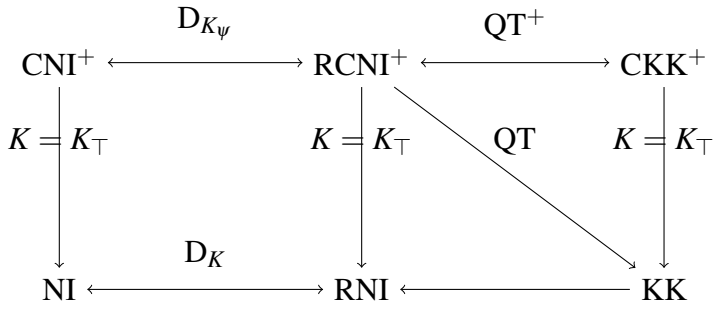


Figure 2.1: Logical relations between KK, NI, RNI, CKK^+, NI^+, and $RCNI^+$.

One advantage of this abductive argument for positive introspection is that there is no analogous argument for negative introspection. Lehrer (1970: 133) suggests that KK should be sustained because it brings a gain in theoretical simplicity, providing a simple account of what it takes to know that one knows.[43] The problem with this, and similar arguments is that the overgenerate, pushing us to accept also K¬K ($\neg K\phi \supset K\neg K\phi$).[44] K¬K provides a similarly simply account of what it takes to know one doesn't know. And in fact, it is not moving from KT to S4, but to S5 where a significant gain in theoretical simplicity is achieved.[45] This is why computer scientists tend to like S5.[46] But of course the 5 axiom is philosophically unacceptable (cf. Stalnaker 2006), and so the argument shows too much.

In this subsection, I have shown how CNI can motiviated from an anti-Moorean requirement for conditional attitudes. QT also seems well-supported. Perhaps we should accept KK after all.

43 Cf. Lehrer (1970: 133) for a (dubious) abductive argument for KK: "[T]he gain in theoretical simplicity that [KK] effects makes it worth some effort to sustain. Any theory of knowledge must confront the problem of explaining under what conditions we know that we know, and it is simplifying indeed to be able to respond that the conditions are exactly the same as those under which we know simpliciter."

44 Cf. San (ms) for discussion how arguments for KK overgenerate, supporting K¬K.

45 For example, only in S5 collapse iterated modalities, only in S5 are accessibility relations superfluous for the model theory, and only in S5 become nice normal form theorems available.

46 Cf. Aucher (2014), Fagin et al. (1995), Meyer & Van Der Hoek (2004).

2.5.2 Take 2: An argument against CNI

What can we do if don't want to accept KK? Straightaway rejecting CNI without substitute won't do — after all there is quite a bit of evidencce for CNI. But one might look for weakenings of CNI that are consistent with KK-denial. Here is one variant of CNI that is consistent with denying KK:

CNI*. $K\neg K_\psi\phi \supset \neg K_\psi\phi$

Like NI, CNI* is an instance of factivity. CNI looks plausible, one might contend, but only CNI* is true. We could try a similar argument for *General Anti-Moore*. If we think of conditional knowledge as knowing the indicative conditional, then one might reject *General Anti-Moore* and endorse only the following instance of *Anti-Moore*:

Anti-Moore$_\rightarrow$*:* Don't hold S towards $(\phi \rightarrow \psi) \wedge \neg S(\phi \rightarrow \psi)$!

Anti-Moore$_\rightarrow$ corresponds to CNI*, given the identification of conditional knowledge with knowing the indicative conditional. So one way to reject CNI and *General Anti-Moore* is to assume only these substitute principles.

If we explain the plausibility of NI for knowledge in terms of factivity, why does NI look plausible for belief and other non-factive notions? In particular, how do we explain the badness of Moorean conjunctions of the form 'ϕ but I don't believe that ϕ'? Here is a standard option:[47] It seems bad to believe things of the form 'ϕ and I don't believe that ϕ' because one's belief is guaranteed to be false: If one truly believes the conjunction, one believes both conjuncts, so the second conjunct is false. And while factivity doesn't hold for beliefs, it ought to hold, while beliefs are not generally true, they *aim* at truth. This is why NI is plausible as a *normative* requirement for belief.

This line of defence derives the plausibility of NI as a descriptive requirement on knowledge (from factivity), and as a normative requirement on belief (from a

47 See Sorensen (1988), Williamson (2000).

truth-norm). Can we derive CNI in a similar way? No. As mentioned, conditional knowledge is clearly not factive, one may know false things on false suppositions. And conditional belief does not aim at truth — at best conditional belief aims at conditional truth. So if we go for this line of defence, we do get rid of CNI.

The plausibility of this strategy depends on how we think of conditional attitudes. If we think of conditional attitude states as attitude states in their own right, CNI has considerable purchase. Compare multi-agent doxastic logic: an intrapersonal version of *Anti-Moore* is plausible here, whereas an interpersonal version looks implausible:

Interpersonal Anti-Moore $\forall i, j : \neg B_i(\phi \wedge \neg B_j\phi)$
Intrapersonal Anti-Moore $\forall i : \neg B_i(\phi \wedge \neg B_i\phi)$

Sara may well believe 'p but Tom doesn't believe that p.' What Sara shouldn't, and perhaps can't do is to believe 'p but I don't believe that p.' On this way of thinking, it is natural to formulate anti-Moorean requirements such as to hold fixed the index of knowledge and belief operators. By analogy, one might think, it is CNI and not CNI* that we should endorse. For CNI holds the index of the knowledge operator fixed, whereas CNI* varies it.

On the other hand, we might think conditional knowledge states reduce to knowing conditionals. Then such analogies with multi-agent cases will seem dubious. Conditional knowledge states are not like agents in their own right — they just reduce to one agent believing various conditionals. In that case, *Anti-Moore*$_{\rightarrow}$ + CNI* is a more natural package to adopt than *General Anti-Moore* + CNI.

2.5.3 Take 3: Rejecting QT

Rejecting QT is costly, as it is ingrained in standard frameworks for conditional attitudes. We better have good reasons before we object to QT. But perhaps the plausibility of CNI and the implausibility of KK is such a reason.

The left-to-right direction of QT effectively requires one to hold on to beliefs/knowledge whenever consistency permits. Stalnaker (2009b: 194) appeals to this idea in an argument for the left-to-right direction of QT for belief:

> To fully accept something (to treat it as knowledge) is [...] to continue accepting it unless evidence forces one to give up something.

QT says that one should hold on to knowledge/beliefs whenever consistency permits. But maybe we shouldn't care only about consistency. Some belief/knowledge states are consistent but undesirable nevertheless, like believing/knowing "p but I don't believe/know p." This is what *General Anti-Moore* encodes. If updating on something compatible with our knowledge would leave us in a Moorean state, perhaps we are permitted or even obligated to remove something from our beliefs/knowledge to avoid the Moorean state.[48] *General Anti-Moore* might thus be viewed as a principled reason to resist QT. Some beliefs are consistent, but don't fit with one another anyway.

If one likes this diagnosis why QT is false, one might support it by arguments from the belief revision literature, where *Preservation*, a consequence of QT, has been discussed. It is usually formulated in terms of belief:

Preservation: If one believes q without believing $\neg p$, one believes q conditional on p.

Preservation is rather controversial in the belief revision literature. Here is a well-known counterexample:[49]

Composers: You believe on independent grounds that Bizet is French, that Satie is French, and that Verdi is Italian. You are first told by a reliable source

48 A parallel, more ambitious requirement is that no knowledge state should be akratic, thinking both 'p' and 'probably I don't know that p.' This gives you undercutting defeat. To do: Clarify this.

49 Lin (2019) uses the example to argue against *Preservation*. Stalnaker (1994) used it to argue against a related (but slightly different) principle, inspired by a similar example involving counterfactuals by Ginsberg, who apparently took it from Quine.

> that Bizet and Verdi are compatriots. You are later also told by a reliable source that Bizet, Satie, and Verdi are all compatriots.

What should you believe after being told that Bizet and Verdi are compatriots? Intuitively, you believe that Satie is French, and that Bizet and Verdi are either both French or both Italian. After being told that all three composers are compatriots, you should only believe that the three composers are either all French or all Italian. If this description is correct, we have a counterexample to *Preservation*: After receiving the first bit of information and before receiving the second one, you (still) believe that Satie is French, but you have no belief as to whether Bizet and Verdi are both French or both Italian. By *Preservation*, then, you believe that Satie is French on the supposition that all three composers are compatriots. But this precisely doesn't seem right — upon being told that all three composers are compatriots, you leave open that Satie is Italian.

Many similar counterexamples to *Preservation* have been proposed in the belief revision literature.[50] The more general point here is that *Preservation* is controversial in belief revision because it does not sit well with nonmonotonic belief change.[51] This point has been extensively appreciated in the literature:[52]

> If I initially believe that B but my reasons for doing so are undermined by learning that A, I should be free to give up my belief in B even if it is neither logically nor epistemically inconsistent to believe both A and B. (Bradley 2007: 155)

50 For similar examples, see Cross (1990), Rabinowicz (1996), Bradley (2012, 2017b), Chandler (2017), Fermé & Hansson (2018: 45). See Rott (2017) for discussion.

51 Another common argument against *Preservation* are dynamic triviality results for languages containing modals or conditionals, e.g. Fuhrmann (1989) and Gärdenfors (1986). I find such arguments against *Preservation* unconvincing; the second paper of this thesis explains why.

52 For similar sentiments, see Rott (1989, 2017), Cross (1990), Bradley (2012, 2017b), Chandler (2017), Koons (2017), Fermé & Hansson (2018: 44ff.), Lin (2019).

Rational belief change, one might think, is highly nonmonotonic. For parallel reasons, knowledge change might be nonmonotonic, leading to failures of QT.[53] This way of thinking fits well with the first take on may result, as it also involves rejecting QT, and the monotonic frameworks of belief revision supporting it.

2.6 Conclusion

In this paper, I have examined level-bridging principles for conditional attitudes. My main result is that assuming QT, CNI entails KK. This is a puzzling result. CNI is a natural way to generalise the popular principle NI to conditional knowledge states, but KK is considered false by many epistemologists. It is not entirely clear what to make of this result: Should we accept KK, reject CNI, or reject the formal frameworks that allow us to derive KK from CNI?

One way to view my result is as providing the basis for a new abductive argument for KK, and other positive introspection principles: If we want to hold on to QT, we must assume KK if we want to preserve harmony between our logic for knowledge and our logic for conditional knowledge. Evaluating this argument would require getting clear on whether such harmony is desirable.

A different way to view my result is as a new argument for defeat: My result shows that if KK can fail, and some anti-Moorean constraint such as CNI or RCNI is right, then there is a kind of defeat, i.e. one can know one thing p and leave open another thing q, but fail to know p conditional on q.

2.7 Formal appendix

$$\mathscr{L} := p \mid \neg\phi \mid (\phi \supset \chi) \mid K\phi \mid K_\psi\phi$$

A modal logic is a set of $\mathscr{L}$-sentences containing all classical truth-functional tautologies (PC) closed under modus ponens (MP) and uniform substitution (US).

53 Rothschild & Spectre (2018b) and Holguín (2019) have (independently) rejected 'or-to-if', endorsing only the weakening $K(K(\phi \vee \psi) \wedge \neg K\phi) \supset K(\neg\phi \rightarrow \psi)$. This weakening does not suffice for my result. I explore these connections in other work.

$\vdash_{\mathrm{L}} \phi$ abbreviates $\phi \in \mathrm{L}$, and $\phi_1, ..., \phi_n \vdash_{\mathrm{L}} \psi$ iff $\vdash_{\mathrm{L}} (\phi_1 \wedge ... \wedge \phi_n) \supset \psi$. Where context disambiguates, we use $\vdash$.

Let L be the smallest *normal* modal logic, i.e. closed under RN_K $(\phi / K\phi)$ and RN_{K_ψ} $(\phi / K_\psi\phi)$, and containing all instances of K_K $(K(\phi \supset \psi) \supset (K\phi \supset K\psi))$ and K_{K_ψ} $(K_\psi(\phi \supset \chi) \supset (K_\psi\phi \supset K_\psi\chi))$. L is then closed under the montonicity rules RM_K $(\phi \supset \psi / K\phi \supset K\psi)$ and RM_{K_ψ} $(\phi \supset \chi / K_\psi\phi \supset K_\psi\chi)$, as well as RE_K $(\phi \equiv \psi / K\phi \equiv K\psi)$ and RE_{K_ψ} $(\phi \equiv \chi / K_\psi\phi \equiv K_\psi\chi)$.[54] We write '$\mathrm{L} + \mathrm{X}_1 + ... + \mathrm{X}_n$' for the smallest extension of L containing the axioms $\mathrm{X}_1 + ... + \mathrm{X}_n$.

We introduce labels for a bunch of principles:

KK.	$K\phi \supset KK\phi$	(= 4_K in modal logic lingo)
CKK.	$K_\psi\phi \supset K_\psi K_\psi\phi$	(= 4_{K_ψ} in modal logic lingo)
NI.	$K\neg K\phi \supset \neg K\phi$	(= $5c_K$ in modal logic lingo)
CNI.	$K_\psi\neg K_\psi\phi \supset \neg K_\psi\phi$	(= $5c_{K_\psi}$ in modal logic lingo)
QT.	$M\psi \supset (K_\psi\phi \equiv K(\psi \supset \phi))$	

Lemma 2.1. $\mathrm{L} + \mathrm{QT}$ contains Inclusion $(K_\psi\phi \supset K(\psi \supset \phi))$ and Success $(M\phi \supset K_\phi\phi)$.

Proof. Start with Inclusion. By PC $\vdash \neg\psi \supset (\psi \supset \phi)$, and thus $\vdash K\neg\psi \supset K(\psi \supset \phi)$ by RM_K. By Duality, $\vdash \neg M\psi \supset K(\psi \supset \phi)$, so $\vdash \neg M\psi \supset (K_\psi\phi \supset K(\psi \supset \phi))$ by PC. The other case, $\vdash M\psi \supset (K_\psi\phi \supset K(\psi \supset \phi))$, is an instance of QT.

On to Success. $\vdash \phi \supset \phi$ by PC, so $\vdash K(\phi \supset \phi)$ by RN_K. By QT, $M\phi \vdash K_\phi\phi \equiv K(\phi \supset \phi)$, and thus $M\phi \vdash K_\phi\phi$. □

Fact 2.1. $\mathrm{L} + \mathrm{QT} + \mathrm{CNI}$ contains KK.

Proof. Immediate from fact 2.2. □

Fact 2.2. $\mathrm{L} + \mathrm{QT} + \mathrm{RCNI}$ contains KK.

54 This is a familiar fact (Chellas 1980: 114). To illustrate, suppose L is normal for $\Box$, and $\vdash_{\mathrm{L}} \phi \supset \psi$. Then $\vdash_{\mathrm{L}} \Box(\phi \supset \psi)$ by $\mathrm{RN}_\Box$, and so $\vdash_{\mathrm{L}} \Box\phi \supset \Box\psi$ by MP and $\mathrm{K}_\Box$, as required by $\mathrm{RM}_\Box$. Similarly, if $\vdash_{\mathrm{L}} \phi \equiv \psi$ then $\vdash_{\mathrm{L}} \Box(\phi \equiv \psi)$ by $\mathrm{RN}_\Box$, and so $\vdash_{\mathrm{L}} \Box\phi \equiv \Box\psi$ by MP and $\mathrm{K}_\Box$, as required by $\mathrm{RE}_\Box$.

RCNI. $\neg K_\psi \bot \supset (K_\psi \neg K_\psi \phi \supset \neg K_\psi \phi)$

Proof. By lemma 1, we can use Inclusion and Success.

1.	$\phi \equiv ((\phi \supset \neg K\phi) \supset \phi)$	PC
2.	$K\phi \equiv K((\phi \supset \neg K\phi) \supset \phi)$	RE_K 1
3.	$K_{(\phi \supset \neg K\phi)}\phi \supset K((\phi \supset \neg K\phi) \supset \phi)$	Inclusion
4.	$\neg K\phi \supset \neg K_{(\phi \supset \neg K\phi)}\phi$	PC 2, 3
5.	$K_{(\phi \supset \neg K\phi)}(\neg K\phi \supset \neg K_{(\phi \supset \neg K\phi)}\phi)$	RN_{K_ψ} 4
6.	$K\phi \wedge \neg KK\phi$	Assumption
7.	$M\neg K\phi$	$\wedge$E, Duality 6
8.	$M(\phi \supset \neg K\phi)$	RM_K 7
9.	$K_{(\phi \supset \neg K\phi)}\bot$	Asssumption
10.	$K((\phi \supset \neg K\phi) \supset \bot)$	Inclusion 9
11.	$K\neg(\phi \supset \neg K\phi)$	RM_K 10
12.	$M(\phi \supset \neg K\phi) \wedge \neg M(\phi \supset \neg K\phi)$	$\wedge$I, Duality 8, 11
13.	$\bot$	PC 12
14.	$\neg K_{(\phi \supset \neg K\phi)}\bot$	reductio 9-13
15.	$K((\phi \supset \neg K\phi) \supset \phi)$	MP 2, 6
16.	$M(\phi \supset \neg K\phi) \supset (K_{(\phi \supset \neg K\phi)}\phi \equiv K((\phi \supset \neg K\phi) \supset \phi))$	QT
17.	$K_{(\phi \supset \neg K\phi)}\phi \equiv K((\phi \supset \neg K\phi) \supset \phi)$	MP 8, 16
18.	$K_{(\phi \supset \neg K\phi)}\phi$	MP 15, 17
19.	$M(\phi \supset \neg K\phi) \supset K_{(\phi \supset \neg K\phi)}(\phi \supset \neg K\phi)$	Success
20.	$K_{(\phi \supset \neg K\phi)}(\phi \supset \neg K\phi)$	MP 8, 19
21.	$K_{(\phi \supset \neg K\phi)}\neg K\phi$	MP, K_{K_ψ} 18, 20
22.	$K_{(\phi \supset \neg K\phi)}\neg K_{(\phi \supset \neg K\phi)}\phi$	MP, K_{K_ψ} 5, 21
23.	$\neg K_{(\phi \supset \neg K\phi)}\bot \supset (K_{(\phi \supset \neg K\phi)}\neg K_{(\phi \supset \neg K\phi)}\phi \supset \neg K_{(\phi \supset \neg K\phi)}\phi)$	RCNI
24.	$K_{(\phi \supset \neg K\phi)}\neg K_{(\phi \supset \neg K\phi)}\phi \supset \neg K_{(\phi \supset \neg K\phi)}\phi$	MP 14, 23
25.	$\neg K_{(\phi \supset \neg K\phi)}\phi$	MP 22, 24
26.	$K_{(\phi \supset \neg K\phi)}\phi \wedge \neg K_{(\phi \supset \neg K\phi)}\phi$	$\wedge$I 18, 25
27.	$\bot$	PC 26
28.	$K\phi \supset KK\phi$	reductio 6-27

□

$$\mathscr{L}^+ := p \mid \neg\phi \mid (\phi \supset \chi) \mid K_{\psi,\ldots,\psi_n}\phi \quad (n \geq 0)$$

Let L^+ be the smallest normal modal logic over $\mathscr{L}^+$. We write '$\mathrm{L}^+ + \mathrm{X}_1 + \ldots + \mathrm{X}_n$' for the smallest extension of L^+ containing the axioms $\mathrm{X}_1 + \ldots + \mathrm{X}_n$.

Lemma 2.2. $\mathrm{L}^+ + \mathrm{QT}^+$ contains Inclusion$^+$ ($K_{\psi_1,\ldots,\psi_n,\phi}\chi \supset K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi)$) and Success$^+$ ($M_{\psi_1,\ldots,\psi_n}\phi \supset K_{\psi_1,\ldots,\psi_n,\phi}\chi$).

Proof. Start with Inclusion$^+$. By PC $\vdash \neg\phi \supset (\phi \supset \chi)$, and thus $\vdash K_{\psi_1,\ldots,\psi_n}\neg\phi \supset K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi)$ by RM_{K_ψ}. By Duality, $\vdash \neg M_{\psi_1,\ldots,\psi_n}\phi \supset K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi)$, so $\vdash \neg M_{\psi_1,\ldots,\psi_n}\phi \supset (K_{\psi_1,\ldots,\psi_n,\phi}\chi \supset K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi))$ by PC. The other case, $\vdash M_{\psi_1,\ldots,\psi_n}\phi \supset (K_{\psi_1,\ldots,\psi_n,\phi}\chi \supset K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi))$, is an instance of QT^+.

On to Success$^+$. $\vdash \phi \supset \phi$ by PC, so $\vdash K_{\psi_1,\ldots,\psi_n}(\phi \supset \phi)$ by RN_{K_ψ}. By QT^+, $M_{\psi_1,\ldots,\psi_n}\phi \vdash K_{\psi_1,\ldots,\psi_n,\phi}\phi \equiv K_{\psi_1,\ldots,\psi_n}(\phi \supset \phi)$, and thus $M_{\psi_1,\ldots,\psi_n}\phi \vdash K_{\psi_1,\ldots,\psi_n,\phi}\phi$. □

Fact 2.3. $\mathrm{L}^+ + \mathrm{QT}^+$ contains RCNI$^+$ iff it contains CKK$^+$.

CKK$^+$. $K_{\psi_1,\ldots,\psi_n}\phi \supset K_{\psi_1,\ldots\psi_n}K_{\psi_1,\ldots\psi_n}\phi$ $(n \geq 1)$

RCNI$^+$. $\neg K_{\psi_1,\ldots\psi_n}\bot \supset (K_{\psi_1,\ldots\psi_n}\neg K_{\psi_1,\ldots\psi_n}\phi \supset \neg K_{\psi_1,\ldots\psi_n}\phi)$ $(n \geq 1)$

QT^+. $M_{\psi_1,\ldots\psi_n}\phi \supset (K_{\psi_1,\ldots\psi_n,\phi}\chi \equiv K_{\psi_1,\ldots\psi_n}(\phi \supset \chi))$ $(n \geq 1)$

Proof. To avoid clutter, we write ψ to abbreviate $\psi_1, \ldots, \psi_n$. We first show that $\mathrm{L}^+ + \mathrm{CKK}^+$ contains RCNI$^+$. By CKK$^+$, $K_\psi\neg K_\psi\phi \wedge K_\psi\phi \vdash K_\psi\neg K_\psi\phi \wedge K_\psi K_\psi\phi$. By normality of K_ψ, $K_\psi\neg K_\psi\phi \wedge K_\psi K_\psi\phi \vdash K_\psi(\neg K_\psi \wedge K_\psi) \vdash K_\psi\bot$. Putting everything together, $K_\psi\neg K_\psi\phi \wedge K_\psi\phi \vdash K_\psi\bot$. By the deduction theorem and contraposition, $\vdash \neg K_\psi\bot \supset (K_\psi\neg K_\psi\phi \supset \neg K_\psi\phi)$, which is just RCNI$^+$.

We now show that $\mathrm{L}^+ + \mathrm{QT}^+ + \mathrm{RCNI}^+$ contains CKK$^+$. By lemma 2, we can use Inclusion$^+$ and Success$^+$.

1. $\phi \equiv ((\phi \supset \neg K_\psi\phi) \supset \phi)$ PC

2.	$K_\psi\phi \equiv K_\psi((\phi \supset \neg K_\psi\phi) \supset \phi)$	RE_{K_ψ} 1
3.	$K_{\psi\wedge(\phi\supset\neg K_\psi\phi)}\phi \supset K_\psi((\phi \supset \neg K_\psi\phi) \supset \phi)$	Inclusion$^+$
4.	$\neg K_\psi\phi \supset \neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi$	PC 2, 3
5.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}(\neg K_\psi\phi \supset \neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi)$	RN_{K_ψ} 4
6.	$K_\psi\phi \wedge \neg K_\psi K_\psi\phi$	Assumption
7.	$M_\psi\neg K_\psi\phi$	$\wedge$E, Duality 6
8.	$M_\psi(\phi \supset \neg K_\psi\phi)$	RM_{K_ψ} 7
9.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\bot$	Asssumption
10.	$K_\psi((\phi \supset \neg K_\psi\phi) \supset \bot)$	Inclusion 9
11.	$K_\psi\neg(\phi \supset \neg K_\psi\phi)$	RM_{K_ψ} 10
12.	$M_\psi(\phi \supset \neg K_\psi\phi) \wedge \neg M_\psi(\phi \supset \neg K_\psi\phi)$	$\wedge$I, Duality 8, 11
13.	$\bot$	PC 12
14.	$\neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\bot$	reductio 9-13
15.	$K_\psi((\phi \supset \neg K_\psi\phi) \supset \phi)$	MP 2, 6
16.	$M_\psi(\phi \supset \neg K_\psi\phi) \supset (K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi \equiv K_\psi((\phi \supset \neg K_\psi\phi) \supset \phi))$	QT$^+$
17.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi \equiv K_\psi((\phi \supset \neg K_\psi\phi) \supset \phi)$	MP 8, 16
18.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi$	MP 15, 17
19.	$M_\psi(\phi \supset \neg K_\psi\phi) \supset K_{\psi,(\phi\supset\neg K_\psi\phi)}(\phi \supset \neg K_\psi\phi)$	Success$^+$
20.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}(\phi \supset \neg K_\psi\phi)$	MP 8, 19
21.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\neg K_\psi\phi$	MP, K_{K_ψ} 18, 20
22.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi$	MP, K_{K_ψ} 5, 21
23.	$\neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\bot \supset (K_{\psi,(\phi\supset\neg K_\psi\phi)}\neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi \supset \neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi)$	RCNI
24.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi \supset \neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi$	MP 14, 23
25.	$\neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi$	MP 22, 24
26.	$K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi \wedge \neg K_{\psi,(\phi\supset\neg K_\psi\phi)}\phi$	$\wedge$I 18, 25
27.	$\bot$	PC, 26
28.	$K_\psi\phi \supset K_\psi K_\psi\phi$	reductio 6-27

Unpacking ψ into $\psi_1, ..., \psi_n$ yields $K_{\psi_1,...,\psi_n}\phi \supset K_{\psi_1,...,\psi_n}K_{\psi_1,...,\psi_n}\phi$. $\mathrm{L}^+ + \mathrm{QT}^+ + \mathrm{RCNI}^+$ thus contains CKK^+.

Since $\mathrm{L}^+ + \mathrm{CKK}^+$ contains RCNI^+, and $\mathrm{L}^+ + \mathrm{QT}^+ + \mathrm{RCNI}^+$ contains CKK^+, $\mathrm{L}^+ + \mathrm{QT}^+$ contains CKK^+ iff it contains RCNI^+. □

Corollary 2.1. $\mathrm{L}^+ + \mathrm{QT}^+ + \mathrm{CNI}^+$ contains CKK^+.

CNI^+. $\quad K_{\psi_1,\ldots\psi_n}\neg K_{\psi_1,\ldots\psi_n}\phi \supset \neg K_{\psi_1,\ldots\psi_n}\phi \qquad (n \geq 1)$

Proof. Since CNI^+ strengthens RCNI^+, this is immediate from fact 2.3. □

Fact 2.4. $\langle W, R, \restriction \rangle$ validates QT iff whenever $R(w) \cap p \neq \varnothing$, $R \restriction_p (w) = R(w) \cap p$.

Proof. Let $\langle W, R, \restriction, V \rangle$ be any model such that $R \restriction_p (w) = R(w) \cap p$ whenever $R(w) \cap p \neq \varnothing$ for all $p \subseteq W$, and let $w \in W$. If $R(w) \cap [\![\psi]\!] = \varnothing$, then $M\psi \supset (K_\psi\phi \equiv K(\psi \supset \phi))$ is trivially true at w. So suppose $R(w) \cap [\![\psi]\!] \neq \varnothing$. Then $R(w) \cap [\![\psi]\!] = R \restriction_{[\![\psi]\!]} (w)$, and so $R(w) \cap [\![\psi]\!] \subseteq [\![\phi]\!]$ just in case $R \restriction_{[\![\psi]\!]} (w) \subseteq [\![\phi]\!]$, and thus $[\![K_\psi\phi \equiv K(\psi \supset \phi)]\!]^w = 1$. So either way, $M\psi \supset (K_\psi\phi \equiv K(\psi \supset \phi))$ is true at w.

Now let $\langle W, R, \restriction \rangle$ be a frame such that $R(w) \cap p \neq \varnothing$, but $R \restriction_p (w) \neq R(w) \cap p$. Then either there is $w \in W$ and $v \in R \restriction_p (w)$ such that $v \notin R(w) \cap p$, or there is $w \in W$ and $v \in R(w) \cap p$ such that $v \notin R \restriction_p (w)$. Either way, we choose $V(A) = \{v\}$ and $V(B) = p$ such that $[\![MB \supset (K_BA \equiv K(B \supset A))]\!]^w = 0$. □

Fact 2.5. $\langle W, R, \restriction \rangle$ validates KK iff R is transitive, validates NI iff R is mildly transitive, and validates RNI ($\neg K\bot \supset (K\neg K\phi \supset \neg K\phi)$) iff R is very midly transitive.

Proof. The proof that KK corresponds to transitivity is trivial.

Regarding NI, suppose $\langle W, R \rangle$ invalidates NI. Then there is a model $\langle W, R, \restriction, V \rangle$ and world $w \in W$ such that $[\![K\phi \wedge K\neg K\phi]\!]^w = 1$. Hence $R(w) \subseteq [\![\phi]\!]$, but $\forall v \in R(w) : R(v) \nsubseteq [\![\phi]\!]$. Then R isn't mildly transitive, since $\forall v \in R(w) \exists u \in R(v) : u \notin R(w)$.

Now suppose R is not mildly transitive. Then there is $w \in W$ such that $\forall v \in R(w) : \exists u \in R(v) : u \notin R(w)$. Choose valuation V such that $V(A) = R(w)$, and then $[\![K(A) \wedge K\neg K(A)]\!]^w = 1$ and so NI is invalidated.

On to RNI. Suppose $\langle W,R,\restriction\rangle$ invalidates RNI. Then there is a model $\langle W,R,\restriction,V\rangle$ and world $w\in W$ such that $[\![\neg K\bot\wedge K\phi\wedge K\neg K\phi]\!]^w=1$. Hence $\varnothing\neq R(w)\subseteq[\![\phi]\!]$, but $\forall v\in R(w):R(v)\nsubseteq[\![\phi]\!]$. R then is not even very mildly transitive, since all $v\in R(w)\exists u\in R(v):u\notin R(w)$, and $R(w)\neq\varnothing$.

Now suppose R is not very mildly transitive. Then there is $w\in W$ such that $R(w)\neq\varnothing$ and $\forall v\in R(w):\exists u\in R(v):u\notin R(w)$. Choose valuation V such that $V(A)=R(w)$, so $[\![\neg K\bot\wedge K(A)\wedge K\neg K(A)]\!]^w=1$ and RNI is invalidated. □

Fact 2.6. $\langle W,R,\restriction\rangle$ validates CKK iff $R\restriction_p$ is transitive for all $p\subseteq W$, validates CNI iff $R\restriction_p$ is mildly transitive for all $p\subseteq W$, and validates RCNI iff $R\restriction_p$ is very mildly transitive for all $p\subseteq W$.

Proof. Focus on CKK first. Suppose $\langle W,R,\restriction\rangle$ invalidates CKK. Then there is some model $\langle W,R,\restriction,V\rangle$ and world $w\in W$ such that $[\![K_\psi\phi\wedge\neg K_\psi K_\psi\phi]\!]^w=1$. This means $R\restriction_{[\![\psi]\!]}(w)\subseteq[\![\phi]\!]$, but $\exists v\in R\restriction_{[\![\psi]\!]}(w):R\restriction_{[\![\psi]\!]}(v)\nsubseteq[\![\phi]\!]$. Thus $wR\restriction_{[\![\psi]\!]}v$ and $vR\restriction_{[\![\psi]\!]}u$ but not $wR\restriction_{[\![\psi]\!]}u$ for some w,v,u, so $R\restriction_{[\![\psi]\!]}$ is not transitive.

Now suppose $R\restriction_p$ is not transitive for some $p\subseteq W$. Then there are $w,v,u\in W$ with $wR\restriction_p v$ and $vR\restriction_p u$ but not $wR\restriction_p u$. Choose valuation V such that $V(A)=R\restriction_p(w)$ and $V(B)=p$. This means $[\![K_B(A)\wedge\neg K_BK_B(A)]\!]^w=1$, invalidating CKK.

Now consider CNI. Suppose $\langle W,R,\restriction\rangle$ invalidates CNI. Then there is some model $\langle W,R,\restriction,V\rangle$ and world $w\in W$ such that $[\![K_\psi\phi\wedge K_\psi\neg K_\psi\phi]\!]^w=1$. This means $R\restriction_{[\![\psi]\!]}(w)\subseteq[\![\phi]\!]$, but $\forall v\in R\restriction_{[\![\psi]\!]}(w):R\restriction_{[\![\psi]\!]}(v)\nsubseteq[\![\phi]\!]$. So for any world $v\in R\restriction_{[\![\psi]\!]}(w)$ there is $u\in R\restriction_{[\![\psi]\!]}(v)$ such that $u\notin R\restriction_{[\![\psi]\!]}(w)$. So $R\restriction_{[\![\psi]\!]}$ isn't mildly transitive.

Now suppose $R\restriction_p$ is not mildly transitive. Then there is w such that $\forall v\in R\restriction_p(w)\exists u\in R\restriction_p(v):u\notin R\restriction_p(w)$. Choose valuation V such that $V(A)=R\restriction_p(w)$, and $V(B)=p$. This means $[\![K_B(A)\wedge K_B\neg K_B(A)]\!]^w=1$ and hence NI is invalidated.

Now consider RCNI. Suppose $\langle W,R,\restriction\rangle$ invalidates RCNI. Then there is some model $\langle W,R,\restriction,V\rangle$ and world $w\in W$ such that $[\![\neg K_\psi\bot\wedge K_\psi\phi\wedge K_\psi\neg K_\psi\phi]\!]^w=1$. Hence $\varnothing\neq R\restriction_{[\![\psi]\!]}(w)\subseteq[\![\phi]\!]$, but $\forall v\in R\restriction_{[\![\psi]\!]}(w):R\restriction_{[\![\psi]\!]}(v)\nsubseteq[\![\phi]\!]$. So $R\restriction_{[\![\psi]\!]}$

$(w) \neq \varnothing$ and for any $v \in R\restriction_{[\![\psi]\!]}(w)$ there is a world $u \in R\restriction_{[\![\psi]\!]}(v)$ such that $u \notin R\restriction_{[\![\psi]\!]}(w)$. So $R\restriction_{[\![\psi]\!]}$ isn't even very mildly transitive.

Now suppose $R\restriction_p$ is not very mildly transitive. Then there is w such that $R\restriction_p(w) \neq \varnothing$ and $\forall v \in R\restriction_p(w) \exists u \in R\restriction_p(v) : u \notin R\restriction_p(w)$. Choose valuation V such that $V(A) = R\restriction_p(w)$, and $V(B) = p$. This means $[\![\neg K_B \bot \wedge K_B(A) \wedge K_B \neg K_B(A)]\!]^w = 1$ and hence RCNI is invalidated. □

Fact 2.7. If $\langle W, R, \restriction \rangle$ validates QT, and $R\restriction_p$ is mildly or at least very mildly transitive for all $p \subseteq W$, then R is transitive.

Proof. Suppose R is not transitive, so there are $w, v, u \in W$ with wRv and vRu but not wRu. Let $p = \{v, u\}$. Then $R(w) \cap p \neq \varnothing$ and $R(v) \cap p \neq \varnothing$, and so by the validity of QT and fact 2.4, $R\restriction_p(w) = R(w) \cap p$ and $R\restriction_p(v) = R(v) \cap p$. So $R\restriction_p(w) \neq \varnothing$ and for all $x \in R\restriction_p(w) \exists y \in R\restriction_p(x) : y \notin R\restriction_p(w)$ (since $R\restriction_p(w) = \{v\}$, and $u \in R\restriction_p(v)$ but $u \notin R\restriction_p(w)$). So $R\restriction_p$ isn't very mildly transitive.

This establishes that if $R\restriction_p$ is very mildly transitive for all $p \subseteq W$ then R is transitive. The parallel claim for mild transitivity follows immediately since any mildly transitive relation is very mildly transitive. □

Fact 2.8. If $\langle W, R, \restriction \rangle$ validates QT^+, then $R\restriction_{p_1,..,p_n}$ is very mildly transitive for all $p_1, .., p_n \subseteq W$ just in case $R\restriction_{p_1,..,p_n}$ is transitive for all $p_1, .., p_n \subseteq W$.

Proof. Let $\langle W, R, \restriction \rangle$ validate QT^+. Clearly, any transitive relation is very mildly transitive, so the '$\Rightarrow$' direction is trivial. As to the '$\Leftarrow$' direction, suppose $R_{p_1,...,p_n}$ is not transitive, i.e. there are $w, v, u \in W$ such that $R_{p_1,...,p_n}wv$ and $R_{p_1,...,p_n}vu$ but not $R_{p_1,...,p_n}wu$. Let $q = \{v, u\}$, and consider $R_{p_1,...,p_n,q}$. Since $R_{p_1,...,p_n}(w) \cap q \neq \varnothing$ and $R_{p_1,...,p_n}(v) \cap q \neq \varnothing$, by QT^+ and (a trivial extension of) fact 2.4, $R_{p_1,...,p_n,q}(w) = R_{p_1,...,p_n}(w) \cap q$ and $R_{p_1,...,p_n}(v) = R_{p_1,...,p_n}(v) \cap q$. But then $R_{p_1,...,p_n,q}$ is not very mildly transitive, for $R_{p_1,...,p_n,q}(w) \neq \varnothing$ (as it contains v), and for all all $x \in R_{p_1,...,p_n,q}(w)$, $R_{p_1,...,p_n,q}(x) \nsubseteq R_{p_1,...,p_n,q}(w)$ (for v is the only such x, and $R_{p_1,...,p_n,q}vu$ but not $R_{p_1,...,p_n,q}wu$). □

Corollary 2.2. If $\langle W, R, \restriction \rangle$ validates QT$^+$, then if $R \restriction_{p_1,..,p_n}$ is mildly transitive for all $p_1, .., p_n \subseteq W$ then $R \restriction_{p_1,..,p_n}$ is transitive for all $p_1, .., p_n \subseteq W$.

Proof. Immediate from the previous fact, since any mildly transitive relation is very mildly transitive. □

Fact 2.9. QT is valid on all probability frames.

Proof. $K_\psi \phi$ is true at w just in case $P_w([\![\phi]\!] \mid [\![\psi]\!]) = 1$, and $K(\psi \supset \phi)$ is true at w iff $P_w([\![\psi \supset \phi]\!]) = 1$. Provided $M\psi$ is true at w, that is provided $P_w([\![\psi]\!]) > 0$, these hold in the same conditions:

$$\begin{aligned}
& P_w([\![\phi]\!] \mid [\![\psi]\!]) = 1 && \\
\Leftrightarrow\ & P_w([\![\neg\phi]\!] \mid [\![\psi]\!]) = 0 && \\
\Leftrightarrow\ & P_w([\![\psi \wedge \neg\phi]\!])/P_w([\![\psi]\!]) = 0 && P_w([\![\psi]\!]) > 0 \\
\Leftrightarrow\ & P_w([\![\psi \wedge \neg\phi]\!]) = 0 && P_w([\![\psi]\!]) > 0 \\
\Leftrightarrow\ & P_w([\![\psi \supset \phi]\!]) = 1 &&
\end{aligned}$$

□

Fact 2.10. QT$^+$ is valid on all probability frames.

Proof. By an analogous argument, $[\![QT]\!]^+$ is valid: $K_{(\psi_1,...,\psi_n,\phi)}\chi$ is true at w just in case $P_w([\![\chi]\!] \mid [\![\psi]\!], ..., [\![\psi_n]\!], [\![\phi]\!]) = 1$, and $K_{\psi_1,...,\psi_n}(\phi \supset \chi)$ is true at w iff $P_w([\![\phi \supset \chi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!]) = 1$. Provided $M_{\psi_1,...,\psi_n}\phi$ is true at w, that is provided (*) $P_w([\![\phi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!]) > 0$, these are equivalent:

$$\begin{aligned}
& P_w([\![\chi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!], [\![\phi]\!]) = 1 && \\
\Leftrightarrow\ & P_w([\![\neg\chi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!], [\![\phi]\!]) = 0 && \\
\Leftrightarrow\ & P_w([\![\phi \wedge \neg\chi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!])/P_w([\![\phi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!]) = 0 && \text{by (*)} \\
\Leftrightarrow\ & P_w([\![\phi \wedge \neg\chi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!]) = 0 && \text{by (*)} \\
\Leftrightarrow\ & P_w([\![\phi \supset \chi]\!] \mid [\![\psi_1]\!], ..., [\![\psi_n]\!]) = 1 &&
\end{aligned}$$

□

Fact 2.11. QT is valid on all conditional Kripke frames where f fulfils Success, Weak Centering and the Indicative Constraint.

Proof. QT$_{\Rightarrow}$ first. Suppose $K_\psi\phi$ and so $K(\psi \to \phi)$ is true at w. Let $v \in R(w)$. Either $v \in [\![\psi]\!]$, or $v \notin [\![\psi]\!]$. If $v \notin [\![\psi]\!]$, then $\psi \supset \phi$ is trivially true at v. If $v \in [\![\psi]\!]$, then $v \in f([\![\psi]\!], v)$ by Weak Centering, and so ϕ is true at v. So $\psi \supset \phi$ is true at all $v \in R(w)$, and so $K(\psi \supset \phi)$ at w.

Now QT$_{\Leftarrow}$. Suppose $K(\psi \supset \phi)$ and $M\psi$ are true at w. Let $v \in R(w)$. Then by the indicative constraint $f([\![\psi]\!], v) \subseteq R(w)$, and thus $f([\![\psi]\!], v) \subseteq [\![\psi \supset \phi]\!]$. By Success, $f([\![\psi]\!], v) \subseteq [\![\psi]\!]$. So $f([\![\psi]\!], v) \subseteq [\![\phi]\!]$ and so $\psi \to \phi$ is true at v. So $K(\psi \to \phi)$ and $K_\psi\phi$ are true at w. □

Fact 2.12. Any partial meet revision fulfils QT.

Proof. QT $_{\Rightarrow}$ first: Suppose that $K_\phi\psi$ is true, and thus $\psi \in K * \phi$. By Inclusion, $\psi \in Cl(K \cup \{\phi\})$ and hence $(\phi \supset \psi) \in Cl(K)$, and hence $(\phi \supset \psi) \in K$ by closure of K. Then $K(\phi \supset \psi)$ is also true.

Now QT $_{\Leftarrow}$: Suppose that $K(\phi \supset \psi) \wedge M\phi$. Then $(\phi \supset \psi) \in K$, and $\neg\phi \notin K$, so by Vacuity $K * \phi = Cl(K \cup \{\phi\})$. Combining both points, $\psi \in K * \phi$, and thus $K_\phi\psi$. □

Fact 2.13. Any *transitively relational* partial meet revision fulfils QT$^+$.

Proof. QT$^+_{\Rightarrow}$ first: Suppose that $K_{\psi_1,\ldots,\psi_n,\phi}\chi$ is true, and thus $\chi \in K * (\psi_1 \wedge \ldots \wedge \psi_n \wedge \phi)$. By Superexpansion, $\chi \in Cl(K * (\psi_1 \wedge \ldots \wedge \psi_n) \cup \{\phi\})$ and hence $(\phi \supset \chi) \in Cl(K * (\psi_1 \wedge \ldots \wedge \psi_n))$, and hence $(\phi \supset \chi) \in K * (\psi_1 \wedge \ldots \wedge \psi_n)$ by closure of $K * (\psi_1 \wedge \ldots \wedge \psi_n)$. Then $K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi)$ is also true.

Now QT$^+_{\Leftarrow}$: Suppose that $K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi) \wedge M_{\psi_1,\ldots,\psi_n}\phi$. Then $(\phi \supset \chi) \in K * (\psi_1 \wedge \ldots \wedge \psi_n)$, and $\neg\phi \notin K * (\psi_1 \wedge \ldots \wedge \psi_n)$, so by Subexpansion $Cl(K * (\psi_1 \wedge \ldots \wedge \psi_n) \cup \{\chi\}) \subseteq K * ((\psi_1 \wedge \ldots \wedge \psi_n) \wedge \phi)$. Combining both points, $\chi \in K * ((\psi_1 \wedge \ldots \wedge \psi_n) \wedge \phi)$, and thus $K_{\psi_1,\ldots,\psi_n,\wedge\chi}\chi$. □

Fact 2.14. Any simple revision frame $\langle W, \geq \rangle$ validates QT.

Proof. Let $\langle W, \geq \rangle$ be a simple revision frame, i.e. W a set of points, and $\geq$ is a function from worlds w to total preorders $\geq_w$ over W, such that for all worlds $w \in W$ and consistent propositions $p \neq \varnothing$, $\min_{\geq_w}(p) := \{v \in p \mid \nexists u \in p : u >_w v\} \neq \varnothing$. Suppose $[\![M\psi]\!]^w = 1$ for some model $\langle W, \geq, V \rangle$, so $\min_{\geq_w}(W) \cap [\![\psi]\!] \neq \varnothing$. Then:

$$
\begin{aligned}
[\![K_\psi \phi]\!]^w = 1 &\Leftrightarrow \min_{\geq_w}([\![\psi]\!]) \subseteq [\![\phi]\!] \\
&\Leftrightarrow (\min_{\geq_w}(W) \cap [\![\psi]\!]) \subseteq [\![\phi]\!] \qquad \text{since } \min_{\geq_w}(W) \cap [\![\psi]\!] \neq \varnothing \\
&\Leftrightarrow \min_{\geq_w}(W) \subseteq ([\![\neg\psi]\!]) \cup [\![\phi]\!]) \\
&\Leftrightarrow \min_{\geq_w}(W) \subseteq [\![\psi \supset \phi]\!] \\
&\Leftrightarrow [\![K(\psi \supset \phi)]\!]^w = 1
\end{aligned}
$$

□

Fact 2.15. Any extended revision frame $\langle W, \geq, \restriction \rangle$ where $\restriction$ meets DP1 validates QT^+.

Proof. Let $\langle W, \geq, \restriction \rangle$ be an extended revision frame satisfying DP1. That is, W is a set of points, $\geq$ a function from worlds w to total preorders $\geq_w$ over W, such that for all worlds $w \in W$ and consistent propositions $p \neq \varnothing$, $\min_{\geq_w}(p) := \{v \in p \mid \nexists u \in p : u >_w v\} \neq \varnothing$, and $\restriction$ fulfils

(DP1) For $u, v \in p$, $u \geq_w v$ iff $u \geq_{w,p} v$.

Note the following consequence of DP1:

(*) If $\min_{\geq, w, p_1, \ldots, p_n}(W) \cap q \neq \varnothing$, $\min_{\geq, w, p_1, \ldots, p_n, q}(W) = \min_{\geq, w, p_1, \ldots, p_n}(W) \cap p$.

Now suppose $[\![M_{\psi_1, \ldots, \psi_n} \phi]\!]^w = 1$ for some model $\langle W, \geq, \restriction, V \rangle$,. By the semantic clause for K, $\min_{\geq_w, [\![\psi_1]\!], \ldots, [\![\psi_n]\!]}(W) \cap [\![\phi]\!] \neq \varnothing$. By (*):

(**) $\quad \min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!],[\![\phi]\!]}(W) = \min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!]}(W) \cap [\![\phi]\!]$.

From (**), $[\![K_{\psi_1,\ldots,\psi_n,\phi}\chi \equiv K_{\psi_1,\ldots,\psi_n}(\phi \supset \chi)]\!]^w$ follows easily:

$$
\begin{aligned}
[\![K_{\psi_1,\ldots,\psi_n,\phi}\chi]\!]^w = 1 &\Leftrightarrow \min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!],[\![\phi]\!]}(W) \subseteq [\![\chi]\!] \\
&\Leftrightarrow (\min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!]}(W) \cap [\![\phi]\!]) \subseteq [\![\chi]\!] \qquad \text{by (**)} \\
&\Leftrightarrow \min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!]}(W) \subseteq ([\![\neg\phi]\!]) \cup [\![\chi]\!]) \\
&\Leftrightarrow \min_{\geq_w,[\![\psi_1]\!],\ldots,[\![\psi_n]\!]}(W) \subseteq [\![\phi \supset \chi]\!] \\
&\Leftrightarrow [\![K_{\psi_1,..,\psi_n}(\phi \supset \chi)]\!]^w = 1
\end{aligned}
$$

□

3 Preservation and Reflection

Abstract Fuhrmann (1989) pointed out that in standard AGM belief revision frameworks, preservation fails for reflective agents. Various drastic conclusions have been drawn from this observation. I distinguish preservation for *sentences* from preservation for *propositions*. Preservation for *sentences* fails for reflective agents, but for languages involving context-sensitive sentences, only preservation for *propositions* is plausible. Once we disambiguate belief operators by allowing both belief and conditional belief operators in the object language, we can accept preservation even for sentences. However, it then turns out that preservation becomes problematic for *unreflective* agents, in particular for agents for whom positive introspection fails. This is new argument against preservation is much more convincing.

3.1 Introduction

A common belief in the AGM literature is that preservation cannot be sustained for reflective agents.[1] Preservation is the principle that if one believes p and does not believe that not q, then one believes p conditional on q. A number of authors have drawn radical conclusions from the (seeming) incompatibility of preservation with reflectivity. Levi (1988) denies that beliefs about one's own beliefs can be elements of belief sets, Fuhrmann (1989) and Segerberg (2006) redefine revision, Gillies (2006) concludes that consequence is not persistent, and Willer (2010) argues that our semantics should use pairs of information states, one keeping track of one's beliefs, the other of one's suppositions.

1 Cf. Fuhrmann (1989), Rott (1989, 2011, 2017).

These radical conclusions seem like overreactions. We can distinguish preservation for *sentences* from preservation for *propositions*. Preservation for *sentences* fails for reflective agents, but is implausible anyway for languages involving context-sensitive sentences.[2] If we insist on a purely syntactic framework, we can keep preservation for sentences provided we eliminate context-dependency (Duca & Leitgeb 2012). Ironically, once we eliminate context-dependency by explicitly distinguishing belief and conditional belief in our object language, one can prove that preservation is inconsistent with *failures* of reflectivity.

Here is the plan: §3.2 sets out the alleged conflict between preservation and reflection, §3.3 resolves the conflict, and §3.4 establishes the 'opposite' conflict between preservation and failures of reflection.

3.2 The problem

Following Alchourrón et al. (1985), much of the belief revision literature models belief states as sets of sentences K of a propositional language $\mathscr{L}_{PC} := p \mid \neg\phi \mid (\phi \supset \psi)$. Revision operations can then be modelled as functions taking a belief state K and a sentence ϕ into another belief state $K * \phi$.[3] Revision is usually contrasted with contraction and extension. Contraction with ϕ is minimal belief change with the aim of removing ϕ from one's beliefs. Extension amounts to adding ϕ to one's beliefs and closing them under logical consequence.

The following fairly intuitive constraints on revisions $*$ are usually accepted:

Preservation: If $\phi \in K$ and $\neg\psi \notin K$, then $\phi \in K * \psi$.
Success: $\phi \in K * \phi$.

Preservation encodes the thought that revising your beliefs in the light of information consistent with your beliefs should not lead you to give up any beliefs. Success captures the idea that revising with ϕ will lead one to believe ϕ.

2 Cf. Bacon (2015), Khoo & Mandelkern (2019), Mandelkern & Khoo (2019) for a similar point for preservation for conditionals.
3 Cf. Alchourrón et al. (1985) for the actual definitions.

We extend the object language with an operator $\Box$ ('must') to the language $\mathscr{L}_{\Box} := p \mid \neg\phi \mid (\phi \supset \psi) \mid \Box\phi$. We call K inconsistent when there is $\phi \in K$ such that $\neg\phi \in K$, too. For any set of sentences K, $Poss(K)$ is the smallest superset such that if $\phi \in K$, then $\Box\phi \in K$, and if $\phi \notin K$, then $\neg\Box\phi \in K$.

Full reflectivity: $Poss(K) \subseteq K$ and for all sentences ϕ, $Poss(K * \phi) \subseteq K * \phi$.
Non-Triviality: There is ϕ such that $\phi \notin K$ and $\neg\phi \notin K$, and $K * \phi$ is consistent.

Fact 3.1 ($\approx$ Fuhrmann (1989))**.** No set of sentences K and revision operation $*$ fulfil all of Preservation, Success, Full Reflectivity, and Non-Triviality.

Proof. By Non-Triviality, there is ϕ such that $\phi \notin K, \neg\phi \notin K$ and $K * \phi$ is consistent. Since $\phi \notin K$, $\neg\Box\phi \in K$ by Full Reflectivity. Since $\neg\phi \notin K$ and $\neg\Box\phi \in K$, by Preservation $\neg\Box\phi \in K * \phi$. On the other hand, by Success $\phi \in K * \phi$, and thus by Full Reflectivity also $\Box\phi \in K * \phi$. Since both $\Box\phi \in K * \phi$ and $\neg\Box\phi \in K * \phi$, $K * \phi$ is inconsistent, contradicting Non-Triviality. $\Box$

One thing to note immediately is that while the result is formulated using full reflectivity, all one needs really is reflectivity for one sentence, namely the ϕ whose existence is guaranteed by Non-Triviality. Full reflectivity seems like an implausibly strong constraint on belief sets in general, but this weaker assumption seems plausible enough.

A number of authors have drawn drastic conclusions from this triviality result. Levi (1988) denies that sentences involving $\Box$ (or the dual $\Diamond$) can be elements of belief sets. Levi was attracted to this idea because he thought sentences such as $\Box\phi$ don't bear truth-values. He adopts the same attitude towards conditionals, thereby avoiding related triviality results involving conditionals (Gärdenfors 1986). Hardly anyone is happy with this way of resolving the triviality result, partly because one can reformulate the result for commitments, which Levi allows to include modal formulas (cf. Gillies 2006: for details). And in any case, since we ordinarily seem to have beliefs about what might and must be just as about other things, banning them from belief sets seems like a last resort.

Fuhrmann (1989) suggests that his impossibility result should make us reconsider how to understand revision. For languages including epistemic modalities, the revised belief state $K * \phi$ should be understood as $K - \neg\Box\phi$, where '$-$' is an AGM contraction. Revising with ϕ thus comes down to removing the information that $\neg\phi$ might be true. The reaction seems ad hoc. And Fuhrmann's suggestion does not work for agents who are not fully reflective, requiring us to work with different notions of revision for different languages.[4]

Segerberg (2006) considerably improves on Fuhrmann's suggestion by assuming an AGM revision operation $\star$ in the background, and defining revision $*$ in terms of $\star$. His proposal is that $K * \phi = K \star (\phi \wedge C\phi)$, where $C\phi$ is a operator expressing that one believes ϕ, one believes that one believes ϕ, and so on. (In effect, C is a common belief operator for the single agent case.) Simplifying a bit, Segerberg's idea is that when one revises with ϕ, one should also revise with the information that one believes ϕ, and the information that one believes that one believes ϕ, and so on. Compared to Fuhrmann's suggestion, Segerberg's proposal has the advantage that it works even in the case of belief sets that are not fully reflective. Nevertheless, the proposal seems ad hoc (cf. Enqvist & Olsson 2013).

Stalnaker foreshadowed Segerberg's idea in a discussion of Thomason conditionals, but developed it in a slightly different direction. Thomason conditionals are a kind of (alleged) counterexample to the claim that accepting a conditional goes along with a disposition to accept the consequent upon learning the antecedent:[5]

(1) If my employees dislike me, I will never know it.

Stalnaker (1984) thinks that these aren't really counterexamples to the claim, because if I were to learn that my employees dislike me, I would generally additionally learn that I believe that my employees dislike me:

4 For $K = Cl(\varnothing)$, since $\neg\Box\phi \notin K$ on Fuhrmann's suggestion $K * \phi = K$. This is highly undesirable. Fuhrmann's suggestion could be adjusted to something like $(K - \neg\Box\phi) + \phi$. This still doesn't do well for $K = Cl(\neg\phi)$, giving us $K * \phi = Cl(\bot)$. Arguably, the best way to deal with these issues in Fuhrmann's spirit is to define $K * \phi = K \star (\phi \wedge \Box\phi)$, where $\star$ is an AGM revision in the classical sense. This is very close to the proposal of Segerberg (2006) discussed in the main text.

5 Richmond Thomason is credited by van Fraassen (1980) for a similar example.

> Normally — perhaps always — when we learn A we also learn that we have learned it [...]. Perhaps this is essential to belief. If so, then there are some propositions — propositions which are not in part about the believer's own state of belief — which cannot be the total information received. (Stalnaker 1984: 105)

The idea here is very close to Segerberg's suggestion. But instead of constructing a notion of revision with consist of AGM-revising with both the information ϕ and the information that one believes ϕ, Stalnaker seems to toy with the idea that one simply can't revise with purely factual information — one must also revise with the information that one believes or has learnt that factual information. This would be a fairly radical conclusion in its own right.

Without going into any detail, let me note that Gillies (2006) takes Fuhrmann's result to motivate adopting non-persistent (and thus non-monotonic) consequence relation. Willer (2010) uses it to argue for a double-indexed dynamic semantics, on which information states are pairs $\langle s_1, s_2 \rangle$ of one set of possible worlds $s_1 \subseteq W$ state keeping track of what's believed, and another set of possible worlds $s_2 \subseteq W$ keeping track of what's supposed.

All of these conclusions require a significant revision of existing belief revision frameworks, and I think all of them are overreactions. The next section describes a much simpler and very natural alternative way out of Fuhrmann's result.

3.3 An alternative way out

On standard views, epistemic modals are used to talk about what is entailed, compatible, or likely given our knowledge (Kratzer 1981, 1991b). Roughly, 'must ϕ' is true at a world w iff ϕ is true at all worlds compatible with the relevant knowledge at w. And very roughly, 'probably ϕ' is true at a world w iff ϕ is sufficiently probable given the relevant knowledge at w.[6] But what knowledge

6 This ignores complications orthogonal to my concern, e. g. indirectness (von Fintel & Gillies 2010).

is relevant may differ from context to context, and hence sentences involving epistemic modals express different propositions in different contexts.

Given that sentences involving epistemic modals express different propositions in different contexts, one needs to be very careful when formulating principles such as Preservation. There is no particular reason why knowing the proposition $[\![\Diamond\phi]\!]^c$ expressed by a sentence $\Diamond\phi$ at a context c while leaving open q should guarantee knowing conditional on q a different proposition $[\![\Diamond\phi]\!]^{c'}$ expressed by $\Diamond\phi$ at a different context c'. This is especially pressing when considering whether the *sentence* $\Diamond\phi$ should be preserved under belief revision, as the revised state will generally incorporate more information, such that $\Diamond\phi$ is interpreted relative to different information states.

A simple option is to formulate principles such as Preservation in terms of propositions: If one believes a proposition p and leaves open another proposition q, then one knows p conditional on q. I think this is a perfectly viable option, and I will show in a second how to make this more precise. But it requires us to move from the usual syntactic framework of AGM belief revision to a framework that allows explicit representation of propositions, i.e. sets of possible worlds. Can we somehow stick to the syntactic setting?

Another option would be to formulate principles like Preservation in a way that holds the context of evaluation fixed.[7] The problem with this option is that standard belief revision frameworks do not contain any representation of context. Formulating preservation in a way that holds context fixed would thus require us to significantly complicate our models. This is what Lindström & Rabinowicz (1999a,b) do, effectively introducing a two-dimensional semantics for epistemic operators. In principle, I have little to object to this approach. But it introduces all kinds of complications, and lots of room to go wrong. One of the strengths of the AGM models is that they are very simple, and there is little room to go wrong.

7 Cf. Bacon (2015), Khoo & Mandelkern (2019), Mandelkern & Khoo (2019) for similar points related to conditionals.

Here is a simple fix (cf. Duca & Leitgeb 2012): Instead of holding context fixed when formulating principles like Preservation, we make sure our object language does not contain context-dependent expressions. Instead of working with catch-all modal operators $\Box$ and $\Diamond$, we introduce one epistemic operator for every belief state (as in multiagent doxastic logic):

$$\mathscr{L}_{B,B_\psi} = p \mid \neg\phi \mid \phi \wedge \psi \mid B\phi \mid B_\psi\phi$$

We correspondingly adjust our notion of full reflectivity: For any belief state K, $Dox(K)$ is the smallest superset such that if $\phi \in K$, then $B\phi \in K$, and if $\phi \notin K$, then $\neg B\phi \in K$; and $Dox_\psi(X)$ is the smallest superset of X such that if $\phi \in X$, then $B_\psi\phi \in X$, and if $\phi \notin X$, then $\neg B_\psi\phi \in X$.

Full reflectivity:* $Dox(K) \subseteq K$ and for all sentences ϕ, $Dox_\phi(K * \phi) \subseteq K * \phi$.

This form of full reflectivity is very close to the notion of full reflectivity Fuhrmann (1989), Levi (1988) and so on were interested it, and it is perfectly consistent with Success, Preservation, and Non-Triviality.

Fact 3.2. There are sets of sentences K and revision operations $*$ that fulfil all of Preservation, Success, Full Reflectivity*, and Non-Triviality.

Before we prove this fact, it is instructive to show how our extended language $\mathscr{L}_{B,B_\psi}$ can be interpreted using extended Kripke frames. An extended Kripke frame is a triple $\langle W, R, \restriction \rangle$ where W is a set of points, informally called worlds, $R \subseteq W \times W$ is an accessibility relation, and $\restriction$ is a function from accessibility relations and sets of worlds to accessibility relations.[8] Our models $\langle W, R, \restriction, V \rangle$ are frames extended with a valuation function $V : At \to \mathscr{P}(W)$ from atoms to subsets of W. Validity ($\models$) is truth at all worlds in all models.

We assume the usual semantic clauses for atoms and the connectives, plus

8 Cf. Boylan & Schultheis (msb) for models of this kind. A similar technique is also employed in dynamic epistemic logic, cf. van Benthem (2004).

- $[\![K\phi]\!]^w = 1$ iff $R(w) \subseteq [\![\phi]\!]$
- $[\![K_\psi\phi]\!]^w = 1$ iff $R \restriction_{[\![\psi]\!]} (w) \subseteq [\![\phi]\!]$

Here and later, $[\![\phi]\!] = \{w \in W \mid [\![\phi]\!]^w = 1\}$. All of this is relatively familiar from modal logic. Using these frames allows us to prove the previous consistency claim in an instructive way:

Proof. Let $\langle W, R, \restriction, V\rangle$ be an extended Kripke model such that (i) if $R(w) \cap p \neq \varnothing$, then $R \restriction_p (w) = R(w) \cap p$, (ii) $R \restriction_p \subseteq p$, (iii) R and R_p are transitive and euclidian (for all $p \subseteq W$), and (iv) $\exists p \subseteq W : p \nsubseteq R(w)$ and $R \restriction_p (w) \neq \varnothing$. (To see that there are such frames, just take a transitive euclidean Kripke frame where $|R(w)| > 1$ for some $w \in W$ and let $R \restriction_p (w) := R(w) \cap p$.)

Now choose $w \in W$ such that $|R(w)| > 1$ (exists by (iv)), and let

- $\phi \in K$ iff $[\![K\phi]\!]^w = 1$
- $\phi \in K * \psi$ iff $[\![K_\psi\phi]\!]^w = 1$

Preservation then follows immediately from (i), Success from (ii), Full reflectivity* from (iii). Non-Triviality requires that there is ϕ such that $\phi \notin K$ and $\neg\phi \notin K$ and $K * \phi$ is consistent. Since $|R(w)| > 1$, we can choose $p \subset R(w)$ such that $p \neq \varnothing$, and the first part then holds for ϕ chose as $[\![\phi]\!] = p$. For the second part, note that $R(w) \cap p \neq \varnothing$, so by (i) $R \restriction_p (w) = R(w) \cap p$, and then since $R(w) \cap p \neq \varnothing$, it follows that $K * \phi$ is consistent. □

Let me also briefly note that it is very easy to formulate a preservation condition for propositions in a semantic setting like this.[9] If one has an original belief state, represented as a set of worlds (in our case $R(w)$), and a revision operation like $\restriction$ which in effect takes belief states and propositions p into revised belief states $R \restriction_p (w)$, then one can formulate preservation simply as the constraint that if $R(w) \cap p \neq \varnothing$, then $R \restriction_p (w) \subseteq R(w)$. Semantic ways to formulate preservation

9 There are various ways to do belief revision in a semantic setting. Subsection 2.4.3 presented one option, but see also (Grove 1988, Katsuno & Mendelzon 1991, Stalnaker 2009b, Lin 2019).

generally seem preferable to me, as they completely avoid the issues with context-dependent expressions encountered in the previous section.

3.4 A result in the opposite direction

The belief revision tradition has it that preservation does not sit well with full reflectivity. We have seen above that this impression is wrong, or at least trades on a very implausible *syntactic* version of preservation for languages containing context-dependent expressions. In this section, I show a result in the opposite direction: given weak side-conditions, preservation *entails* reflection principles (in particular BB and BK). This means that if one wishes to reject reflection principles, then must reject preservation.

Fact 3.3 (corollary of fact 3.4)**.** Any normal modal logic over $\mathscr{L}_{B,B_\psi}$ which includes

Preservation$_B$: $(B\phi \wedge \neg B\neg\psi) \supset B_\psi\phi$
Success$_B$: $B_\phi\phi$
Inclusion$_B$: $B_\psi\phi \supset B(\psi \supset \phi)$
Anti-Moore: $\neg B_\psi\bot \supset \neg(B_\psi\phi \wedge B_\psi\neg B_\psi\phi)$

also includes

BB: $\quad B\phi \supset BB\phi$

Proof. Immediate corollary of fact 3.4, by a translation τ from $\mathscr{L}_{B,B_\psi,K,K_\psi}$ into $\mathscr{L}_{B,B_\psi}$ that identifies K with B, and K_ψ with B_ψ. (That is $\tau(K\phi) = B\tau(\phi)$, and $\tau(K_\psi\phi) = B_{\tau(\psi)}\tau(\phi)$, $\tau(\phi \supset \psi) = \tau(\phi) \supset \tau(\psi)$, $\tau(\neg\phi) = \neg\tau(\phi)$, and $\tau(A) = A$ for atoms A.) □

For reasons of proof economy, we infer fact 3.3 as a corollary from the more general fact 3.4. We extend the language $\mathscr{L}_{B,B_\psi,K,K_\psi}$ with sentential operators K ('one knows that') and K_ψ ('one knows conditional on ψ that') to the language $\mathscr{L}_{B,B_\psi,K,K_\psi} = \phi \mid \neg\phi \mid \phi \supset \psi \mid B\phi \mid B_\psi\phi \mid K\phi \mid K_\psi\phi$.

Fact 3.4. Any normal modal logic over $\mathscr{L}_{B,B_\psi,K,K_\psi}$ which includes

Preservation$_B$: $(B\phi \wedge \neg B\neg\psi) \supset B_\psi\phi$
Success$_B$: $B_\phi\phi$
Inclusion$_B$: $B_\psi\phi \supset B(\psi \supset \phi)$
Inclusion$_K$: $K_\psi\phi \supset K(\psi \supset \phi)$
Anti-Moore*: $\neg B_\psi\bot \supset \neg B_\psi(\phi \wedge \neg K_\psi\phi)$

also includes

BK: $B\phi \supset BK\phi$

Proof.

1.	$\phi \equiv ((\phi \supset \neg K\phi) \supset \phi)$	PC
2.	$K\phi \equiv K((\phi \supset \neg K\phi) \supset \phi)$	RE$_K$ 1
3.	$K_{(\phi\supset\neg K\phi)}\phi \supset K((\phi \supset \neg K\phi) \supset \phi)$	Inclusion$_K$
4.	$\neg K\phi \supset \neg K_{(\phi\supset\neg K\phi)}\phi$	PC 2, 3
5.	$B_{(\phi\supset\neg K\phi)}(\neg K\phi \supset \neg K_{(\phi\supset\neg K\phi)}\phi)$	RN$_{B_\psi}$ 4
6.	$B\phi \wedge \neg BK\phi$	Assumption
7.	$\neg B\neg(\phi \supset \neg K\phi)$	$\wedge$E, RM$_B$
8.	$B_{(\phi\supset\neg K\phi)}\phi$	Preservation$_B$ 6, 7
9.	$B_{(\phi\supset\neg K\phi)}(\phi \supset \neg K\phi)$	Success$_B$
10.	$B_{(\phi\supset\neg K\phi)}\neg K\phi$	K$_{B_\psi}$, MP 8, 9
11.	$B_{(\phi\supset\neg K\phi)}\neg K_{(\phi\supset\neg K\phi)}\phi$	K$_{B_\psi}$, MP 5, 10
12.	$B_{(\phi\supset\neg K\phi)}\bot$	Assumption
13.	$B((\phi \supset \neg K\phi) \supset \bot)$	Inclusion$_B$ 12
14.	$B\neg(\phi \supset \neg K\phi)$	RM$_B$ 13
15.	$\bot$	PC 7, 14
16.	$\neg B_{(\phi\supset\neg K\phi)}\bot$	reductio 12-15
17.	$\neg B_{(\phi\supset\neg K\phi)}(\phi \wedge \neg K_{(\phi\supset\neg K\phi)}\phi)$	Anti-Moore 16
18.	$B_{(\phi\supset\neg K\phi)}(\phi \wedge \neg K_{(\phi\supset\neg K\phi)}\phi)$	AGG$_{B_\psi}$ 8, 11

19.	$\bot$	PC 17, 18
20.	$B\phi \supset BK\phi$	reductio 6-19

□

The results show that if we accept Preservation$_B$ and certain side-conditions, then introspection principles BB and BK follow. This means that if one wants to deny BB or BK, one needs to reject preservation, or one of the side-conditions. Success$_B$, Inclusion$_B$, and Inclusion$_K$ are very weak background conditions. They are part of (almost) all frameworks for belief revision, including the classic framework of Alchourrón et al. (1985). The choice here is really between Preservation$_B$, the constraints Anti-Moore and Anti-Moore*, and denial of the iteration principles BB and BK.

There is no doubt that some authors will be willing to accept BB and BK, for example Lenzen (1979) and Stalnaker (2006). These authors like a KD45 logic for belief, and accept the interaction principle $B\phi \equiv \neg K \neg K\phi$. This immediately excludes $B\phi \wedge \neg BB\phi$ (by the 4 axiom) and $B\phi \wedge \neg BK\phi$ (by the interaction principle). But many others, including myself, are more sceptical of BB and BK. For one thing, it has been argued that margin-for-error arguments against KK (Williamson 2000) apply just as much to BB, or parallel principles for what one presupposes to be true (Hawthorne & Magidor 2009, 2010). I don't rehearse these arguments for space reasons.

Against BK, note that it makes error a necessary precondition for ignorance. Let me explain: it is generally acknowledged that an agent may have true beliefs that fall short of knowledge. Plausibly, this kind of situation can arise in a very general way: it may in principle be that one believes only truths, and yet one does not know everything one believes. Natural models of Gettier cases allow for this possibility (Williamson 2013a). Fake-barn cases are often taken to be examples of gettierised beliefs that are not necessarily accompanied by any false beliefs (Goldman 1976). Importantly, however, this possibility is precluded by BK. According to BK, whenever one believes p, one believes that one knows p. Hence

BK entails that one can't have any beliefs falling short of knowledge unless one also has some false belief(s).

Let me take stock. Fuhrmann (1989) pointed out allowing higher-order beliefs into standard frameworks for AGM belief revision leads to triviality for reflective agents. Various drastic conclusions have been drawn from this observation. We have distinguished preservation for *sentences* from preservation for *propositions*, and seen that in languages involving context-sensitive sentences, only preservation for *propositions* is plausible. This takes the bite out of Fuhrmann's result. What's more, if we disambiguate belief operators by allowing both belief and conditional belief operators in the object language, we can accept preservation even for sentences. Ironically, however, it then turns out that preservation becomes problematic for *unreflective* agents, in particular for agents for whom positive introspection fails. I find this new argument against preservation much more convincing.

4 Known Ignorance and KK

Abstract KK is the controversial principle that if one knows p, one knows that one knows p. KK-defenders routinely argue that if KK could fail, one could know dubious conjunctions of the form 'p but I don't know whether I know p.' I show that assuming KK overgenerates, predicting that not only dubious conjunctions, but also 'I don't know whether I know p' on its own is unknowable. This prediction is undesirable because 'I don't know whether I know p' is not infelicitous, as evidenced by actual usage of such constructions.

It is worse still to be ignorant of your ignorance.

Saint Jerome, Letter 53.7

4.1 Introduction

KK is the controversial principle that if one knows that p, one knows that one knows that p.[1] A popular argument for KK exploits the connection between knowledge and assertion. If KK could fail, nothing would prevent one from knowing dubious conjunctions of the form 'p but I don't know whether I know p.' Since dubious conjunctions are unassertable, KK-fans conclude that KK cannot fail.[2]

1 Recent KK-defences include Stalnaker (1999, 2009c), Greco (2014a, 2017), Goodman & Salow (2018), and Das & Salow (2018); anti-KK arguments include Williamson (2000), Dorr et al. (2014), and Carter (2019).

2 See Sosa (2009a), Smithies (2012), Cohen & Comesaña (2013), Greco (2014b, 2015a, 2017), and Das & Salow (2018). Cf. Dorst (2019) for a related argument involving conditionals.

Assertion-based arguments for KK focus on a dangerously narrow range of assertions. Consider for example the right conjunct of dubious assertions, 'I don't know whether I know p.' If KK holds, then one can't know, and hence can't say 'I don't know whether I know p.' But 'I don't know whether I know p.' is perfectly fine, as evidenced by actual usage. I argue based on these, and similar data that dubious conjunctions don't really support KK.

§4.2 gives a more careful exposition of arguments for KK from dubious conjunctions. §4.3 shows that the KK-based account for dubious assertions overgenerates, predicting 'I don't know whether I know p' on its own to be unknowable. §4.4 presents evidence that 'I don't know whether I know p' is commonly used. §4.5 argues that if arguments for KK from ordinary language are to be successful, we need more systematic evidence. §4.6 concludes.

4.2 Assertion-based arguments for KK

According to Malcolm, Wittgenstein "once remarked that the only work of Moore's that greatly impressed him was his discovery of the peculiar kind of nonsense involved in such a sentence as 'It's raining but I don't believe it"' (Malcolm et al. 2001: 56). What puzzled Wittgenstein so much is that conjunctions of the form 'p but I don't believe/know that p.' sound very weird, but are perfectly consistent. This puzzle has since become known as Moore's paradox.[3]

Following Shoemaker (1995), a host of KK-fans have argued that we must assume introspection principles for belief, knowledge, and justification to fully solve Moore's 'paradox'.[4] Clearly, one does not need introspection to account for simple Moorean conjunctions of the form 'p but I don't know that p'. This is because simple Moorean conjunctions cannot be *known* to be true (Hintikka 1962). (If one knew 'p but I don't know that p', one would know the first conjunct p, and

3 The earliest discussions in print seem to be MacDonald (1937) and MacIver (1938). MacDonald (1937) references Moore's lectures. Cf. Williams (2015a,b) for an overview article.

4 Cf. Cohen & Comesaña (2013), Das & Salow (2018), Greco (2014b, 2015a, 2017), and Smithies (2012). For responses and discussion, see Williamson (2005, 2013b), Benton (2013), Montminy (2013), Marušić (2013: fn. 13), Holguín (2019), Liu (forthcoming), and San (ms).

so the second conjunct would be false, and thus the conjunction unknown.) Given that Moorean conjunctions cannot be known, one can explain their infelicity if one assumes that assertion in some sense commits one to knowledge:[5]

(KNA) One must: assert p only if one knows p. (Williamson 2000: 243)

For the purposes of this paper, I will assume the knowledge account of assertion, although similar issues probably arise on other accounts of assertion as well.

This explanation extends to variants of simple Moorean conjunctions:

(1) #He has toothache but I am not sure whether he has. (MacDonald 1937: 30)

(2) #Mussolini is having a bath but I do not know whether he is or not. (MacIver 1938: 48)

(3) #It isn't raining in Chicago, but it may be raining there (Karttunen 1972: 9).

Assuming that ⌜I'm not sure whether ϕ⌝, ⌜I don't know whether ϕ⌝, and ⌜it may be that not ϕ⌝ all entail ⌜I don't know that ϕ⌝, the standard explanation extends to these conjunctions.

However, not just simple Moorean conjunctions, but also conjunctions of the form 'p but I don't believe that I believe that p' sound weird (Sorensen 2000).[6] We focus on the knowledge version here, that is 'p but I don't know whether I know that p', and call them *dubious conjunctions* (Sosa 2009a). Slight variations of dubious conjunctions still sound bad:

(4) a. #It is raining but I might/may not know that it is raining.
b. #It is raining but perhaps/possibly I don't know that it is raining.
c. #It is raining but for all I know I don't know that it is raining.

5 Moore (1962: 277) says that "by asserting p positively, you imply, though you don't assert, that you know that p." See Unger (1975), Gazdar (1979), Williamson (1996, 2000) for arguments for the knowledge norm of assertion along these lines, and Benton (2014) for an overview. Cf. Sorensen (1988) for a more general account of Moorean 'paradoxes' in the same spirit.

6 Sosa (2009a) is usually credited for this observation, but Sorensen (2000), Williams (2007), and Hájek (2007) already discuss iterated Moorean conjunctions.

d. #It is raining but I'm not sure if I know that it is raining.

e. #It is raining but it could be that I don't know that it is raining.

Note that the infelicity is not due to the order of the conjuncts:

(5) a. #I might not know that it is raining, but it is.

b. #Perhaps I don't know that it is raining, but it is.

Sentences (4)-(5) all share that they combine an assertion with a proposal to leave open that the speaker does not know what is asserted. I will be somewhat loose in grouping together weak epistemic modals (⌜might ϕ⌝,⌜may ϕ⌝, ⌜could ϕ⌝), weak modal adverbs (⌜perhaps ϕ⌝, ⌜maybe ϕ⌝, ⌜possibly ϕ⌝), and avowals of ignorance or uncertainty (⌜I don't know whether ϕ⌝, ⌜For all I know, ϕ⌝, ⌜I'm not sure if ϕ⌝) as *proposals to leave open that* ϕ, abbreviated as ⌜$\textsc{open}(\phi)$⌝.[7] Symbolising knowledge that ϕ in the usual way as ⌜$K\phi$⌝, we can then formulate the generalisation underlying (4)-(5) as follows:

> Assertions of *dubious conjunctions* of the form ⌜$\phi \wedge \textsc{open}(\neg K\phi)$⌝ and the form ⌜$\textsc{open}(\neg K\phi) \wedge \phi$⌝ are typically infelicitous.

As KK-fans like to point out, KK plus the knowledge norm of assertion, KNA, predict this generalisation. For if KK holds, one cannot know conjunctions of the form ⌜$\phi \wedge \textsc{open}(\neg K\phi)$⌝ or ⌜$\textsc{open}(\neg K\phi) \wedge \phi$⌝, assuming that ⌜$K\phi$⌝ entails ⌜$\neg\textsc{open}(\neg\phi)$⌝.[8] And if knowledge is required for assertion, we predict that assertions of dubious conjunctions will be infelicitous.

While assuming KK provides an easy account of what's wrong with dubious conjunctions, various authors have suggested that KK-deniers have nothing to say.[9]

7 I take the terminology from Mandelkern (2019b).

8 Suppose $K(\phi \wedge \textsc{open}(\neg K\phi))$. By distribution $K\phi \wedge K(\textsc{open}(\neg K\phi))$. KK plus the first conjunct entail $KK\phi$. By the assumption that $K\phi$ implies $\neg\textsc{open}(\neg\phi)$, the second conjunct entails $K\neg KK\phi$, so by factivity $\neg KK\phi$. Contradiction. The order of the conjuncts does not matter for the proof.

9 Cf. Sosa (2009a), Smithies (2012), Cohen & Comesaña (2013), Greco (2014b, 2015a, 2017), and Das & Salow (2018). For responses and discussion, see Williamson (2005, 2013b), Benton (2013), Montminy (2013), Marušić (2013: fn. 13), Holguín (2019), Liu (forthcoming), and San (ms).

The general worry is that denying KK will allow us to know things which, on the face of it, cannot be known. Upshot: KK-deniers must explain why dubious conjunctions are unassertable if not because they are unknowable. While I do not provide such an account, I argue in the following sections that KK does not account for the infelicity of dubious conjunctions.

4.3 Known ignorance about knowledge

Assertion-based arguments for KK focus on an extremely narrow set of data. This is dangerous, because it might turn out upon closer examination that KK does not account for the full range of data, either. This is what I argue in this section.

Consider the following avowals of ingorance about whether one knows:

(6) a. I don't know whether I know that it is raining.
b. Maybe/perhaps I know that it is raining, maybe/perhaps I don't.
c. I might know that it is raining, and I might not know that it is raining.
d. I'm not sure whether I know that it is raining.

Perhaps these are not completely ordinary, but by no means as weird as dubious conjunctions. While it is weird to say p while expressing uncertainty as to whether you know p, there is nothing obviously wrong with expressing uncertainty whether you know p on its own. Strikingly, however, KK entails that one can't know that one doesn't know whether one knows. So KK seems to over-generate, predicting that not just dubious conjunctions, but sentences like those in (6) are bad.

Let's go over this a bit more carefully. Recall that KK says that whenever you know, you know that you know. So by the lights of KK, the only way you may fail to know whether you know is if you don't know. This latter fact is a logical truth, so you'll know it. But then, if you know that you don't know whether you know, you'll be able to conclude that you don't know, and so after all you do know whether you know. So given KK, it turns out that there is no way for you to know that you don't know whether you know. Assuming that $\ulcorner K\phi \urcorner$ entails $\ulcorner \neg \textsc{open}(\neg\phi) \urcorner$, this means that $\ulcorner \textsc{open}(K\phi) \wedge \textsc{open}(\neg K\phi) \urcorner$ is a *blindspot* in the sense of Sorensen (1988),

i. e. a consistent but unknowable propositions.[10] Since the propositions (6) are equivalent to something of the form $\ulcorner \textsc{open}(K\phi) \wedge \textsc{open}(\neg K\phi)\urcorner$, KK entails that they are blindspots, too. If KK is right, one should then not be able to assert the examples in (6).

Upshot: KK predicts that not just 'p but I don't know whether I know p' but also that 'I don't know whether p' on its own to be a blindspot. KK precludes known ignorance about one's knowledge.

The point that KK precludes known ignorance whether one knows is very close to Fine's observation that KK makes second-order ignorance unknowable. One is (first-order) ignorant whether p just in case one knows neither p nor $\neg p$. One is second-order ignorant whether p *iff* one is ignorant whether one is ignorant whether p. More generally, nth order ignorance whether p is ignorance whether one is $n-1$th order ignorant whether p[11]. Fine (2018) proves that if KK holds, then nth order ignorance can't be known for any $n \geq 2$.

Fine accepts KK and concludes from his results that higher-order ignorance is unknowable. I want to draw precisely the opposite conclusion; since higher-order ignorance seems to be knowable, we should reject KK. How could one resolve this dispute? In the next section, I argue that one can *say* 'I don't know whether I know p.' by showing that people actually say things like this. If what we do say is any guide to what we can say, and what we can say is a guide to what we can know, then pace Fine there is known second-order ignorance.

4.4 What we actually say

In this section, I first survey some evidence that people do express higher-order ignorance by constructions very close to those in (6). I then discuss a curious pref-

10 Contraposing KK, $\vdash \neg KK\phi \supset \neg K\phi$. By PC, $\vdash (\neg KK\phi \wedge \neg K\neg K\phi) \supset \neg K\phi$. By rule RM for K, $\vdash K(\neg KK\phi \wedge \neg K\neg K\phi) \supset K\neg K\phi$, and so $K(\neg K\neg K\phi \wedge \neg KK\phi) \vdash K\neg K\phi$. But also $K(\neg K\neg K\phi \wedge \neg KK\phi) \vdash \neg K\neg K\phi$ by distribution over the conjunction and factivity. So $K(\neg K\neg K\phi \wedge \neg KK\phi) \vdash \bot$. Given closure and $K\phi \vdash \neg\textsc{open}(\neg\phi)$, $K(\textsc{open}(K\phi) \wedge \textsc{open}(\neg K\phi)) \vdash K(\neg K\neg K\phi \wedge \neg KK\phi)$. So $\ulcorner \textsc{open}(K\phi) \wedge \textsc{open}(\neg K\phi)\urcorner$ can't be known, given KK.

11 Orders of ignorance are the epistemic analogue of 'orders of vagueness' (Williamson 1999).

erence for professing ignorance about what one knows-*wh* rather than professing ignorance about what one knows-*that*.

4.4.1 Natural language data

A 1955 Washington D. C. court case concerning illegal narcotics trade discussed the possibility of uncertainty about what one knows at comic length:[12]

(7) Senator Daniel. Now, I want to go back to Melvin Sutton. You told the committee you did not know a man by the name of Melvin Sutton, by that name.
Mr. Douglas. Not knowingly.
Senator Daniel. What?
Mr. Douglas. Not knowingly.
Senator Daniel: Not knowingly. Well, that is the whole thing; I ask you whether or not you know him. Do you know a man by that name?
Mr. Douglas. I am not sure whether I know him or not.
Senator Daniel: You are not sure?
Mr. Douglas: I am not sure.
Senator Daniel. That you know him?
Mr. Douglas. That is correct.
Senator Daniel. Are you sure you did not know a man by that name?
Mr. Douglas. I am not sure whether I didn't know him.
Senator Daniel: Well, let's see if I can identify him for you, a man by the name of Robert Melvin Sutton, who was arrested at your house, 1452 Fairmont Street, for a narcotic violation on January 17 of this year. Does that recall him to you?
Mr. Douglas. No; it does not.
Senator Daniel. It does not?
Mr. Douglas. No; it does not.

12 *Illicit Narcotics Traffic*, United States Government Printing Office, Washington 1955, p. 1053ff.

> Senator Daniel. Did you ever hear of a Robert Melvin Sutton? I am just asking if you say so.
> Mr. Douglas. I am not saying so because I am not sure.
> Senator Daniel. You are not sure?
> Mr. Douglas. No; I am not.
> Senator Daniel. You are not sure whether you know Robert Melvin Sutton or not?
> Mr. Douglas. No; I am not.

The questioning later returns to the issue if Mr. Douglas knows Melvin Sutton, and Mr. Douglas admits knowing "quite a few Suttons", but doesn't know if any of them are Melvin Sutton. After consulting his counsel, he affirms his position that he does not know if he knows Melvin Sutton. Although this example is especially extreme, there are many other more mundane examples:

(8) I don't know if I know exactly the definition of rigged. — Rudy Giuliani[13]

(9) I still don't know whether I know how to write a sentence.[14]

(10) What if I'm not sure whether I know the password?[15]

(11) Sometimes I have moments where I'm not sure whether I know what I'm doing as a parent.[16]

(12) How Do I Know If I Know? (book on religion)

(13) Q: Ed. W. Young: do you know him? — A: I might; I don't know that I know him. — Q: E. T. Taylor? — A: I don't know whether I know him or not; I know a young boy Taylor but don't know if it is him or not.[17]

(14) How do I know if I know what I'm doing? (title of an abstract in *Gerontologist* 49 (2009): 360.)

13 See https://twitter.com/CNNPolitics/status/792114944963665920.
14 Apparently by Ethan Canin, see https://www.allgreatquotes.com/quote-133768/.
15 https://support.tigertech.net/lost-email-password.
16 https://www.watchingyougrow.co.uk/2015/10/5-reasons-why-i-dont-want-your.html.
17 *Contested-election Case of James I. Campbell V. Robert L. Doughton* from the eight congressional district of Northern Carolina. Washington Government printing office, 1921, p. 1066.

(15) I don't know if I know what I'm doing with this race mod.[18]

(16) I don't know whether I know what I'm wanting to sing or not.[19]

These examples show that it can be fine to express ignorance or uncertainty about what one knows, provided there is sufficient contextual setup. Admittedly, without such setup it is usually odd for people to express ignorance about what they know (cf. Dorr & Hawthorne 2013: fn. 37). But having *some* cases of known ignorance about what one knows is sufficient for my argument. For if KK is true, knowledge of ignorance whether one knows would be *impossible*. Some further examples are noted in a footnote.[20]

18 https://community.playstarbound.com/threads/i-dont-know-if-i-know-what-im-doing-with-this-race-mod.151874/.

19 http://web.lyon.edu/wolfcollection/songs/sutterfieldbarbara1273.html.

20 Here are some more examples to prove my point:

(17) Q: Benjamin Jones; what about Benjamin Jones? — A: I know his father. I don't know whether I know him or not. [...] — Q: And Patrick McDade? — A: I don't know about that; I don't know whether I know him personally or not. (*Contested Election Case of William Connell Vs. George Howell*, Part 2, Washington Printing Office 1903, p. 1503-5.)

(18) Honestly, I am not sure whether I know what's best for Nick. (Ben Price from Coronation Street, see https://www.femalefirst.co.uk/tv/news/ben-prices-easy-coronation-street-return-1167656.html.)

(19) I am not sure whether I know what a common cultural heritage is, but I suspect that it is grounded in the ideas and values of the Renaissance, in the Enlightenment Project and in liberal political thought. (Mrs. Clwyd, House of Commons Jan. 1993, cf. https://publications.parliament.uk/pa/cm199293/cmhansrd/1993-01-19/Debate-2.html.)

(20) I'm not sure whether I know exactly why you see that effect with the crab legs and moon snails. (seehttps://www.sciencebuddies.org/science-fair-projects/ask-an-expert/viewtopic.php?t=13005.)

(21) I am not sure whether I know what their ascertainable wishes and feelings are. (Mrs. Justice Parker in a judgement about four children, see https://www.familylawweek.co.uk/site.aspx?i=ed121127)

(22) I'm not sure whether I know the particular track or not, and I've never seen whatever LP it's on. (https://trojanrecords.com/community/music-music-music-ja-other/gene-rondo/)

(23) [I]t's only been three months and I'm not sure if I know that I'm ready. (https://www.washingtonpost.com/lifestyle/style/ask-amy-pot-smoking-husband-wants-to-toke-freely/2018/11/19/c7daee0a-e85f-11e8-a939-9469f1166f9d_story.html)

4.4.2 Know-that and presuppositions

Curiously, people much more commonly express ignorance about whether they know *someone*, or know-*wh*, or know *the answer* over saying that they don't know if they know *that* p.[21]. This suggests a reply to my argument: Perhaps, although one can know that one doesn't know whether one knows someone, or whether p, it is not possible to know that one doesn't know if one knows that p. In this section, I argue that this response is inadquate, and propose an explanation of the effect.

A first problem with the objection is that sometimes there are known connections between propositional and objectual knowledge or knowledge-*wh*. For example, we could stipulate that you think the password is *secret*, and that you thus know that you know the password *iff* you know that the password is *secret*. In that case, uncertainty about whether you know the password will engender uncertainty about whether you know that the password is *secret*.

A second problem with the objection is KK makes even 'I don't know whether I know *whether* p' a blindspot. We assume that you know whether p *iff* you know either p, or know its negation. KK then entails that if you know whether p, you know that you know whether p.[22]

Independently, there is a natural explanation why speakers avoid saying 'I don't know whether I know *that* ϕ'. 'I know that ϕ' generally presupposes ϕ (Kiparsky & Kiparsky 1971), and thus saying 'I don't know whether I know *that* ϕ' involves presupposing ϕ while saying that one doesn't know whether one knows ϕ. This is very close to asserting a dubious conjunction 'ϕ but I don't know whether I

(24) "You mean start the tractor?" His voice was brisk as he stood up. "Gee whiz! Grampa told us kids to leave the tractor alone. It's dangerous for kids. I don't know whether I know how" (http://print.ditd.org/Academy/Come_On_Wagon.pdf)

(25) I don't know whether I know the answer, basically. (https://academic.oup.com/humrep/article/20/3/810/2356635)

21 Holguín (2019) observes a similar preference for ascriptions of knowledge-*wh* in cases of uncertainty about what one knows, but does not provide an account of the phenomenon.

22 $K_w\phi = K\phi \vee K\neg\phi \vdash KK\phi \vee KK\neg\phi$ by KK. $KK\phi \vee KK\neg\phi \vdash K(K\phi \vee K\neg\phi) = KK_w\phi$ by closure. Clearly, $KK_w\phi \vdash K_wK_w\phi$. Putting everything together, given KK and closure, $\vdash K_w\phi \supset K_wK_w\phi$.

know ϕ' (with the only difference that ϕ is part of the presupposed content, not the asserted content). I propose that speakers prefer to say 'I don't know whether I know *wh* ...' because interrogative complements don't have this problematic presupposition.

To summarise, we prefer to say 'I don't know whether I know' when the knowledge in question is knowledge-*wh*, or knowing someone. This is surprising because known ignorance about knowledge-*wh* engenders known ignorance about knowledge-*that*. The obvious explanation for this phenomenon is that 'I don't know whether I know *that p*' presupposes that *p*, and this presupposition does not sit well with one's asserted ignorance whether one knows *p*.

4.5 Methodological afterword

So far, I have uncritically followed the methodology of assertion-based defences of KK, and argued that it should lead us to the opposite conclusion. It's time to question this methodology. Hintikka (1970: 141), an staunch defender of KK himself, was *extremely* dismissive:

> In discussing whether the KK-thesis holds one finds plenty of material but little guidance in ordinary discourse. In no case can the acceptability of the thesis be decided by appeal to 'ordinary language'.

Natural language data like those quoted above need to be handled with caution. What leads non-philosophers to say, for example, that they don't know whether they know, is not so different from considerations brought up in epistemology. Philosophers have come up with all kinds of counterexamples against KK from the early days onwards, such as Radford (1966)'s unconfident examinee, Danto (1967)'s sceptic who thinks no one knows anything, Lemmon (1959, 1967)'s cases of temporarily inaccessible knowledge, and Danto (1967)'s examples of people who misunderstand what knowledge is. Clearly, there is some temptation, for philosophers and non-philosophers alike, to *say* that these agents know, but don't

know that they know. Many KK-fans are unmoved by such examples. I don't think our ordinary language data should suddenly change their mind.

To illustrate that the kinds of considerations that lead non-philosophers to say, for example, that they don't know whether they know, are not so different from the kinds of considerations philosophers have brought up, let me quote an interview with *Game of Thrones* actor John Bradley from *The Ellen Show*:[23]

(26) Ellen DeGeneres: Do you know how it [i.e. the TV show] ends?
John Bradley: Well I thought I knew how it was going to end, and we definitely shot an ending. But I read an interview with our show runners not too long ago when they said, "The actors think they know how it's going to end." So, you know, actor's paranoia instantly kicked in and you think, "What does that mean?"
E.D.: So you don't know?
J. B.: I don't know what I know, is true.
E. D.: Because he may be just saying that to make you think you don't know, but you may know?
J. B.: I may ultimately know.
E. D.: But you may not know.
J. B.: But I'm not going to know that I know until I know that I know.
E. D.: Right, exactly. I'm with you.

What's motivating John Bradley to say that he does not know what he knows is clearly the nagging higher-order doubts induced by the statement of the show-runners. The case is very close to a number of standard examples in the debate on the role of higher-order counter-evidence.[24] Finding out that Bradley is pulled to say he doesn't know what he knows should not surprise us — epistemologists have had the same judgement about similar hypothetical cases.

23 Cf. https://www.youtube.com/watch?v=qPHlaLti3g0.
24 Cf. Horowitz (2014), Lasonen-Aarnio (2010, 2014), Schoenfield (forthcoming).

If we want guidance on KK from ordinary language, we need to do better. We need to examine examples involving iterated expressions of epistemic modality in a much more systematic way. Quite often, nested possibility modals seem to 'collapse' into a single possibility modal, and necessity modals can be strengthened with another necessity modal:

(27) a. Could this possibly be a pink fairy armadillo?
b. ≈ Could this be a pink fairy armadillo?

(28) a. My eyes are certainly deceiving me.
b. ≈ My eyes must certainly be deceiving me. (Huitink 2012: 404)

However, such data should not be taken to support KK. The semantics literature uniformly analyses such examples as involving *modal concord*. Concord is the phenomenon that several operators occurring in a sentence are interpreted as if there was only one such operator in the sentence. Although there is disagreement about how to analyse modal concord, there is a consensus that it does not arise from an entailment from (27a) to (27b), or from (28a) to (28b). One standard argument to this effect is that modal concord occurs with deontic modals, too, where the 4 axiom ($\Box\phi \supset \Box\Box\phi$) is considered implausible:[25]

(29) a. In this case a resumptive pronoun is optionally allowed for objects and obligatorily required for subjects.[26]
b. ≈ In this case a resumptive pronoun is optional for objects and required for subjects.

Of course, modal concord readings are not always the most natural. To take an example from Huitink (2012: 407), saying that everyone should study logic is very different from saying that everyone should be *obliged* to study logic. A natural direction for examinations of KK in ordinary language would thus be to study similar examples of iterated epistemic modals that disallow modal concord. Such

25 Cf. Geurts & Huitink (2006), Huitink (2012), Zeijlstra (2007).
26 Example originally from Bodomo & Hiraiwa (2004: 58), but I take it from Huitink (2012: 406).

examples have been discussed for example by Moss (2015) and in a recent corpus study by Lassiter (2018). One general difficulty here is that the easiest way to prevent modal concord readings is to use modals of different strengths, or to negate one of them. Here are two examples:

(30) a. The time is now near at hand which must probably determine, whether Americans are to be, Freemen, or Slaves. (G. Washington)[27]

b. "I believe in a 'yes' victory, but I'm definitely not certain," said Prime Minister Goran Persson. (Lassiter 2018: 16)

But it is difficult to get evidence about of KK out of examples like these. The real test cases for KK is usually whether $\Box\phi$ entails $\Box\Box\phi$, or whether $\Diamond\Diamond\phi$ entails $\Diamond\phi$ (or variations of these involving negation). But these are precisely the kinds of cases where modal concord is hard to control for.

Upshot: If ordinary language is to provide us with reliable evidence for or against KK at all, much more work is needed.

4.6 Conclusion

We started with a puzzle for KK-deniers: If KK can fail, nothing prevents one from knowing dubious conjunctions of the form 'p but I don't know whether I know p.' But dubious conjunctions are unassertable. If they can be known, why can't they be asserted? KK-fans conclude that they can't be known, and KK holds.

The main insight of this paper is that KK can't plausibly explain what's wrong with dubious conjunctions. This is because if KK holds, then 'I don't know whether I know p' on its own is a blindspot. Since the argument from dubious conjunctions is one of the most commonly cited considerations in favour of KK, this is a small but significant point.

Let's return briefly to the consequences for higher-order ignorance. Fine (2018) proves given KK that higher-order ignorance permeates upwards. If KK fails, there

27 Address to the Continental Army before the Battle of Long Island, 27 August 1776, quoted by Moss (2015: 4).

is in general no such guarantee that ignorance at a given level leads to ignorance at higher levels. There may still be interesting things to say about the structure of ignorance in particular cases. Take the *transparent* sentence "This sentence is known." If it can be known at all, plausibly what it takes to know the transparent sentence is just what it takes to know that one knows it.[28] So for the transparent sentence, ignorance will still permeate upwards.[29] Or, to take a less paradoxical example, some situations, like the unmarked clock case from Williamson (2011), are commonly modelled in such a way that B ($\phi \supset K\neg K\neg\phi$) is validated. By analogy with a well-known argument from the literature on higher-order vagueness (Williamson 1999), second-order ignorance (or above) can then be shown to induce ignorance at all higher orders.[30]

As quoted in the epigraph at the beginning of the paper, St. Jerome remarks (in a letter to Paulinus) that "it is worse still to be ignorant of your ignorance." One might wonder: Is it even worse to be ignorant of your ignorance of your ignorance? Fine (2018) proves that if KK holds, then any second-order ignorance is necessarily third-order ignorance; any third-order ignorance is necessarily fourth-order ignorance; and so on. If KK holds, then ignorance of ignorance of ignorance is no worse than ignorance of ignorance; they are one and the same thing. If KK fails, on the other hand, ignorance of ignorance is not automatically the worst it can get. If you think ignorance of ignorance is bad, you haven't come across ignorance of ignorance of ignorance. And even if you're in that situation, don't worry, you could be even worse off by being ignorant of your ignorance of your ignorance of your ignorance. And so on. Things could always be worse.

28 What the transparent sentence says is that it is known, and so the only way to know^n the sentence is to know^n that one knows the transparent sentence, i.e. to know^{n+1} the transparent sentence.

29 Of course, such self-referential sentences are problematic in many ways. Cf. Kaplan & Montague (1960), Anderson (1983) for some discussion.

30 Basically, B (plus normality) entail $\vdash K\phi \equiv K\neg K\neg K\phi$, so by failure to know ϕ by normality induces failure whether one knows $\neg K\neg K\phi$. Cf. Williamson (1999), Mahtani (2008), Dorr (2015).

5 Defeating dubious assertions

Abstract KK is the controversial principle that if one knows p, one knows that one knows p. KK-defenders argue that if KK could fail, one could know dubious conjunctions and conditionals of the form 'p but I might not know p' and 'If I don't know that p, p.' But the underlying phenomenon is more general. Whenever p is a defeater for one's knowledge that q, it sounds weird to assert 'q but it might be that p', or 'If p, then q.' I argue that this undermines assertion-based arguments for KK, and suggest a contextualist account.

KK is the controversial principle that if one knows that p, one knows that one knows that p.[1] A popular argument for KK exploits the connection between knowledge and assertion. If KK could fail, nothing would prevent one from knowing dubious conjunctions and conditionals of the form 'p but I might not know p' and 'If I don't know that p, p.' Since dubious conjunctions and conditionals are unassertable, KK-fans conclude that KK cannot fail.[2]

Assertion-based arguments for KK disregard that dubious assertions are an instance of a more general phenomenon: When p is a defeater for one's knowledge that q, it sounds weird to assert 'q but it might be that p', or 'If p, then q.' For example, it is odd to say 'The wall is red but there might be trick lighting' or 'If there is trick lighting, the wall is red' in normal circumstances. Call these *defeat*

1 Recent KK-defences include Stalnaker (1999, 2009c), Greco (2014a, 2017), Goodman & Salow (2018), and Das & Salow (2018); anti-KK arguments include Williamson (2000), Dorr et al. (2014), and Carter (2019).

2 See Sosa (2009a), Smithies (2012), Cohen & Comesaña (2013), Greco (2014b, 2015a, 2017), and Das & Salow (2018) on dubious conjunctions, and Dorst (2019) for abominable conditionals.

conjunctions and *defeat conditionals*. Before jumping to grand conclusions from a narrow range of dubious assertions, we should examine the broader phenomenon. This is the project of this paper.

Here goes the plan for the paper. §5.1 gives a brief exposition of assertion-based arguments for KK. §5.2 argues that dubious assertions are a special case of defeat conjunctions and conditionals. §5.3 argues that the right account of this more general phenomenon must have a contextualist component. §5.4 explores some problems for a contextualist account in this spirit. §5.5 concludes.

5.1 Assertion-based arguments for KK

A familiar argument for the knowledge norm of assertion starts from the observation that Moorean conjunctions of the form 'p but I don't know that p' cannot be felicitously asserted. Of course, many truths are unknown, and so many Moorean conjunctions are true. Crucially, though, Moorean conjunctions cannot be *known* to be true (Hintikka 1962). If one knew 'p but I don't know that p', one would know the first conjunct p, and so the second conjunct would be false and thus unknown, ensuring that the conjunction isn't known. Given that Moorean conjunctions cannot be known, a natural explanation of their infelicity is that assertion in some sense commits one to knowledge:[3]

(KNA) One must: assert p only if one knows p.

For the purposes of this paper, I assume the knowledge account of assertion.

Not just Moorean conjunctions, but also so-called dubious conjunctions of the form 'p but I don't know whether I know that p' sound weird (Sorensen 2000, Sosa 2009a). The underlying phenomenon is robust with respect to minor changes in formulation. For example, saying 'p but I might not know p' is also quite weird.

3 Moore (1962: 277) says that "by asserting p positively, you imply, though you don't assert, that you know that p." See Unger (1975), Gazdar (1979), Williamson (1996, 2000) for arguments for the knowledge norm of assertion along these lines, and Benton (2014) for an overview. Cf. Sorensen (1988) for a more general account of Moorean 'paradoxes' in the same spirit.

(Using examples involving modals is often preferable, because they are easier to compute, and evoke clearer judgements.)

KK plus a knowledge norm for assertion predict these data about iterated Moorean conjunctions. For if KK holds, one can't know 'p but I might not know that p.'[4] And if knowledge is required for assertion, we predict that one can't say 'p but I might not know that p.'

While assuming KK provides an easy account of what's wrong with dubious conjunctions, various authors have suggested that KK-deniers have nothing to say.[5] One subtlety worth noting is that KK is significantly more than is needed. To predict that dubious conjunctions can't be known, all one needs is the principle $K\phi \supset \neg K \neg KK\phi$.[6] Technicalities aside, the general worry is that denying KK will allow us to know things which, on the face of it, cannot be known.

Dorst (2019) presents a variant of the argument from dubious conjunctions that avoids the technical difficulty just mentioned. Instead of conjunctions, Dorst looks at dubious conditionals of the form 'If I don't know that p, then p.' On the one hand, assertions of such conditionals tend to be infelicitous. On the other hand, a relatively plausible stability principle for knowledge entails that dubious conditionals are known in KK-failure scenarios.

K-Stability: If one knows ψ and does not know that $\neg\phi$, one knows $\phi \rightarrow \psi$.[7]

Suppose KK fails, so you know ϕ without knowing that you know ϕ ($K\phi \wedge \neg KK\phi$). Then by K-Stability, you know the conditional that if you don't know ϕ, then ϕ ($K(\neg K\phi \rightarrow \phi)$). So given K-Stability, dubious conditionals are known in KK-

4 $K(\phi \wedge \neg K \neg(\neg K\phi)) \vdash KK\phi \wedge K\neg KK\phi$ by distribition and normality of K. $KK\phi \wedge K\neg KK\phi \vdash KK\phi \wedge K\neg KK\phi \vdash \bot$ by factivity.

5 Cf. Sosa (2009a), Smithies (2012), Cohen & Comesaña (2013), Greco (2014b, 2015a, 2017), and Das & Salow (2018). For responses and discussion, see Williamson (2005, 2013b), Benton (2013), Montminy (2013), Marušić (2013: fn. 13), Holguín (2019), Liu (forthcoming), and San (ms).

6 San (ms) credits Simon Goldstein for this point.

7 Bacon (2015) and Mandelkern & Khoo (2019) point out that if conditionals express different propositions in different contexts, such principles should hold context fixed. So: If the knowledge K^c relevant at a context c leaves open $[\![\phi]\!]^c$ and includes $[\![\psi]\!]^c$, then K^c also includes $[\![\phi \rightarrow \psi]\!]^c$. The contextual parameter is dropped in the main text for readability.

failure scenarios. But since assertions of dubious conditionals sound weird, Dorst (2019) concludes that KK-failure scenarios are impossible.

Upshot: KK-deniers have some explaining to do. Either they have to give us some principled reason to resist K-STABILITY, or they need to say why dubious conditionals are unassertable. And even if K-STABILITY were false, this would do nothing to explain why dubious *conjunctions* are bad. So KK-deniers are ultimately perhaps better advised to give an unified account why dubious conjunctions and conditionals are unassertable albeit knowable.

5.2 Defeat conjunctions and conditionals

My central claim is that the repulsiveness of dubious assertions is a special case of a more general phenomenon: when p is a defeater for one's knowledge that q, it sounds weird to say 'q but it might be that p', or 'If p, then q.' I call the former *defeat conjunctions*, and the latter *defeat conditionals*. I argue that defeat conjunctions and conditionals undermine assertion-based arguments for KK.

Let me give an example (variation of Chisholm 1977: 48):

(WALL) You are standing in front of a red wall in normal conditions. By eyesight, you know that the wall is red. However, you leave open that there is trick lighting. You say:

(1) #The wall is red, but there might be trick lighting.

(2) #If there is trick lighting, the wall is red.

Asserting (1) and (2) in WALL sounds odd. This is puzzling. Surely, if conditions are in fact normal with no trick lighting, you can know that the wall is red simply by looking. And if circumstances are suitable, you will not completely rule out the possibility of trick lighting, and you will be aware that you can't rule it out. But then, how can it be that (1) and (2) sound weird, if you can know either of them?

Slightly more carefully, if you can know that the wall is red, and know that it is compatible with your knowledge that there is trick lighting, then it is unclear what

stands in the way of you knowing the conjunction (1). Similarly, if you know that the wall is red and you leave open that there is trick lighting, then by K-STABILITY you'll know the conditional (2).

Defeat cases like WALL undermine assertion-based arguments for KK. KK-fans like to point out that dubious conjunctions and conditionals are known in KK-failure scenarios. However, for parallel reasons defeat conjunctions and conditionals are known in defeat cases like WALL. And defeat conjunctions and conditionals are unassertable, too. Concluding that WALL is impossible is hardly plausible. Something must have gone wrong. Whatever goes wrong in the bad argument against WALL likely also goes wrong in the standard arguments against KK-failure.

If you like KK, you might object to my description of WALL: if you know that the wall is red and KK holds, you know that you know that the wall is red. You also know that if there was trick lighting, you wouldn't know that the wall is red. So you can conclude that there is no trick lighting, contradicting the case description that you leave open that there is trick lighting.

The objection can be avoided by building similar cases where you can't exclude the relevant possibility because it happens to be realised:

(MANCHESTER) You trust *The Times* and *The Guardian* equally. Reading *The Times*, you know that Manchester won yesterday. You don't know what result the *The Guardian* reported as you haven't checked it. (Hawthorne 2003)

Modifying Hawthorne's case, suppose you go on to assert:

(3) #Manchester won, but *The Guardian* might report that they lost.

(4) #If *The Guardian* reports that Manchester lost, Manchester won.

Assuming that you do know on the basis of the report in *The Times* that Manchester won, and that you know that you don't know what *The Guardian* reported, it is hard to see why you can't say (3). And if you do in fact know that Manchester won from *The Times* and leave open that the *The Guardian* reports otherwise, then by

K-STABILITY you'll know that if *The Guardian* reports that Manchester lost, they won. Crucially, you can't know that *The Guardian* (also) reported that Manchester won on the basis of the report in *The Times*, because in fact *The Guardian* reported that Manchester lost.

Why think that the proposition that one doesn't know that p is a defeater for one's knowledge that p? Here is a quick and dirty argument. If you learned that you don't know, then you'd learn that you'd be violating the knowledge norm if you continued believing, and so you should stop believing. So learning that you don't know ϕ should make you stop believing ϕ.[8] (For a more careful argument, see 2.5.3.)

In conclusion, dubious conjunctions and conditionals seem to be an instance of a more general phenomenon. Whenever p is a defeater for one's knowledge that q, it sounds weird to assert 'q but it might be that p', or 'If p, then q.'

5.3 Why we need contextualism

In this section, I give some reasons to think the right account of defeat conjunctions and conditionals must have a contextualist component.

First, recall WALL: Conditions are normal, you're standing in front of a red wall, but for all you know there is trick lighting (you haven't checked). Observe the following contrast between different things you can say:

(5) a. ✓The wall is red.
b. ✓There might be trick lighting.
c. #The wall is red and there might be trick lighting.
d. #There might be trick lighting and the wall is red.

You can say each conjunct, but you can't say the conjunction. This is the kind of phenomenon that cries out for an account in terms of context-shifting. Clearly, saying one conjunct does something to the context that makes it infelicitous to say the other conjunct.

8 Montminy (2013: 826) argues that justificatied belief in $\neg Kp$ is a defeater for knowledge that p.

Second, note that dubious conditionals exhibit *order effects*. Suppose that we have checked neither *The Times* nor *The Guardian*, but we trust both of them, and to equal extent. Then I might say, in the following order, the two conditionals

(6) a. ✓If *The Times* reports that Manchester won, they won.
b. ✓But if *The Times* reports that Manchester won and *The Guardian* reports that they lost, they might have lost.

Strikingly, the opposite sequences crashes:

(7) a. ✓If *The Times* reports that Manchester won and *The Guardian* reports that they lost, they might have lost.
b. #But if *The Times* reports that Manchester won, they won.

Again, a very natural hypothesis is that something about asserting (6b) that changes the context in such a way that it is not felicitous to say (6a) anymore.

In fact, the underlying phenomenon is quite general. In almost any defeat case, you first get some positive evidence for ϕ that allows you to know ϕ, and then get further, negative evidence that defeats your knowledge that ϕ. I think it is fair to say that it is generally fine to say, before you get any evidence, that

(8) a. ✓If POSITIVE EVIDENCE, then ϕ.
b. ✓But if POSITIVE EVIDENCE and NEGATIVE EVIDENCE, then it might be that $\neg\phi$.

The opposite order generally crashes.

Analogous order effects have been observed for *Sobel sequences* (Sobel 1970), that is counterexamples to antecedent strengthening that sound good in one order, but bad in the other.

Antecedent Strengthening: $A \to C \vdash (A \wedge B) \to C$

While traditional discussed for counterfactuals, Sobel sequences work just as well for indicatives (Willer forthcoming):

(9) a. If Alice comes to the party, it will be fun.
b. But if Alice and Bert come, it will not be fun.
c. But if Charles comes as well, it will be fun...

This sequence seems true if Alice on her own usually makes for a fun party, while Alice and Bert together generally spoil parties, except when Bert is around. Classically, Sobel sequences have been taken to motivate variably strict analyses in the style of Stalnaker (1968), Lewis (1973), and Kratzer (1986), until it was noticed that Sobel sequences cannot felicitously be reversed:[9]

(10) a. If Alice and Bert come to the party, it will not be fun.
b. #But if Alice comes to the party, it will be fun.

The irreversibility of Sobel sequences is generally explained by context-shifting. Some people think that widening the live possibilities is part of the meaning of conditionals, formalised as a context change potential (von Fintel 2001, Gillies 2007, Willer forthcoming), others prefer pragmatic accounts (Moss 2012, Lewis 2018). On the semantic view, the 'opposing' conditionals in reverse Sobel sequences are semantically inconsistent. On the pragmatic view, the 'opposing' conditionals are consistent, but one cannot be warranted in asserting both. Both camps invoke context-shifting. Presumably we should do the same for our Sobel-like sequences.

(On a side-note, our Sobel-like sequences may shed light on the debate between semantic and pragmatic accounts of Sobel sequences. 'Opposing' conditionals in reverse Sobel sequences feel inconsistent because their consequents are inconsistent. My sequences of defeat conditionals show that there are Sobel-like sequences where the consequents are (classically) consistent. For example, "Manchester won" and "Manchester might have lost" can both be true in the same circumstances. Similarly, there seems to be no inconsistency between "If *The Times* reports Manchester won, they won." and "If *The Times* reports Manchester won and *The Guardian* reports otherwise, they might have lost." And yet the sequence of these two conditionals behaves much like a Sobel sequence.)

9 Irene Heim is credited for this observation by von Fintel (2001).

5.4 Towards a positive account

In this section, I transfer accounts of Sobel sequences to see how much work they can do with our 'defeat sequences'. The problems are instructive.

Here is a sketch of the pragmatic account of Sobel sequences due to Moss (2012): By asserting (10a), speakers raise into salience the possibility that both Alice and Bert might come to the party.[10] Further, by saying (10a) speakers establish that if Alice and Bert both come, the party will be terrible. Now, that the possibility of both Alice and Bert coming is salient, and known to lead to a terrible party, I must be able to rule out this possibility in order to be able to assert (10b). Since I am unable to rule out this possibility, I cannot felicitously assert (10b). One bit of positive evidence for this proposal is that when I am able to rule this possibility, I can say (10b). (Imagine I know Bert very well, and he told me he'll avoid Alice in the next few weeks. Now you wonder if you should go to the party, reporting concerns about both Alice and Bob attending. I agree that if somehow Alice and Bert both attend, it'll be a nightmare. But I also assure you that if Alice attends, Bert won't come and the party will be nice. In fact, in this case the reverse sequence sounds better than the non-reversed sequence.)

Why does raising certain possibilities affect what one can assert? Although Moss doesn't do this, it is natural to cash this story out is in terms of a contextualist account about knowledge plus a knowledge norm of assertion. To get the rough idea on the table, suppose knowing p requires the ability to rule out *relevant* possibilities where p is false (Dretske 1970, Stine 1976, Lewis 1996). The reason why raising possibilities affects what you can say is that it affects what you count as "knowing", and speakers agree that you must "know" in order to permissibly assert something. This is *very* sketchy, but it'll do for our purposes.

What happens when we transfer Moss' story about Sobel sequences to our cases involving defeat?[11] One might suggest that you can assert in WALL that the wall is

10 In the case of indicative conditionals, this is naturally construed as an instance of presupposition accommodation, since indicatives are standardly thought to presuppose that their antecedent is compatible with the common ground (Stalnaker 1975, von Fintel 1998, Gillies 2009).

11 See Greco (2017), Neta (2004), Salow (2019) for contextualist accounts of defeat.

red because your eyesight puts you in a position to exclude all ordinarily relevant possibilities where the wall isn't red. You cannot say that *if there is trick lighting*, the wall is red, because the antecedent of this conditional raises the possibility of trick lighting.[12] Among the possibilities where there is trick lighting, there are some where the wall is not red that you can't exclude. So by uttering the antecedent, you change the context to the effect that you do not "know" that the wall is red anymore. Similarly, you can't both say that there might be trick lighting and that the wall is red because saying the former changes the context to the effect that you don't count as "knowing" the latter anymore.

But there is a gap in this account: Assuming that it is originally accepted that the wall is red, why should raising the possibility of lighting introduce possibilities where the wall isn't red, rather than just possibilities where the wall is red but there is also trick lighting? Or take the MANCHESTER case. If we initially accept that Manchester won, why should raising the possibility that *The Times* and *The Guardian* disagree also raise the possibility that Manchester lost? Why doesn't it just raise the possibility that *The Times* and *The Guardian* disagree, but *The Times* is right and Manchester won?

Here is my very tentative answer to this problem: When someone successfully raises the possibility that p, as a matter of fact this seems to have the effect of raising all possibilities that we would leave open upon revising with p, that is raising the possibility that p will raise all the possibilities that we leave open conditional on p. Conditional on there being trick lighting, you leave open that the wall is not red. That's why even merely *raising* the possibility of trick lighting also raises the possibility that the wall isn't red. We might try to codify this in the form of a constraint:

Inheritance: If "Might ϕ" is salient in c, and "If ϕ, might ψ" is accepted in c, then there is strong pressure to accept "Might ψ" in c as well.

12 Again, if indicative conditionals presuppose that their antecedent is a live possibility (Stalnaker 1975, von Fintel 1998, Gillies 2009), this is an instance of accommodation.

The constraint is motivated by examples like this one:

(11) a. Sara might be playing with her dog.
b. If Sara is playing with her dog, she might be outside.
c. #But it's not the case that Sara might be outside.

Inheritance would give us a story about defeat conjunctions. If I say "The wall is red but there might be trick lighting", I clearly make "There might be trick lighting" salient. And the way the case is set up, we will accept "If there is trick lighting, the wall might not be red". By Inheritance, this creates pressure to accept "The wall might not be red". But of course, I can't felicitously *assert* that the wall is red and *accept* that the wall might not be red.

Getting back to dubious conjunctions, the thought is this: It sounds weird to say something of the form 'p but I might not know that p' because saying 'I might not know that p' raises the possibility that I don't know that p into salience. Conditional on the supposition that I don't know p, I leave open $\neg p$. Hence even merely *raising* the possibility that I don't know p in the course of a conversation will raise previously ignored $\neg p$-possibilities into salience as well.

Much is missing from this account both in terms of detail and in terms of explanatory power. In particular, it would be important to have some account of why Inheritance holds. This is a project for another time.

5.5 Conclusion

We started with a puzzle for KK-deniers: If KK can fail, nothing prevents one from knowing dubious conjunctions and conditionals of the form 'p but I might not know p' and 'If I don't know that p, p.' But dubious conjunctions and conditionals are unassertable. If they can be known, why can't they be asserted?

The main insight of this paper is that the underlying phenomenon is more general. Whenever p is a defeater for one's knowledge that q, it sounds weird to assert 'q but it might be that p', or 'If p, then q.' I call these *defeat conjunctions* and *defeat conditionals*.

If the puzzle about dubious assertions is an instance of the puzzle about defeat conjunctions and conditionals, this is a significant result. There are two main lessons. First, a negative lesson: Pace many KK-defenders, dubious assertions are not evidence for KK. Presumably dubious conjunctions and conditionals sound bad for the same reasons as defeat conjunctions and conditionals. Defeat conjunctions and conditionals have nothing to do with knowledge iteration. So we should not explain what's wrong with dubious assertions by appeal to KK.

Second, there is a positive, more tentative lesson. There is some reason to think that the right account of defeat conjunctions and conditionals has to be contextualist. Roughly, you can say in ordinary circumstances that the wall is red because your eyesight puts you in a position to exclude all ordinarily relevant possibilities where the wall isn't red. However, by saying that there might be trick lighting, you raise the possibility of trick lighting into salience, changing the context to the effect that you do not 'know' that the wall is red anymore. This suggests that the right account of dubious conjunctions and conditionals should also be contextualist.

6 Modally qualified counterparts

Abstract KK is the claim that knowing entails knowing that one knows. If at all, epistemologists usually endorse not KK itself, but weakenings. Liu (2020) and San (2019, ms) challenge this strategy: It seems that any satisfactory weakening of KK should at least as strong as $KK^{\Diamond}$, the claim that knowing entails *possibly* knowing that one knows. But $KK^{\Diamond}$ can be shown to be just as implausible as KK. This paper generalises the challenge. Where X is of the form $\phi \supset \psi$, we call $\phi \supset \Diamond\psi$ the modally qualified counterpart $X^{\Diamond}$ of X. Modally qualified counterparts of lots of closure, transmission, and preservation principles are just as implausible as the full-strength principles. However, *pace* Liu (2020, ms) and San (2019, ms), this isn't evidence that all these principles are wrong. Modally qualified counterparts do not constitute a lower bound on satisfactory weakenings.

6.1 Introduction

According to KK, knowing entails knowing that one knows. KK seems false for the boring reason that one might lack the concept of knowledge, fail to consider whether one knows, or believe one knows on a guru's says so. All this is well-known, and the standard response is to endorse not KK itself, but suitable weakenings. Liu (2020) and San (2019, ms) challenge this strategy: It seems that any satisfactory weakening of KK should at least as strong as $KK^{\Diamond}$, the claim that knowing entails *possibly* knowing that one knows. However, $KK^{\Diamond}$ can be shown to be just as implausible as KK. So is there any satisfactory way to weaken KK?

The pessimistic conclusion that there are no satisfactory weakenings of KK is tempting, but premature. Parallel problems infect closure, transmission, and preservation principles. Where X is of the form $\phi \supset \psi$, we call $\phi \supset \Diamond\psi$ the modally

qualified counterpart $X^{\Diamond}$ of X. For a wide range of closure, transmission, and preservation principles, their modally qualified counterparts are just as implausible as the full-strength principles. If successful, arguments from modally qualified counterparts thus prove many (perhaps all) plausible structural constraints on knowledge wrong. Such a radical conclusion seems implausible.

Instead, we should conclude that modally qualified counterpart principles do not constitute a lower bound on satisfactory weakenings of KK, closure, transmission, and preservation principles. These constraints capture our abilities to introspect, deduce from, transmit, or preserve our knowledge. The full-strength principles can fail because we do not always make use of our epistemic abilities. And even the modally qualified counterparts fail because some things can only be known when one does not use one's epistemic abilities. I can know *No one deduces anything* only if I don't deduce *I don't deduce anything*. And if I did deduce this, I would not know *No one deduces anything*. Carefully avoiding such counterfactual trickery, one can weaken KK, closure, transmission, and preservation principles without committing to their modally qualified counterparts. In fact, there are well-known such weakenings in the literature.

Here is the plan for the paper. Section 6.2 reviews common weakenings of KK. Section 6.3 explains why $KK^{\Diamond}$ is implausible, and why this threatens the project of weakening KK. Section 6.4 generalises the threat to closure, transmission, and preservation principles, while section 6.5 provides an alternative route to the underlying formal results. Section 6.6 explains how we can circumvent the threat and weaken KK, closure, transmission, and preservation. Section 6.7 concludes.

6.2 Boring reasons to reject KK

KK is the claim that if one knows that p, one knows that one knows that p. KK has been a bone of contention in epistemology ever since the 1950s, and remains controversial. In the early days, much of the debate evolved around the so-called BK principle — the claim that knowing entails believing that one knows — so

much so that Lehrer (1970) contended the two theses stand and fall together.[1] The later literature has brought out other, more principled reasons to reject KK that do not depend on the possibility of knowing without believing that one knows.[2] For the purposes of this paper, I set these considerations aside, and focus on more traditional reasons to reject KK.

Popular counterexamples to KK and BK include Radford (1966)'s unconfident examinee, who remembers Queen Elizabeth died in 1603, but thinks he is just guessing, Danto (1967)'s sceptic who thinks no one knows anything, but in fact knows she has hands, Lemmon (1959, 1967)'s cases of temporarily inaccessible knowledge, Danto (1967)'s examples of people who do not understand what knowledge is, and, relatedly, agents who lack the concept of knowledge (cf. Alston 1980, Dretske 2004, Feldman 1981).

There is a range of responses available to defenders of KK that vary greatly in ambition. The most ambitious responses deny that the presented cases are counterexamples. Lehrer (1968, 1970), for example, denies that the unconfident examinee knows, and that inaccessible 'knowledge' really is knowledge (cf. Harker 1980: for discussion). Stalnaker (1999) and Greco (2014b), on the other hand, argue that agents who haven't considered whether they know, would not say that they know, or lack the concept of knowledge may still know that they know.[3] We will not consider these strategies any further here.

A somewhat less ambitious line of response it to distinguish different kinds of knowledge, and accept KK only for a demanding form of knowledge. Sosa (2009b),

1 Early papers on KK include Brown (1957), Castañeda (1970), Danto (1967), Gerber (1956), Ginet (1970), Harker (1980), Hilpinen (1970, 1973), Hintikka (1962, 1970), Lehrer (1968, 1970), Lemmon (1959, 1967), Malcolm (1952), Prichard (1950), Radford (1966), Taylor (1955), Woozley (1952). For an overview on earlier discussions of KK, see Hintikka (1962: 106ff.).

2 This is mainly the margin-for-error argument from Williamson (2000). See also Carter (2019), Dorr et al. (2014), Hawthorne & Magidor (2009, 2010), Williamson (2000, 2011) for other objections to KK that don't depend on failures to believe.

3 Cf. Stalnaker (1999: 259): "ALICE knows that BOB does not know [whether ALICE knows that Bob knows that ϕ]. [...] ALICE knows this proposition, even though she does not have anything like the concept of knowledge, or a representation of BOB, or of the proposition. The concepts are the theorist's concepts: they describe, but are not attributed to the participants of the system."

for example, accepts KK for 'reflective knowledge', a certain elevated form of knowledge, but not for 'animal knowledge' (cf. Kornblith 2009: for criticism). Malcolm (1952) and Hintikka (1970), on the other hand, endorse KK for a strong notion of indefeasible knowledge (Hintikka (1970) calls it knowledge on conclusive grounds), and Voorbraak (1990) defends KK for a "rather idealised notion" of 'objective knowledge.' Again, I shall not consider this line of argument here.

A related line of defence is to endorse KK only for a certain subgroup of agents. In philosophy, Shoemaker (1994: 282) defends KK for agents "equal in intelligence, rationality, and conceptual capacity to a normal person". Computer scientists, in contrast, apply epistemic logic mostly to reason about the knowledge of non-human agents, including "thermostats and television-receivers" (Voorbraak 1993). This makes it tempting to react to worries based on human psychological deficiencies by restricting KK to the kinds of idealised non-human agents computer scientists tend to be more interested in.[4] I'm more interested here in versions of KK applicable to ordinary human agents.

The focus of this paper is on a third strategy, perhaps the most popular one, which rejects full-strength KK in favour of some weakening. There are roughly two versions of this strategy. A first way to weaken KK is to strengthen the antecedent, for example to the requirement that one knows and *considers* the proposition that one knows (Chisholm 1977), or to the requirement that one knows and *believes* that one knows (Chisholm 1982, Ginet 1970). A second way to weaken KK is to qualify the consequent, for example requiring that we "can, by reflecting, directly know that we are knowing" (Prichard 1950: 86), or that we are *in a position* to know that we know (Goodman & Salow 2018). In a similar vein, Hilpinen (1970: 119f.) suggests weakening KK to the requirement that being in a position to know implies being in a position to know that one is in a position to know.

The different ways to weaken KK can be, and have been combined. For example, McHugh (2010) defends the claim that if one knows and grasps the proposition that one knows, and "the normal conditions for psychological self-knowledge are

4 Cf. Aucher (2014). Cf. Fagin et al. (1995) for a computer science textbook on epistemic logic.

in place", then one is in a position to know that one knows. And Das & Salow (2018) defend the principle that "for any agent who is able to apply [the rule 'if p, believe you know p'] to the premise that p, if the agent knows that p, she is in a position to know that she knows that p."

Upshot: A standard objection to KK is that one may know without believing that one knows, for example because one is insufficiently confident, lacks the concept of knowledge, has not considered whether one knows, or has unreasonably high standards for knowledge. The most common reaction to this worry is to weaken KK. These alleged counterexamples to KK, and the obvious rejoinders may seem straightforward to the point of being boring. In the next section, I discuss a new challenge for the retreat line of defence.

6.3 KK and KK$^\Diamond$

This section is about a certain challenge to those who wish to retreat from KK to a weakened principle. Let KK$^\Diamond$ be the claim that knowing entails *possibly* knowing that one knows ($K\phi \supset \Diamond KK\phi$). Liu (2020) and San (2019, ms) suggest that KK$^\Diamond$ is weaker than virtually all extant weakenings of KK, and constitutes a lower bound on any satisfactory weakening of KK. However, they also show that KK$^\Diamond$ is just as implausible as KK. The aim of this section is to go through the argument against KK$^\Diamond$ presented by Liu (2020).[5]

Before we get going, let me mention some formal preliminaries. We use a propositional language, extended with one-place operators $\Box$ ('necessarily') and K ('one knows that'). $\Diamond$ ('possibly') is used to abbreviate $\neg\Box\neg$, and M ('for all one knows') abbreviates $\neg K\neg$.

A modal logic in the sense of this paper is a set of sentences containing all tautologies of propositional calculus (PC) closed under modus ponens (MP) and uniform substitution (US). For a given modal logic L, we call ϕ a theorem of L

5 The argument by San (2019, ms) is slightly more complex. The basic point is that in suitably strong logics, if both KK$^\Diamond$ and $M\phi \supset \Diamond KM\phi$ are theorems, then KK and $M\phi \supset KM\phi$ are theorems as well. I think this result is interesting, but it seems somewhat less pertinent as it is unclear why one would want to accept $M\phi \supset \Diamond KM\phi$.

($\vdash_{\mathrm{L}} \phi$) iff $\phi \in \mathrm{L}$. We extend this notation to sets of sentences Γ, writing $\Gamma \vdash_{\mathrm{L}} \psi$ whenever there are $\phi_1, ..., \phi_n \in \Gamma$ such that $\vdash_{\mathrm{L}} (\phi_1 \wedge ... \wedge \phi_n) \supset \psi$. Where the logic L is clear from context, I use $\vdash$. I will often speak loosely of accepting principles, or operators obeying certain axioms. What is at stake is whether we should endorse a logic containing all instances of the corresponding *schema*.

It will be useful to introduce a few axioms and rules for ease of reference:

$$(\mathrm{RN}_{\Box}) \quad \frac{\phi}{\Box\phi} \qquad (\mathrm{K}_{\Box}) \quad \Box(\phi \supset \psi) \supset (\Box\phi \supset \Box\psi)$$

$$(5\mathrm{c}_K) \quad KM\phi \supset M\phi \qquad (\mathrm{KK}) \quad K\phi \supset KK\phi \qquad (\mathrm{RM}_K) \quad \frac{\phi \supset \psi}{K\phi \supset K\psi}$$

$$(\mathrm{M}_K) \quad K(\phi \wedge \psi) \supset (K\phi \wedge K\psi) \qquad (\mathrm{C}_K) \quad (K\phi \wedge K\psi) \supset K(\phi \wedge \psi)$$

As our base logic L, we choose the smallest modal logic containing $\mathrm{K}_{\Box}$ and $5\mathrm{c}_K$ closed under $\mathrm{RN}_{\Box}$. We will consider various extensions of L, where 'L + X_1 + ... + X_n' denotes the smallest extension of L which contains the axioms and is closed under the rules X_1, ..., X_n.

The assumptions built into L are fairly innocuous. $\mathrm{K}_{\Box}$ and $\mathrm{RN}_{\Box}$ ensure that the $\Box$-sublogic of L be normal. $5\mathrm{c}_K$ is equivalent to what's sometimes called the principle of *negative infallibility*, $K\neg K\phi \supset \neg K\phi$. On the interpretation of K in terms of knowledge, $5\mathrm{c}_K$ follows from the fact that what's known is true. But $5\mathrm{c}_K$ is plausible on other interpretations as well. In particular, $5\mathrm{c}_K$ is widely accepted for belief[6] and sometimes for justification.[7] Among other reasons, this is because $5\mathrm{c}_K$ is needed to prohibit Moorean beliefs, that is beliefs which can be true but only if they aren't believed.[8]

Wherever X is a principle of the form $\phi \supset \psi$, we let $\mathrm{X}^{\Diamond}$ be $\phi \supset \Diamond\psi$, and we call $\mathrm{X}^{\Diamond}$ the *modally qualified counterpart* of X (cf. San 2019). For example, the modally

6 See especially Stalnaker (2006) and Rieger (2015), as well as Lenzen (1979), Aucher (2014), and Chalki et al. (2018).

7 Cf. Rosenkranz (2018: 327) and Smithies (2012).

8 To be precise, call ϕ Moorean just in case $\nvdash_{\mathrm{KD}} \phi$ but $\vdash_{\mathrm{KD}} \neg(\phi \wedge K\phi)$. Rieger (2015) shows that an extension L of KD contains $5\mathrm{c}_K$ if $\vdash_{\mathrm{L}} \neg K\phi$ for all Moorean ϕ.

qualified counterpart of KK ($K\phi \supset KK\phi$) is $\mathrm{KK}^{\Diamond}$, the principle $K\phi \supset \Diamond KK\phi$. We occasionally extend this notation to M, the dual of K, such that for example KK^{M} denotes the principle $K\phi \supset MKK\phi$.

The core point of Liu (2020) is that $\mathrm{KK}^{\Diamond}$ is incompatible with *strong* KK-failures, that is cases where one knows something and knows that one does not know that one knows it ($K\phi \wedge K\neg KK\phi$). Strong KK-failures are precisely hte kinds of cases excluded by the principle KK^{M} (that is $K\phi \supset MKK\phi$). The following is then a slight variant of Liu's result:[9]

Fact 6.1 (Variant of Liu (2020))**.** $\mathrm{L} + \mathrm{RM}_K + \mathrm{C}_K^{\Diamond} + \mathrm{M}_K^{\Diamond} + \mathrm{KK}^{\Diamond}$ contains KK^{M}.

Proof. By $\mathrm{C}_K^{\Diamond}$, $K\phi \wedge K\neg KK\phi \vdash \Diamond K(\phi \wedge \neg KK\phi)$. By $\mathrm{KK}^{\Diamond}$ and normality of $\Box$, $\Diamond K(\phi \wedge \neg KK\phi) \vdash \Diamond\Diamond KK(\phi \wedge \neg KK\phi)$. Applying $\mathrm{M}_K^{\Diamond}$ twice (and given normality of $\Box$), $\Diamond\Diamond KK(\phi \wedge \neg KK\phi) \vdash \Diamond\Diamond\Diamond\Diamond(KK\phi \wedge KK\neg KK\phi)$. By $5\mathrm{c}_K$, $K\neg KK\phi \vdash \neg KK\phi$, and so by RM_K (and given normality of $\Box$), $\Diamond\Diamond\Diamond\Diamond(KK\phi \wedge KK\neg KK\phi) \vdash \Diamond\Diamond\Diamond\Diamond(KK\phi \wedge \neg KK\phi) \vdash \Diamond\Diamond\Diamond\Diamond\bot$, and $\Diamond\Diamond\Diamond\Diamond\bot \vdash \bot$ by $\mathrm{RN}_{\Box}$.[10] Putting all implications together, $K\phi \wedge K\neg KK\phi \vdash \bot$. □

The interest of this result lies in the fact that it reveals a conflict between $\mathrm{KK}^{\Diamond}$ and the kinds of boring counterexamples to KK discussed in the previous section. Those counterexamples show that one can know something while failing to believe that one knows it. However, a number of these counterexamples plausibly also show that one might know p while *knowing* that one does not believe that one knows p, and hence while knowing that one does not know that one knows p. Radford (1966)'s unconfident examinee, who thinks she is just guessing, plausibly knows that she does not believe she knows (in fact, she believes she doesn't know). Danto (1967)'s sceptic who thinks no one knows anything thinks she doesn't know anything, and so plausibly knows she does not believe she knows something. And

9 Liu (2020) assumes stronger distribution and agglomeration principles, which allows him to drop $\mathrm{K}_{\Box}$. Since $\mathrm{K}_{\Box}$ is innocuous anyway, weakening distribution and agglomeration seemed preferable. I have weakened factivity (T_K: $K\phi \supset \phi$) to $5\mathrm{c}_K$ ($KM\phi \supset M\phi$), to keep the result applicable to belief. As a result of this, I need RM_K, which Liu (2020)'s proof does not require.

10 $\vdash \neg\bot$ by PC. By $\mathrm{RN}_{\Box}$ then $\vdash \Box\Box\Box\Box\neg\bot$, and so $\vdash \neg\Diamond\Diamond\Diamond\Diamond\bot$ by duality of $\Box$ and $\Diamond$.

Lemmon (1959, 1967)'s people who temporarily can't access their knowledge might well know that they don't believe they know. Liu (2020)'s result tells us that these possibilities are incompatible with $\mathrm{KK}^{\Diamond}$.

In fact, as pointed out by Liu (2020) himself, the result can be strengthened. Writing K^n for n iterations of K, and $\Diamond^n$ for n iterations of $\Diamond$: [11]

Fact 6.2. $\mathrm{L} + \mathrm{RM}_K + \mathrm{C}_K^{\Diamond} + \mathrm{M}_K^{\Diamond} + \mathrm{KK}^{\Diamond}$ contains $K\phi \supset MK^n\phi$ for all n.

Proof. By $\mathrm{C}_K^{\Diamond}$, $K\phi \wedge K\neg K^n\phi \vdash \Diamond K(\phi \wedge \neg K^n\phi)$. By $n-1$ applications of $\mathrm{KK}^{\Diamond}$ and normality of $\Box$, $\Diamond K(\phi \wedge \neg K^n\phi) \vdash \Diamond^n K^n(\phi \wedge \neg K^n\phi)$. Applying $\mathrm{M}_K^{\Diamond}$ n times (and $\Box$-normality), $\Diamond^n K^n(\phi \wedge \neg K^n\phi) \vdash \Diamond^{2n}(K^n\phi \wedge K^n\neg K^n\phi)$. By $5\mathrm{c}_K$, $K\neg K^i\phi \vdash \neg K^i\phi$, and so by RM_K (and given normality of $\Box$), $\Diamond^{2n}(K^n\phi \wedge K^n\neg K^n\phi) \vdash \Diamond^{2n}(K^n\phi \wedge \neg K^n\phi) \vdash \Diamond^{2n}\bot$, and $\Diamond^{2n}\bot \vdash \bot$ by $\mathrm{RN}_{\Box}$. Putting all implications together, $K\phi \wedge K\neg K^n\phi \vdash \bot$. □

Surely, if belief does not automatically iterate, one can know p while knowing for some large n that one does not have n iterations of belief that p. But then, since one knows that knowledge requires belief, one can know that p while knowing for some large n that one does not have n iterations of knowledge that p. The extended result tells us that this possibility is in conflict with $\mathrm{KK}^{\Diamond}$.

I think these results are a serious reason to reject $\mathrm{KK}^{\Diamond}$, and thus any weakening of KK which entails $\mathrm{KK}^{\Diamond}$. In the next section, I show that this challenge to KK extends to a wide range of closure, transmission, and preservation principles.

6.4 Modally qualified counterparts

Recall that wherever X is a principle of the form $\phi \supset \psi$, we call $\phi \supset \Diamond\psi$ the *modally qualified counterpart* $\mathrm{X}^{\Diamond}$ of X. In this section, I show that Liu (2020)'s argument against $\mathrm{KK}^{\Diamond}$ extends to modally qualified counterparts of a wide range of closure, transmission, and preservation principles and rules.

11 To be precise, $K^n\phi = K^{n-1}K\phi$, where $K^0\phi = \phi$, and likewise $\Diamond^n\phi = \Diamond^{n-1}\Diamond\phi$ where $\Diamond^0\phi = \phi$.

For concreteness, my discussion here will be cast in terms of knowledge, but it applies just as much to belief, justification, evidence, or being sure. In particular, I continue to assume only negative infallibility ($5c_K$: $KM\phi \supset M\phi$), not factivity of the K-operator (T_K: $K\phi \supset \phi$), to ensure applicability to non-factive notions such as belief and justification.

6.4.1 Closure

Closure principles require that knowledge is closed under logical consequence, or more broadly under various inference rules. An example is RM_K, the rule that if p entails q, then knowing p entails knowing q ($\phi \supset \psi / K\phi \supset K\psi$). Closure principles are often said to capture the idea that deduction is a way of extending one's knowledge (Williamson 2000: 117).[12]

Just as there are boring counterexamples to KK, there are boring counterexamples to closure principles. Sometimes one knows ϕ, and ϕ entails (in the relevant sense) ψ, but one fails to believe ψ because one does not consider whether ψ, fails to notice the entailment, lacks some concept required to grasp ψ, or for some other reason.[13] And just as for KK, the standard response to boring worries about closure is to retreat to weakenings, for example the principle that if one knows ϕ and knows ϕ entails ψ, one is in a position to know ψ (cf. Cohen 2002: 312; Kvanvig 2006: 260). There are also more interesting reasons to deny closure, which do not depend on failures to believe the consequences of one's knowledge.[14] We shall continue our policy of focusing on the boring reasons here.

Liu (2020)'s challenge from modally qualified counterpart principles carries over to closure principles (cf. Liu ms, San ms). Take for example the montonicity rule RM_K. We can consider a modally qualified variant $RM_K^{\Diamond}$ of RM_K, which says

12 Cf. Hawthorne (2005), Kvanvig (2006) for agreement.

13 According to Chisholm (1963), Pseudo Scotus was already aware of this issue. Cf. Kvanvig (2006), Luper (2020) for overviews. For a defense of full-strength closure, cf. Stalnaker (1984, 1999).

14 Common arguments against closure rely on the lottery and the preface paradox (Kyburg 1961, Makinson 1965), modal conditions on knowledge (Dretske 1970), and risk-aggregation (Lasonen-Aarnio 2008, Schechter 2013). All of these are controversial, cf. Dretske (2005), Hawthorne (2005), Hawthorne & Lasonen-Aarnio (2009), Holliday (2015), Immerman (forthcoming), Tang (2018).

that if p entails q, then knowing p entails *possibly* knowing q:

$$(\mathrm{RM}_K^{\Diamond}) \quad \frac{\phi \supset \psi}{K\phi \supset \Diamond K\psi}$$

It is easy to generalise Liu (2020)'s result to $\mathrm{RM}_K^{\Diamond}$:

Fact 6.3. $\mathrm{L} + \mathrm{C}_K^{\Diamond} + \mathrm{M}_K^{\Diamond} + \mathrm{RM}_K^{\Diamond}$ is closed under RM_K^M:

$$(\mathrm{RM}_K^M) \quad \frac{\phi \supset \psi}{K\phi \supset MK\psi}$$

Proof. Let $\vdash \phi \supset \psi$. By $\mathrm{C}_K^{\Diamond}$, $K\phi \wedge K\neg K\psi \vdash \Diamond K(\phi \wedge \neg K\psi)$. Since $\vdash \phi \supset \psi$, by PC also $\vdash (\phi \wedge \neg K\psi) \supset (\psi \wedge \neg K\psi)$. By $\mathrm{RM}_K^{\Diamond}$, MP, and $\Box$-normality, $\Diamond K(\phi \wedge \neg K\psi) \vdash \Diamond\Diamond K(\psi \wedge \neg K\psi)$. By $\mathrm{M}_K^{\Diamond}$ and $\Box$-normality, $\Diamond\Diamond K(\psi \wedge \neg K\psi) \vdash \Diamond\Diamond\Diamond(K\psi \wedge K\neg K\psi)$. By $5\mathrm{c}_K$ and $\Box$-normality, $\Diamond\Diamond\Diamond(K\psi \wedge K\neg K\psi) \vdash \Diamond\Diamond\Diamond(K\psi \wedge \neg K\psi) \vdash \Diamond\Diamond\Diamond\bot$. By $\mathrm{RN}_{\Box}$ $\Diamond\Diamond\Diamond\bot \vdash \bot$. Putting everything together, $K\phi \wedge K\neg K\psi \vdash \bot$ □

The point of the boring counterexamples to RM_K is that one can know ϕ while failing to believe a consequence ψ. But some of these counterexamples seem to show more, namely that one might know ϕ while knowing that one fails to believe, and thus fails know ψ. For example, one might not just fail to realise that ϕ entails ψ, one might know that one does not believe that ϕ entails ψ, and know that one does not believe ψ on any other grounds. The above result shows that such cases are counterexamples to $\mathrm{RM}_K^{\Diamond}$.[15]

The problem generalises to more complicated closure principles like Heylen's $\mathrm{RLO}_K^{\Diamond}$:

$$(\mathrm{RLO}_K^{\Diamond}) \quad \text{If } \Gamma \vdash \phi, \text{ then } \{K\psi \mid \psi \in \Gamma\} \vdash \Diamond K\phi$$

Just as $\mathrm{KK}^{\Diamond}$ entails KK^M, and $\mathrm{RN}_K^{\Diamond}$ gives us RN_K^M, $\mathrm{RLO}_K^{\Diamond}$ gives us RLO_K^M:

15 For interesting related discussion, see Liu (ms), San (ms).

Fact 6.4. L + $\mathrm{RLO}_K^{\Diamond}$ is closed under RLO_K^M[16]

$$(\mathrm{RLO}_K^M) \quad \text{If } \Gamma \vdash \phi, \text{ then } \{K\psi \mid \psi \in \Gamma\} \vdash MK\phi.$$

Proof. Suppose $\Gamma \vdash \phi$. By PC $\Gamma \cup \{\neg K\phi\} \vdash \phi \wedge \neg K\phi$. By $\mathrm{RLO}_K^{\Diamond}$, $\{K\psi \mid \psi \in \Gamma\} \cup \{K\neg K\phi\} \vdash \Diamond K(\phi \wedge \neg K\phi)$. By $\mathrm{M}_K^{\Diamond}$ and $\Box$-normality, $K(\phi \wedge \neg K\phi) \vdash \Diamond\Diamond(K\phi \wedge K\neg K\phi)$. By $5\mathrm{c}_K$ and $\Box$-normality, $\Diamond\Diamond(K\phi \wedge K\neg K\phi) \vdash \Diamond\Diamond(K\phi \wedge \neg K\phi) \vdash \Diamond\Diamond\bot$. By $\mathrm{RN}_{\Box}$ $\Diamond\Diamond\bot \vdash \bot$. So if $\Gamma \vdash \phi$, then $\{K\psi \mid \psi \in \Gamma\} \cup \{K\neg K\phi\} \vdash \bot$. □

The boring counterexamples to RLO_K are intended to show that one might know all the premises in Γ, and Γ entails ϕ, but fail to believe ϕ, for example because one fails to notice that Γ entails ϕ. But it seems that such examples show more: One might also know all the premises in Γ, and Γ entails ϕ, but one knows that one does not believe, and thus not know ϕ. Our result shows that this is inconsistent with Heylen's $\mathrm{RLO}_K^{\Diamond}$.

Another application of Liu (2020)'s challenge is to modally qualified agglomeration and distribution principles $\mathrm{C}_K^{\Diamond}$ and $\mathrm{M}_K^{\Diamond}$ used by Liu (2020, ms) himself. One might think these principles are more plausible than their modally unqualified versions. Extensions of Liu's argument suggest otherwise. Here is one way to bring this out: If we accept $\mathrm{C}_K^{\Diamond}$ and $\mathrm{M}_K^{\Diamond}$, presumably we will except generalisations with an arbitrary finite number of conjuncts:

$$(n\text{-}\mathrm{M}_K^{\Diamond}) \quad K(\phi_1 \wedge \ldots \wedge \phi_n) \supset \Diamond(K\phi_1 \wedge \ldots \wedge K\phi_n)$$

16

Fact 6.5 (corollary of fact 6.4)**.** If L is closed under $\mathrm{KR}_K^{\Diamond}$, it is closed under KR_K^M.

$$(\mathrm{KR}_K^{\Diamond}) \quad \frac{(\phi_1 \wedge \ldots \wedge \phi_n) \supset \psi}{(K\phi_1 \wedge \ldots \wedge \phi_n) \supset \Diamond K\psi} \quad (n \geq 0) \qquad (\mathrm{KR}_K^M) \quad \frac{(\phi_1 \wedge \ldots \wedge \phi_n) \supset \psi}{(K\phi_1 \wedge \ldots \wedge \phi_n) \supset MK\psi} \quad (n \geq 0)$$

Proof. Since $\Gamma \vdash \psi$ iff there are $\phi_1, \ldots, \phi_n \in \Gamma$ such that $\vdash (\phi_1 \wedge \ldots \wedge \phi_n) \supset \psi$, a modal logic is closed under $\mathrm{KR}_K^{\Diamond}$ iff it is closed under $\mathrm{RLO}_K^{\Diamond}$, and a modal logic is closed under KR_K^M iff it is closed under $\mathrm{RLO}_K^{\Diamond}$. Then our fact follows immediately from fact 6.4. □

$$(n\text{-}\mathrm{C}_K^{\Diamond}) \quad (K\phi_1 \wedge ... \wedge K\phi_n) \supset \Diamond K(\phi_1 \wedge ... \wedge \phi_n)$$

However, just as $\mathrm{KK}^{\Diamond}$ entails KK^M, $\mathrm{RN}_K^{\Diamond}$ makes RN_K^M admissible, and $\mathrm{RLO}_K^{\Diamond}$ makes RLO_K^M admissible, generalised $\mathrm{M}_K^{\Diamond}$ and $\mathrm{C}_K^{\Diamond}$ entail generalised M_K^M and C_K^M:

Fact 6.6. If L contains n-$\mathrm{M}_K^{\Diamond}$ and n-$\mathrm{C}_K^{\Diamond}$ for all n, then it also contains n-M_K^M and n-C_K^M for all n.

Proof. We first establish n-C_K^M. By $(n+1)$-$\mathrm{C}_K^{\Diamond}$, $K\phi_1 \wedge ... \wedge K\phi_n \wedge K\neg K(\phi_1 \wedge ... \wedge \phi_n) \vdash \Diamond K(\phi_1 \wedge ... \wedge \phi_n \wedge \neg K(\phi_1 \wedge ... \wedge \phi_n))$. By 2-$\mathrm{M}_K^{\Diamond}$ and $\Box$-normality, $\Diamond K(\phi_1 \wedge ... \wedge \phi_n \wedge \neg K(\phi_1 \wedge ... \wedge \phi_n)) \vdash \Diamond\Diamond(K(\phi_1 \wedge ... \wedge \phi_n) \wedge K\neg K(\phi_1 \wedge ... \wedge \phi_n))$. By $5\mathrm{c}_K$ and $\Box$-normality, $\Diamond\Diamond(K(\phi_1 \wedge ... \wedge \phi_n) \wedge K\neg K(\phi_1 \wedge ... \wedge \phi_n)) \vdash \Diamond\Diamond(K(\phi_1 \wedge ... \wedge \phi_n) \wedge \neg K(\phi_1 \wedge ... \wedge \phi_n)) \vdash \Diamond\Diamond\bot$. By $\mathrm{RN}_{\Box}$, $\Diamond\Diamond\bot \vdash \bot$. Putting everything together, $K\phi_1 \wedge ... \wedge K\phi_n \wedge K\neg K(\phi_1 \wedge ... \wedge \phi_n) \vdash \bot$.

On to n-M_K^M. By 2-$\mathrm{C}_K^{\Diamond}$, $K(\phi_1 \wedge ... \wedge \phi_n) \wedge K\neg(K\phi_1 \wedge ... \wedge K\phi_n) \vdash \Diamond K(\phi_1 \wedge ... \wedge \phi_n \wedge \neg(K\phi_1 \wedge ... \wedge K\phi_n))$.By $(n+1)$-$\mathrm{M}_K^{\Diamond}$ and $\Box$-normality, $\Diamond K(\phi_1 \wedge ... \wedge \phi_n \wedge \neg(K\phi_1 \wedge ... \wedge K\phi_n)) \vdash \Diamond\Diamond(K\phi_1 \wedge ... \wedge K\phi_n \wedge K\neg(K\phi_1 \wedge ... \wedge K\phi_n))$. By $5\mathrm{c}_K$ and $\Box$-normality, $\Diamond\Diamond(K\phi_1 \wedge ... \wedge K\phi_n \wedge K\neg(K\phi_1 \wedge ... \wedge K\phi_n)) \vdash \Diamond\Diamond(K\phi_1 \wedge ... \wedge K\phi_n \wedge \neg(K\phi_1 \wedge ... \wedge K\phi_n)) \vdash \Diamond\Diamond\bot$. By $\mathrm{RN}_{\Box}$, $\Diamond\Diamond\bot \vdash \bot$. Putting everything together, $K(\phi_1 \wedge ... \wedge \phi_n) \wedge K\neg(K\phi_1 \wedge ... \wedge K\phi_n) \vdash \bot$. □

Perhaps one can know all of $\phi_1, ..., \phi_n$ without believing their conjunction if the number of conjuncts is sufficiently large. If so, presumably one can know that one fails to know the conjunction, for example by knowing that one has not considered whether the long conjunction is true. Similarly, one might think that one can know the long conjunction $\phi_1 \wedge ... \wedge \phi_n$ without believing one of the conjuncts, say because one fails to realise it is one of the conjuncts. Presumably, one could then also know the long conjunction $\phi_1 \wedge ... \wedge \phi_n$ while knowing that one does not believe a certain conjunct. The foregoing result shows that these possibilities are incompatible with the modally qualified principles $\mathrm{M}_K^{\Diamond}$ and $\mathrm{C}_K^{\Diamond}$, generalised to conjunctions of arbitrary length. This again underlines the generality of challenge from modally qualified counterparts.

Some, like Liu (ms), may be willing to reject all versions of closure. But many epistemologists think some version of closure *has* to be true.[17] But those of us unwilling to give up on closure completely must resist arguments from modally qualified counterparts. This point is further underwritten by noting that not just closure, but many other structural constraints on knowledge are infected by similar problems with modally qualified counterparts.

6.4.2 Perfect recall and Preservation

Knowledge recall and preservation principles require that knowledge is retained through time and preserved under revisions. For example, Perfect Recall requires that whatever one knows at one time one still knows at any later time. There are interesting arguments that perfect recall and preservation principles are not even plausible for elephants, idealised agents who never forget (cf. Williamson 2000: 206). But for actual human beings like us, perfect recall and preservation principles fail for the boring reason that we forget, die, lose concepts, or give up beliefs for no good reason. These imperfections motivate weakening perfect recall and preservation. Again, consideration of modally qualified counterparts spells trouble.

Let's start with perfect recall. We amend our language to include time-indexed knowledge operators K_t ('one knows at t that').[18] Consider:

$$(\mathrm{PR}_K) \quad K_t\phi \supset K_{t'}\phi \qquad\qquad (t > t')$$

Instead of accepting PR_K, one might retreat to the modally qualified counterpart $\mathrm{PR}_K^{\Diamond}$:

$$(\mathrm{PR}_K^{\Diamond}) \quad K_t\phi \supset \Diamond K_{t'}\phi \qquad\qquad (t > t')$$

17 Feldman (1981: 487): "the idea that no version of [closure] is true strikes me, and many other philosophers, as one of the least plausible ideas to come down the philosophical pike in recent years." For other prominent endorsements of closure, see e.g. DeRose (1995: 28), Hawthorne (2003, 2005), Williamson (2000, 2009b).

18 Cf. ch. 8 of Fagin et al. (1995) for more on interactions of knowledge and time. For interesting discussion of related issues with memory retention principles, cf. Fraser & Hawthorne (2015).

But as usual, $\mathrm{PR}_K^{\Diamond}$ entails the implausible principle $\mathrm{PR}_K^M = K_t\phi \supset M_t K_{t'}\phi$ $(t > t')$:

Fact 6.7. L + $\mathrm{M}_K^{\Diamond}$ + $\mathrm{C}_K^{\Diamond}$ + $\mathrm{PR}_K^{\Diamond}$ contains PR_K^M.

Proof. By $\mathrm{C}_K^{\Diamond}$, $K_t\phi \wedge K_t\neg K_{t'}\phi \vdash \Diamond K_t(\phi \wedge \neg K_{t'}\phi)$. By t,t'-$\mathrm{PR}_K^{\Diamond}$ and $\Box$-normality, $\Diamond K_t(\phi \wedge \neg K_{t'}\phi) \vdash \Diamond\Diamond K_{t'}(\phi \wedge \neg K_{t'}\phi)$. By $\mathrm{M}_K^{\Diamond}$ and $\Box$-normality, $\Diamond\Diamond K_{t'}(\phi \wedge \neg K_{t'}\phi) \vdash \Diamond\Diamond\Diamond(K_{t'}\phi \wedge K_{t'}\neg K_{t'}\phi)$. By $5\mathrm{c}_K$ and normality of $\Box$, $\Diamond\Diamond\Diamond(K_{t'}\phi \wedge K_{t'}\neg K_{t'}\phi) \vdash \Diamond\Diamond\Diamond(K_{t'}\phi \wedge \neg K_{t'}\phi) \vdash \Diamond\Diamond\Diamond\bot$. By $\mathrm{RN}_{\Box}$, $\Diamond\Diamond\Diamond\bot \vdash \bot$. Combining all this, $K_t\phi \wedge K_t\neg K_{t'}\phi \vdash \bot$. $\Box$

The boring counterexamples to PR_K illustrate that one might know something at one time but fail to believe it at a later time, for example because one has forgotten. But they show more. Plausibly, one can sometimes know that one will forget, thereby knowing something at one time while knowing that one will not know it at some later time. The above fact establishes that this is incompatible with the modally qualified counterpart $\mathrm{PR}_K^{\Diamond}$ of perfect recall.

A similar result pertains to preservation principles. According to preservation principles, knowledge (or other states such as belief or evidence) are preserved under certain kinds of revisions.[19] According to a particularly popular version, knowledge (or belief, or evidence) is preserved under revisions that are consistent with one's knowledge (or beliefs, or evidence). We add an operator K_ψ ('one knows conditional on ψ that') to our language in order to be able to express this principle:

$$(\mathrm{PRES}) \quad (K\phi \wedge M\psi) \supset K_\psi\phi$$

Again, the full-strength principle is hardly plausible for the boring reason that revising with ψ might cause me to forget ϕ, to lose concepts required to grasp ϕ, to die, or to give up belief in ϕ for some other reason. There are also more interesting

19 Preservation has been extensively studied in belief revision (Alchourrón et al. 1985), cf. Lin (2019) for an overview. It is intimately related to the 'or'-to-'if' principle $(K\phi \vee \psi) \wedge \neg K\phi) \supset K(\phi \rightarrow \psi)$ endorsed by Stalnaker (1975) and many others (where '$\rightarrow$' is the indicative conditional), cf. Boylan & Schultheis (msb). Cf. Williamson (2000: ch. 9.7 & 10) on evidence preservation.

reasons to deny preservation, but we shall set those aside here.[20] One might try to retreat to the modally qualified counterpart PRES$^{\Diamond}$:

$$(\text{PRES}^{\Diamond}) \quad (K\phi \wedge M\psi) \supset \Diamond K_{\psi}\phi$$

This move is problematic, for reasons that should be familiar by now. We let L* be the smallest modal logic containing $\text{K}_{\Box}$, 5c_K, and 5c_{K_ψ} and closed under $\text{RN}_{\Box}$. It is easy to verify the following result:

Fact 6.8. L* + $\text{M}^{\Diamond}_{K_\psi}$ + C_K + PRES$^{\Diamond}$ contains PRES^M

$$(\text{PRES}^{M}) \quad (K\phi \wedge M\psi) \supset MK_{\psi}\phi$$

Proof. By by C_K, $K\phi \wedge M\psi \wedge K\neg K_\psi\phi \vdash K(\phi \wedge \neg K_\psi\phi) \wedge M\psi$. By PRES$^{\Diamond}$, $K(\phi \wedge \neg K_\psi\phi) \wedge M\psi \vdash \Diamond K_\psi(\phi \wedge \neg K_\psi\phi)$. By $\text{M}^{\Diamond}_{K_\psi}$ and $\Box$-normality, $\Diamond K_\psi(\phi \wedge \neg K_\psi\phi) \vdash \Diamond\Diamond(K_\psi\phi \wedge K_\psi\neg K_\psi\phi)$. By 5c_{K_ψ} and normality of $\Box$, $\Diamond\Diamond(K_\psi\phi \wedge K_\psi\neg K_\psi\phi) \vdash \Diamond\Diamond(K_\psi\phi \wedge \neg K_\psi\phi) \vdash \Diamond\Diamond\bot$. By $\text{RN}_{\Box}$, $\Diamond\Diamond\bot \vdash \bot$. $\Box$

Plausibly, if one can fail to preserve one's knowledge upon learning new things, one can occasionally know that one will fail to preserve one's knowledge. The present result shows that this possibility is ruled out by PRES$^{\Diamond}$, the modally qualified counterpart of preservation.

Just as in the case of closure, some may be willing to reject recall and preservation principles, but for those of us who think that these principles capture an important structural feature of knowledge, there is reason to resist arguments from modally qualified counterparts.

20 More interesting counterexamples to preservation are for example the composers case (Stalnaker 1994, Lin 2019), defeat cases such as the red wall case (Chisholm 1982: 48), or introspection failure cases ($K\phi \wedge M\neg K\phi \wedge \neg K_{\neg K\phi}\phi$). See chapter 2 for more detailed discussion.

6.4.3 Transmission

Transmission theses describe circumstances under which accepting a speaker's testimony yields knowledge on the part of the hearer. A simple transmission thesis requires that whenever one agent knows p and tells another agent that p, the other agent comes to know p. We slightly amend our language to include indexed knowledge operators K_a ('a knows that') and an operator $T_{a,b}$ ('a tells b that'):

$$(\text{TRANS}) \quad (K_a\phi \wedge T_{a,b}\phi) \supset K_b\phi$$

There may be interesting epistemic reasons to reject TRANS (Fraser 2016), but there are also the usual boring reasons why such an unqualified transmission thesis fails: the hearer might fail to pay attention, pay attention but decide not to believe the testifier, lack concepts required to grasp the testimony, or believe what is being testified on unrelated grounds. It is natural to think that these kinds of worry simply call for further restrictions on transmission.

A moment of reflection shows, however, that the challenge from modally qualified counterparts applies to transmission principles as well. While in this case TRANS$^{\Diamond}$ does not entail TRANS^{M}, note that there are nevertheless plausible counterexamples to TRANS$^{\Diamond}$. Plausibly, the testifier can know that the hearer will fail to pay attention, pay attention but decide not to believe the testifier, lacks concepts required to grasp the testimony, or believe what is being testified on unrelated grounds. So it is plausible that sometimes a knows that p but b does not know p ($K_a(\phi \wedge \neg K_b\phi)$). If a now tells b that p and b does not know that p, then b won't and can't know what a tells her ($\Box(\neg K_b(\phi \wedge \neg K_b\phi))$), and so we have a counterexample to TRANS$^{\Diamond}$.

One complication (Hintikka 1962: 69): Couldn't b come to know that before a told him, it was true that p but b did not know that p?[21] Yes, this is perfectly possible. But let's suppose that what a knows and tells b is not just that b does not

21 Cf. Hintikka (1962: 69): "You may come to know that what I said was true, but saying it in so many words has the effect of making what is being said false." See also van Benthem (2004).

know that p. but further that b will never know that p, and the problem resurfaces. Of course, in all such situations, b will in fact not believe what a tells her (or believe so on unrelated bad grounds), for otherwise a would not know $p \wedge \neg K_b p$. But this does not change the fact that such cases are perfectly possible.

There are interesting connections here to some classic epistemic 'paradoxes'. First, the surprise exam:[22] A teacher tells her student that she will hold a surprise exam on one of the next n days. An exam on day k counts as a surprise *iff* the student does not know on the morning of day k that there will be an exam on day k. The clever student reasons via backwards induction: "The exam can't be on the last day, n, because I would know on the morning of day n that it has to be on day n. Having eliminated n, the exam can't be on the penultimate day, $n-1$, because I would know on the morning of day $n-1$ that it has to be on day $n-1$." Repeating the reasoning, the student concludes the exam can't be on any day.[23] One way to resolve the paradox is to say that the student can't know what the teacher tells her (Quine 1953), or somewhat less radically that the student can't know the announcement on the last day anymore (Sorensen 1984). On the last morning, knowing the announcement would require knowing 'There will be an exam on day n, but I don't know that there will be an exam on day n.' Clearly, this Moorean conjunction can't be known. The lesson these authors draw is that Moorean testimony may be unsuccesful in surprising ways.

Enter also Smullyan's island of knights and knaves, inhabited by knights, who speak only truth, and knaves, who speak only falsehood. A native tells you: "You will never know that I am a knight!" You reason: "If the native is a knave, then what he says is false and I will know that he is a knight, and so he is not a knave. So I know he is a knight. But if the native is a knight, then given that I know he is a night, what he said is false. So he is not a knight. Contradiction." Smullyan (1987: 68ff.) points out that it is not impossible for the native to say what he does, provided the addressee is dead, deaf, or does not listen. However, if the visitor

22 I believe O'Connor (1948) was the first to discuss the surprise exam in print.
23 Cf. Holliday (forthcoming) for a recent formalisation.

is stipulated to believe the testimony, to be logically sophisticated, and informed about the rules of the island, then "it is logically impossible that any native will say to him, "You will never know that I am a knight!"" (Smullyan 1987: 70).

Fraser & Hawthorne (2015) discuss interesting related cases of paradoxical testimony. Suppose a knowledgeable testifier tells you at midnight that *None of your beliefs at midnight amount to knowledge*, and you believe *None of my beliefs at midnight amount to knowledge* on that basis. Given some suitable transmission principle, you come to know what the testifier tells you. On the other hand, if you come to know what the testifier tells you, then what the testifier tells you is false, and so you can't know it. I am tempted by the flat-footed conclusion that stipulating that the testifier knows *and* that you believe what the testifier tells you amounts to stipulating the impossible. We have to be extremely careful in stipulating examples.

Much more could be, and has been said about these paradoxical cases. But whether or not this solves the puzzles about paradoxical testimony, it is surely right that one agent a may know, and tell b that 'p but b does not know that p.' Simply imagine that b does not (and could not easily) believe a. When this happens, strong transmission principles, and their modally qualified counterparts fail.

Upshot: Not just for KK, but for many other structural constraints on knowledge, retreating to their modally qualified counterparts is ineffective. However, it is implausible that all of these structural constraints on knowledge are wrong. I think the conclusion we should draw is that modally qualified counterparts are not a lower bound on plausible weakenings. The next section provides alternative arguments against the modally qualified counterpart principles.

6.5 Actually similar results

This section shows that in sufficiently strong epistemic logics containing an 'actually' operator, modally qualified counterparts entail the full-strength principles.

We augment our language with propositional quantifiers (for any wff ϕ, $\forall p\phi$ is a wff), and add a one-place operator @ ('actually'). As usual, $\exists$ abbreviates $\neg\forall\neg$.

Let L be the smallest logic containing all propositional tautologies (PC), closed under modus ponens (MP), uniform substitution (US), and universal generalisation (UG: $\phi/\forall p\phi$), and containing

$$(\mathrm{T}_\Box)\quad \Box\phi \supset \phi \qquad (\mathrm{K}_\Box)\quad \Box(\phi \supset \psi) \supset (\Box\phi \supset \Box\psi)$$

$$(\mathrm{T}_K)\quad K\phi \supset \phi \qquad (\mathrm{UI})\quad \forall p\phi \supset \phi[\psi/p] \text{ where } p \text{ free for } \psi \text{ in } \phi$$

$$(\mathrm{GEN}_\Box)\quad \Box\phi, \text{ where } \phi \text{ follows from the above axioms and rules}$$

$$(\mathrm{T}_@)\quad @\phi \supset \phi \qquad (\mathrm{RIG}_@)\quad \phi \supset \Box@\phi$$

Using a trick from Yli-Vakkuri & Hawthorne (ms, 2020), let $\alpha = \forall p(p \equiv @p)$. α is a theorem.[24] What's more, α is plausibly known — it requires only that one know that actuality obtains, not what actuality is like. Crucially, however, α is true only in the actual world, and thus possibly known only if actually known:

Fact 6.9 (Collapse Lemma)**.** $\vdash_{\mathrm{L}} \Diamond K^n(\phi \wedge \alpha) \supset K^n(\phi \wedge \alpha)$

Proof. Cf. appendix of Yli-Vakkuri & Hawthorne (2020). □

Given this collapse lemma, one can prove given $K\alpha$ that $\mathrm{KK}^\Diamond$ entails KK:

Fact 6.10. $\mathrm{L} + \mathrm{C}_K^\Diamond + \mathrm{M}_K + \mathrm{KK}^\Diamond + K\alpha$ contains KK.

Proof. By the last assumption, $\vdash K\alpha$. So $K\phi \vdash K\phi \wedge K\alpha$. By $\mathrm{C}_K^\Diamond$, $K\phi \wedge K\alpha \vdash \Diamond K(\phi \wedge \alpha)$. By the collapse lemma, $\Diamond K(\phi \wedge \alpha) \vdash K(\phi \wedge \alpha)$. By $\mathrm{KK}^\Diamond$, $\Diamond K(\phi \wedge \alpha) \vdash \Diamond KK(\phi \wedge \alpha)$. By the collapse lemma, $\Diamond KK(\phi \wedge \alpha) \vdash KK(\phi \wedge \alpha)$. By M_K, $KK(\phi \wedge \alpha) \vdash KK\phi$. Putting everything together $K\phi \vdash KK\phi$, so $\vdash K\phi \supset KK\phi$. □

The same trick can be applied widely, sometimes even without assuming $\vdash K\alpha$. Here are some more examples:

Fact 6.11. $\mathrm{L} + \mathrm{M}_K + \mathrm{RM}_K^\Diamond$ is closed under RM_K.

24 $\vdash @\phi \supset \phi$ by $\mathrm{T}_@$, and $\vdash \phi \supset @\phi$ by $\mathrm{RIG}_@$ and $\mathrm{T}_\Box$, so $\vdash \phi \equiv @\phi$. By UG, $\vdash \forall p(p \equiv @p)$.

Proof. Suppose $\vdash \phi \supset \psi$. Since $\vdash \alpha$,[25] also $\vdash \phi \supset (\psi \wedge \alpha)$ by PC. So by $\mathrm{RM}_K^{\Diamond}$, $K\phi \vdash \Diamond K(\psi \wedge \alpha)$, and $\Diamond K(\psi \wedge \alpha) \vdash K(\psi \wedge \alpha)$ by the collapse lemma. By M_K, $K(\psi \wedge \alpha) \vdash K\psi$. So $K\phi \vdash K\psi$. Hence if $\vdash \phi \supset \psi$ then $\vdash K\phi \supset K\psi$. □

Fact 6.12. $\mathrm{L} + \mathrm{M}_K + \mathrm{RLO}_K^{\Diamond}$ is closed under RLO_K.

Proof. Suppose $\Gamma \vdash \psi$. Since $\vdash \alpha$, also $\Gamma \vdash (\psi \wedge \alpha)$. So by $\mathrm{RLO}_K^{\Diamond}$, $\{K\phi \mid \phi \in \Gamma\} \vdash \Diamond K(\psi \wedge \alpha)$. By the collapse lemma, $\Diamond K(\psi \wedge \alpha) \vdash K(\psi \wedge \alpha)$. By M_K, $K(\psi \wedge \alpha) \vdash K\psi$. So if $\Gamma \vdash \psi$, then $\{K\phi \mid \phi \in \Gamma\} \vdash K\psi$. □

Fact 6.13. If $\mathrm{L} + \mathrm{C}_K^{\Diamond} + \mathrm{M}_K + K_t\alpha$ contains $\mathrm{PR}_K^{\Diamond}$, then it also contains PR_K.

Proof. Let $t \leq t'$. Given $\vdash K_t\alpha$, $K_t\phi \vdash \Diamond K_t(\phi \wedge \alpha)$ by $\mathrm{C}_K^{\Diamond}$. By the collapse lemma, $\Diamond K_t(\phi \wedge \alpha) \vdash K_t(\phi \wedge \alpha)$. By $\mathrm{PR}_K^{\Diamond}$, $K_t(\phi \wedge \alpha) \vdash \Diamond K_{t'}(\phi \wedge \alpha)$. By the collapse lemma, $\Diamond K_{t'}(\phi \wedge \alpha) \vdash K_{t'}(\phi \wedge \alpha)$. By M_K, $K_{t'}(\phi \wedge \alpha) \vdash K_{t'}(\phi)$. So $K_t\phi \vdash K_{t'}(\phi)$. □

These results show that in sufficiently strong logics containing an actually-operator, the modally qualified counterparts of many principles collapse into the full-strength principle. One should only endorse the modally qualified counterparts if one is willing to endorse the full-strength principles, too.

What drives these proofs is that $\alpha = \forall p(p \equiv @p)$ is a theorem, or in some cases even that $K\alpha$ is a theorem. This assumption is controversial. In particular, some have argued that the axiom $\mathrm{T}_@$ ($@\phi \supset \phi$) should not be accepted (Hanson 2006). One might argue that $\mathrm{T}_@$ seems plausible, but only some slight variant is a theorem. For example, in certain two-dimensional systems such as B2D endorsed by Fusco (forthcoming) $@\phi \supset \phi$ itself is not a theorem, but its diagonalisation $\dagger(@\phi \supset \phi)$ is. Once one extends B2D with devices for quantification over propositions, only $\dagger\alpha$ and not α itself will be a theorem. This will not do for our purposes; $\dagger\alpha$ cannot play the same role in our proofs as α. For $\dagger\alpha$ is necessarily true, and so the collapse lemma cannot be proved for $\dagger\alpha$. All this being said, the majority of the literature seems to endorse $\mathrm{T}_@$ (Kaplan 1989, Montague 1970), and a corresponding notion

25 See fn. 24 for a proof.

of real-world validity (Crossley & Humberstone 1977). So our proofs will still appeal to many.

Surprisingly, the T_K axiom is controversial for epistemic logic with actually-operators, too. Sometimes the semantics for K is given by an accessibility relation between *pairs* of worlds, allowing knowledge-operators to 'shift' the actual world parameter.[26] Heylen (2016b) points out that this can generate failures of factivity such as $\Diamond(K(p \equiv @p) \wedge (\neg p \wedge @p))$. Arguably this approach is philosophically misguided. Kaplan (1989) famously argued that there are no, and could be no 'monsters' in natural language, that is operators shifting the context of evaluation for an indexical.[27] On Kaplan's view, the context of utterance fixes the value of an indexical once and for all, and so the value cannot be shifted by logical operators in whose scope the indexical appears. If we think of '@' on the model of indexicals in natural language, then we should avoid a semantics for the K operator that shifts the actual world parameter.[28] I tentatively conclude that we should leave T_K untouched.[29]

Despite some recent interesting work, the interaction of epistemic logic with 'actually'-operators remains elusive, and the above arguments dubious.[30] Luckily we can rely on the weaker, but more secure Fitch-like results against modally qualified counterparts outlined above. Arguments from languages involving 'actually'-operators are just a risky shortcut.

Let me briefly review where we're at. As explained in section 6.2, full-strength KK seems false for boring reasons such as failure to believe that one knows. This is

26 Cf. Fritz (2013), Heylen (2016b), Holliday & Perry (2014), Rabinowicz & Segerberg (1994).

27 Schlenker (2002) argues that there *are* monsters in Amharic. So there is an interesting question here whether our theorising about indexicals is shaped by a problematic over-reliance on English.

28 One *could* think of 'actually' on a different model, e.g. like tense as handled in German and Russian. In these languages, tense under an attitude ascription is determined not by the context of utterance, but by the context of the agent to which the attitude is being ascribed (cf. Schlenker 2002). My point is not that one couldn't think of '@' in this way, but that philosophers generally don't.

29 In discussing how to think of the interation of an apriority operator A with indexicals, Fritz (2013) suggest that "we can understand A as standing for ⌜in all contexts '...' is true⌝." Note that this is not plausible as a way of thinking about the K operator; 'know' is not a meta-linguistic predicate.

30 Again, cf. Fritz (2013), Fusco (forthcoming), Heylen (2016b), Holliday & Perry (2014), Rabinowicz & Segerberg (1994), Yli-Vakkuri & Hawthorne (2020, ms).

why many epistemologists are interested in weakening the KK principle. However, we saw in section 6.3 that satisfactory weakenings of KK may not be available. For $KK^{\Diamond}$ is implausible, and yet anything weaker than $KK^{\Diamond}$ seems too weak to be of interest. In the last two sections, I have generalised the challenge. For a wide range of structural constraints on knowledge, the full-strength constraints are implausible, their modally qualified counterparts are no better, and yet anything weaker than the modally qualified counterparts seems too weak to be of interest. In the next section, I argue that the modally qualified counterparts do not constitute a lower bound on satisfactory weakenings. In fact, consideration of literature shows that people generally endorse versions of KK, closure, transmission, and memory retention principles that do not entail the modally qualified principles.

6.6 Epistemic Abilities

$KK^{\Diamond}$ is not all that unnatural as a rendering of Prichard (1950: 86)'s claim that whenever we know, we "can, by reflecting, directly know that we are knowing." Or take Goodman & Salow (2018)'s claim that whenever we know, we are *in a position* to know that we know. If being in a position to know requires possibly knowing, then Goodman & Salow's version of KK entails $KK^{\Diamond}$. It is somewhat unclear how to interpret 'being in a position to know'.[31]

Nevertheless, most weakenings of KK do not entail $KK^{\Diamond}$. Recall that some people strengthen the antecedent of KK, for example to the requirement that one knows and *considers* the proposition that one knows (Chisholm 1977), or to the

31 Talk of 'being in a position to know' goes back to the early days of epistemic logic, e.g. Hintikka (1962: 118n). The entailment from being in a position to know to possibly knowing is sometimes explicitly endorsed (Heylen 2016a, Kelp & Pedersen 2010), and certainly natural (cf. Yli-Vakkuri & Hawthorne 2020). According to Willard-Kyle (forthcoming) being in a position to know entails that someone in the same epistemic position could know, according to Stanley (2008: 49) that one is "disposed to acquire the knowledge that the proposition is true, when one entertains it on the right evidential basis". However, San (ms) mentions Bernhard Salow resisting this entailment in personal communication, because one may know $p \wedge \neg Kq$ where $p \vdash q$, and thus be in a position to know $q \wedge \neg Kq$ but $\neg\Diamond K(q \wedge \neg Kq)$. Hilpinen (1970: 119) considers a definition of being in a position to know p as the conjunction $Ep \wedge p$, where Ep is interpreted as having "complete evidence that p". On this definition, being in a position will not imply possibly knowing for parallel reasons.

requirement that one knows and *believes* that one knows (Chisholm 1982, Ginet 1970), or to the requirement that one knows and grasps the proposition that one knows, and "the normal conditions for psychological self-knowledge are in place" (McHugh 2010). More generally, one might render such weakenings of KK as follows:

$$\text{(C-KK)} \quad (K\phi \wedge C) \supset KK\phi$$

Neither C-KK nor its necessitation $\Box((K\phi \wedge C) \supset KK\phi)$ entails $\text{KK}^{\Diamond}$, for there is no guarantee that $\Diamond(K\phi \wedge C)$. In fact, whenever ϕ entails $\neg C$, it is guaranteed that $\Box\neg(K\phi \wedge C)$ due to $\Box(K\phi \supset \phi)$ and $\Box\neg(\neg C \wedge C)$. So C-KK evades the argument from modally qualified counterparts. (Liu (2020) and San (ms) claim that principles of the form of C-KK entail KK; I believe this is just a mistake.[32])

A number of other weakenings of KK also don't entail $\text{KK}^{\Diamond}$. Take the requirement that being in a position to know implies being in a position to know that one is in a position to know (Hilpinen 1970: 119f.). This at best entails $K\phi \supset \Diamond K\Diamond K\phi$, which is significantly weaker than $\text{KK}^{\Diamond}$. And Das & Salow's principle that "for any agent who is able to apply [the rule 'if p, believe you know p'] to the premise that p, if the agent knows that p, she is in a position to know that she knows that p." also does not entail $\text{KK}^{\Diamond}$. Even granting that being in a position to know implies possibly knowing, it only entails something of the form $(K\phi \wedge C) \supset \Diamond KK\phi$, which is weaker than C-KK, and so does not entail $\text{KK}^{\Diamond}$. While $\text{KK}^{\Diamond}$ sounds very weak, in practice most of the common weakenings of KK do not seem to entail $\text{KK}^{\Diamond}$.

An analogy may help here. Imagine a camera set up in a room. Provided the camera is turned on, it monitors the room perfectly, nothing escapes its lens. However, the camera can be turned off through a switch in the room. *The unqualified principle* says that if something is a fact about the room, the camera sees it ($\phi \supset S\phi$). The unqualified principle is false for the boring reason that the camera is sometimes turned off. *The modally qualified principle* says that if something is

32 San (ms) argues that we could define an operator $\Diamond_C\phi = C \supset \phi$, and claims that its dual would obey necessitation. This is a mistake. $\neg\Diamond_C\neg\phi = \neg(C \supset \neg\phi) \equiv C \wedge \phi$ clearly does not obey necessitation ($\vdash \phi$ doesn't entail $\vdash C \wedge \phi$ unless C were itself a theorem, in which case it would be an ineffective restriction). Please note that San (ms) is unpublished work.

a fact about the room, the camera *could* see it ($\phi \supset \Diamond S\phi$). The modally qualified principle is also false, because the switch controlling the camera is part of the room. When the switch position is *off*, the camera does not, and could not see that the switch position is *off*. For whenever the switch position is *off*, the camera is off. Conversely, if the camera were turned on, the switch position would not be *off* anymore. The best we can do is the *moderate principle*: If the camera is on and some fact about the room obtains, the camera sees it ($(On \wedge \phi) \supset S\phi$).

I propose that the epistemic abilities we are interested are closely analogous. Take closure, which is often said to capture the ability to extend our knowledge through deduction (Williamson 2000: 117). *Unqualified Closure* says that whenever something follows from your knowledge, you know it. Unqualified closure fails for the boring reason that you do not always make use of your ability for deduction. *Modally qualified closure* says that if something follows from your knowledge, then you *could* know it. Even this principle fails, for the reason that what you deduce, or know, may itself be an object of your knowledge. For example, if I know that no one is deducing anything, I could not come to know that I'm not deducing anything by deduction. If I did deduce that I'm not deducing anything, I would not have known that no one is deducing anything (for it would have been false). The best we can do is the *moderate principle*: If I know something and (competently) deduce something else, I will know what I deduced.

In fact, variations on what we just called *Moderate Closure* are a popular, if not the most popular way to formulate closure principles in epistemology. Here are some examples:[33]

> [K]nowing $p_1, \ldots, p_n$, competently deducing q, and thereby coming to believe q is in general a way of coming to know q. (Williamson 2000: 117)

33 See also Anderson & Hawthorne (forthcoming), Fraser & Hawthorne (2015: 165), Fraser (2016: 2805), Hawthorne (2003), Kvanvig (2006: 261), Lasonen-Aarnio (2008), Williamson (2009a: 2).

> If one knows P and competently deduces Q from P, thereby coming to believe Q, while retaining one's knowledge that P, one comes to know that Q. (Hawthorne 2005: 43)

Similar versions of recall and transmission principles have also been discussed.[34]

> Necessarily: If x knows P and stores P in preservative memory so that at some later time t, x believes P solely on that basis, then x knows P at t. (Fraser & Hawthorne 2015: 167)

> Necessarily, if S knows that p and A comes to believe that p solely on the basis of competent understanding of S's testimony that p and A has good, undefeated evidence (i) that A's understanding was competent and (ii) that S is a reliable testifier, then then A knows that p. (Fraser 2016: 2805)

Perhaps these principles need some more qualifications, but they definitely avoid implying their modally qualified counterpart. We can then see that standard ways to formulate closure, recall, and transmission do not imply $KK^{\Diamond}$. The strategy can be extended to other structural constraints on knowledge. For example:[35]

Moderate KK: Necessarily: If one knows p and competently introspects one's knowledge that p, then one knows that one knows p.

While I am reluctant to endorse any version of KK myself, I recommend Moderate KK to those looking for a weakening of KK. Moderate KK is what one gets if one transfers the standard strategy for weakening closure to the case of KK. Moderate KK thereby naturally goes with a view that assimilates KK to closure principles, as it has recently been defended by Das & Salow (2018).[36]

34 For a slight variation on transmission, cf. Fraser & Hawthorne (2015: 167).

35 Regarding Preservation, one could try: If q is compatible with one's knowledge, and one knows p and stores p in preservative memory so that upon revising with/supposing q, one believes p solely on that basis, then one knows p upon revising with/supposing q.

36 Cf. Immerman (forthcoming) for further parallels between KK and closure. He attempts to argue against closure using margin-for-error principles of the kind Williamson (2000) uses against KK.

Das & Salow themselves go for a slightly different version:

> For any agent who is able to apply [the rule 'if p, believe you know p'] to the premise that p, if the agent knows that p, she is in a position to know that she knows that p. (Das & Salow 2018: 8)

Whether or not this way of spelling out KK is tenable will depend on how one understands the notion of 'being in a position to know' (cf. fn. 31). In particular, one will probably need to single out a notion of 'being in a position to know' that does not entail the possibility of knowing. One might take inspiration from Spencer (2017) here, holding that being able to know p does not require that it is possible that one knows p. This implies the revisionary view that being able to ϕ does not require possibly ϕ-ing. In some cases this is intuitive. If Beethoven writes a sonata about Mozart's death, then perhaps Mozart was able to play the sonata, but he could not possibly have played it. However, before such a notion of being able to know can do much theoretical work, one would need a detailed formally constrained account of the truth-conditions of such ability ascriptions.[37]

Formulating KK in terms of being in a position to know has a number of disadvantages. For one thing, it not so obvious why we should *care* about *being in a position to know*. Knowledge is central to our lives, being in a position to know isn't. For another, common arguments for KK in terms of dubious conjunctions or common knowledge do not obviously carry over to being in a position to know, especially if it is disconnected from knowledge (cf. San ms). And the concept of *being in a position to know* is less clear than the concept of *knowledge* (cf. Yli-Vakkuri & Hawthorne 2020). Moderate KK does better on these fronts.

37 One could try the apparatus of *localised possibility* here, e.g. the dual of *truth in virtue of the essence of x*. Cf. Fine (1994) and Vetter (2015).

6.7 Conclusion

Let's sum up. This paper started out with a challenge to the project of weakening KK: It seems that any satisfactory weakening of KK should be at least as strong as $KK^{\Diamond}$, but $KK^{\Diamond}$ is implausible (Liu 2020, San 2019, ms).

Formally, I have extended Liu's Fitch-like argument from KK to a wide range of closure, transmission, recall, and preservation principles. I have shown that the modally qualified counterparts collapse into the full-strength principles in sufficiently strong logics containing an 'actually' operator. I thus agree with (Liu 2020, ms) and San (2019, ms) that the modally qualified counterparts are just as implausible as the full-strength principles.

In contrast to Liu (2020, ms) and San (2019, ms), I don't think this should lead us to reject all versions of KK, or closure, transmission, preservation, and recall for that matter. The fact that the challenge from modally qualified counterparts applies so widely indicates the opposite: Modally qualified counterparts are more demanding than they seem, and do not constitute a lower bound on satisfactory weakenings.

Modally qualified counterparts are nevertheless interesting. They remind us what KK, closure, transmission, preservation, and recall principles were supposed to capture in the first place, namely our ability to gain knowledge through introspection and deduction, transmit knowledge through testimony, and maintain knowledge through time and under revisions. We are incredibly imperfect, failing to make use of these epistemic abilities all the time. But we must not confuse our failures to apply these abilities with a failure of the abilities themselves. A failure to deduce does not make for a failure of deduction; a failure to introspect does not make for a failure of introspection.

Bibliography

Adams, Ernest. 1965. The logic of conditionals. *Inquiry: An Interdisciplinary Journal of Philosophy* 8(1-4). 166–197. https://doi.org/10.1080/00201746508601430.

Adams, Ernest. 1966. Probability and the logic of conditionals. In Jaakko Hintikka & Patrick Suppes (eds.), *Aspects of inductive logic*, 165–316. Amsterdam: North-Holland.

Adams, Ernest. 1975. *The logic of conditionals* 86. Springer Science & Business Media.

Adler, Jonathan E. 1981. Skepticism and universalizability. *Journal of Philosophy* 78(3). 143–156. https://doi.org/10.2307/2025862.

Alchourrón, Carlos E., Peter Gärdenfors & David Makinson. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50(2). 510–530.

Alston, William P. 1980. Level-confusions in epistemology. *Midwest Studies in Philosophy* 5(1). 135–150. https://doi.org/10.1111/j.1475-4975.1980.tb00401.x.

Anderson, C. Anthony. 1983. The paradox of the knower. *Journal of Philosophy* 80(6). 338–355. https://doi.org/10.2307/2026335.

Anderson, Charity & John Hawthorne. forthcoming. Pragmatic encroachment and closure. In Brian Kim & Matthew McGrath (eds.), *Pragmatic encroachment in epistemology*, Routledge.

Aucher, Guillaume. 2014. Principles of knowledge, belief and conditional belief. In *Interdisciplinary works in logic, epistemology, psychology and linguistics*, 97–134. Springer.

Bacon, Andrew. 2014. Giving your knowledge half a chance. *Philosophical Studies* (2). 1–25. https://doi.org/10.1007/s11098-013-0276-6.

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

Baker-Hytch, Max & Matthew A. Benton. 2015. Defeatism defeated. *Philosophical Perspectives* 29(1). 40–66.

Baltag, Alexandru & Bryan Renne. 2016. Dynamic epistemic logic. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University winter 2016 edn.

Bennett, Jonathan. 2003. *A philosophical guide to conditionals*. Oxford University Press.

van Benthem, J. 2004. What one may come to know. *Analysis* 64(2). 95–105. https://doi.org/10.1093/analys/64.2.95.

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

Benton, Matthew A. 2014. Knowledge norms. In *Internet encyclopedia of philosophy*, https://www.iep.utm.edu/kn-norms/ (6.8.19).

Bergmann, Michael. 2006. *Justification without awareness*. Oxford University Press.

Bird, Alexander & Richard Pettigrew. 2019. Internalism, externalism, and the kk principle. *Erkenntnis* 1–20.

Blumberg, Kyle & Ben Holguín. 2019. Embedded attitudes. *Journal of Semantics* https://doi.org/10.1093/jos/ffz004.

Bodomo, Adams & Ken Hiraiwa. 2004. Relativization in dagaare. *Journal of Dagaare Studies* 4. 53–75.

Boylan, David & Ginger Schultheis. msa. How strong is a counterfactual?

Boylan, David & Ginger Schultheis. msb. The qualitative thesis. Manuscript.

Bradley, Richard. 2007. A defence of the Ramsey test. *Mind* 116(461). 1–21.

Bradley, Richard. 2012. Restricting preservation: A response to Hill. *Mind* 121(481). 147–159.

Bradley, Richard. 2017a. *Decision theory with a human face*. Cambridge University Press.

Bradley, Richard. 2017b. Supporters and underminers: Reply to Chandler. *Mind* 126(502). 603–608.

Brown, Robert. 1957. Not knowing what one knows. *Philosophical Quarterly* 7(27). 151–153. https://doi.org/10.2307/2216962.

Cariani, Fabrizio. forthcoming. On Stalnaker's "Indicative Conditionals". In Louise McNally, Yael Sharvit & Zoltan Szabo (eds.), *Studies in linguistics and philosophy, vol 100*, Springer.

Carter, Sam. 2019. Higher order ignorance inside the margins. *Philosophical Studies* 176(7). 1789–1806.

Carter, Sam. forthcoming. A suppositional theory of conditionals. *Mind* .

Cartwright, J. P. W. 1990. Conditional intention. *Philosophical Studies* 60(3). 233–255. https://doi.org/10.1007/BF00367471.
Castañeda, Hector-Neri. 1970. On knowing (or believing) that one knows (or believes). *Synthese* 21(2). 187–203. https://doi.org/10.1007/BF00413545.
Casullo, Albert. 2018. Pollock and sturgeon on defeaters. *Synthese* 195(7). 2897–2906. https://doi.org/10.1007/s11229-016-1073-5.
Chalki, Aggeliki, Costas D. Koutras & Yorgos Zikos. 2018. A quick guided tour to the modal logic S4.2. *Logic Journal of the IGPL* 26(4). 429–451.
Chandler, Jake. 2013. Defeat reconsidered. *Analysis* 73(1). 49–51. https://doi.org/10.1093/analys/ans129.
Chandler, Jake. 2017. Preservation, commutativity and modus ponens: Two recent triviality results. *Mind* 126(502). 579–602.
Chellas, Brian F. 1980. *Modal logic: An introduction*. Cambridge University Press.
Chisholm, Roderick M. 1963. The logic of knowing. *Journal of Philosophy* 60(25). 773–795. https://doi.org/10.2307/2022834.
Chisholm, Roderick M. 1977. *Theory of knowledge*. Englewood Cliffs, N.J., Prentice-Hall 2nd edn.
Chisholm, Roderick M. 1982. *The foundations of knowing*. Univ of Minnesota Press.
Christensen, David. 2010. Higher order evidence. *Philosophy and Phenomenological Research* 81(1). 185–215. https://doi.org/10.1111/j.1933-1592.2010.00366.x.
Cohen, Stewart. 2002. Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research* 65(2). 309–329. https://doi.org/10.1111/j.1933-1592.2002.tb00204.x.
Cohen, Stewart & Juan Comesaña. 2013. Williamson on Gettier Cases and Epistemic logic. *Inquiry: An Interdisciplinary Journal of Philosophy* 56(1). 15–29.
Colyvan, Mark. 2013. Idealisations in normative models. *Synthese* 190(8). 1337–1350. https://doi.org/10.1007/s11229-012-0166-z.
Cross, Charles B. 1990. Belief revision, non-monotonic reasoning, and the Ramsey test. In Henry E. Kyburg, Ronald P. Loui & Greg N. Carlson (eds.), *Knowledge representation and defeasible reasoning*, 223–244. Kluwer.
Crossley, John N & Lloyd Humberstone. 1977. The logic of "actually". *Reports on Mathematical Logic* 8(1). 1–29.
Danto, Arthur C. 1967. On knowing that we know. In Avrum Stroll (ed.), *Epistemology. new essays in the theory of knowledge*, 32–53. New York: Harper & Row.

Darwiche, Adnan & Judea Pearl. 1997. On the logic of iterated belief revision. *Artificial Intelligence* 89. 1–29.
Das, Nilanjan & Bernhard Salow. 2018. Transparency and the KK Principle. *Noûs* 52(1). 3–23.
Davidson, Donald. 1978. Intending. *Philosophy of History and Action* 11. 41–60.
DeRose, Keith. 1995. Solving the skeptical problem. *Philosophical Review* 104(1). 1–52. https://doi.org/10.2307/2186011.
Dietz, Christina H. ms. Conditional emotions. Unpublished manuscript.
van Ditmarsch, Hans, Wiebe van der Hoek & Barteld Kooi. 2007. *Dynamic epistemic logic*, vol. 337. Springer Science & Business Media.
Ditmarsch, Hans P. Van. 2005. Prolegomena to dynamic logic for belief revision. *Synthese* 147(2). 229–275. https://doi.org/10.1007/s11229-005-1349-7.
Dorr, Cian. 2015. How vagueness could cut out at any order. *Review of Symbolic Logic* 8(1). 1–10. https://doi.org/10.1017/s175502031400032x.
Dorr, Cian, Jeremy Goodman & John Hawthorne. 2014. Knowing against the odds. *Philosophical Studies* 170(2). 277–287.
Dorr, Cian & John Hawthorne. 2013. Embedding epistemic modals. *Mind* 122(488). 867–914. https://doi.org/10.1093/mind/fzt091.
Dorst, Kevin. 2017. Lockeans maximize expected accuracy. *Mind* 128(509). 175–211.
Dorst, Kevin. 2019. Abominable KK failures. *Mind* 128(512). 1227–1259.
Dorst, Kevin. 2020. Evidence: A guide for the uncertain. *Philosophy and Phenomenological Research* 100(3). 586–632. https://doi.org/10.1111/phpr.12561.
Dorst, Kevin. forthcoming. Higher-order uncertainty. In Mattias Skipper & Asbjørn Steglich Petersen (eds.), *Higher-order evidence: New essays*, .
Dretske, Fred. 2004. Externalism and modest contextualism. *Erkenntnis* 61(2-3). 173–186. https://doi.org/10.1007/s10670-004-9277-3.
Dretske, Fred I. 1970. Epistemic operators. *Journal of Philosophy* 67(24). 1007–1023. https://doi.org/10.2307/2024710.
Dretske, Fred I. 2005. “the case against closure". In M. Steup & Earnest Sosa (eds.), *Contemporary debates in epistemology*, 13–25. Malden, Ma: Blackwell.
Drucker, Daniel. 2019. Policy externalism. *Philosophy and Phenomenological Research* 98(2). 261–285. https://doi.org/10.1111/phpr.12425.
Duca, Simone & Hannes Leitgeb. 2012. How serious is the paradox of serious possibility? *Mind* 121(481). 1–36. https://doi.org/10.1093/mind/fzs042.
Edgington, Dorothy. 1995. On conditionals. *Mind* 104(414). 235–329. https://doi.org/10.1093/mind/104.414.235.

Edgington, Dorothy. 2014. Indicative conditionals. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University winter 2014 edn.

Enqvist, Sebastian & Erik J Olsson. 2013. Segerberg on the paradoxes of introspective belief change. In Robert Trypuz (ed.), *Krister segerberg on logic of action*, Dordrecht: Springer.

Fagin, Ronald, Joseph Y. Halpern, Yoram Moses & Moshe Vardi. 1995. *Reasoning about knowledge*. MIT Press.

Feldman, Richard. 1981. Fallibilism and knowing that one knows. *Philosophical Review* 90(2). 266–282. https://doi.org/10.2307/2184442.

Fermé, Eduardo & Sven Ove Hansson. 2018. *Belief change*. Springer.

Ferrero, Luca. 2009. Conditional intentions. *Noûs* 43(4). 700–741. https://doi.org/10.1111/j.1468-0068.2009.00725.x.

Fine, Kit. 1994. Essence and modality. *Philosophical Perspectives* 8. 1–16. https://doi.org/10.2307/2214160.

Fine, Kit. 2018. Ignorance of ignorance. *Synthese* 195(9). 4031–4045. https://doi.org/10.1007/s11229-017-1406-z.

von Fintel, Kai. 1998. The presupposition of subjunctive conditionals. In Uli Sauerland & Orin Percus (eds.), *The interpretive tract*, 29–44. Cambridge, MA: MITWPL.

von Fintel, Kai. 2001. Counterfactuals in a dynamic context. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, Cambridge: MIT Press.

von Fintel, Kai & Anthony S. Gillies. 2010. Must . . . stay . . . strong! *Natural Language Semantics* 18(4). 351–383. https://doi.org/10.1007/s11050-010-9058-2.

Fitelson, Branden. 2013. Gibbard's collapse theorem for the indicative conditional: an axiomatic approach. In *Automated reasoning and mathematics*, 181–188. Springer.

Fitelson, Branden. 2015. The strongest possible lewisian triviality result. *Thought* 4(2). 69–74.

Fraassen, Bas C. 1976. Representational of conditional probabilities. *Journal of Philosophical Logic* 5(3). 417–430.

van Fraassen, Bas C. 1980. Review of Brian Ellis, Rational Belief Systems. *Canadian Journal of Philosophy* 10(3). 497–511.

Fraser, Rachel. 2016. Risk, doubt, and transmission. *Philosophical Studies* 173(10). 2803–2821. https://doi.org/10.1007/s11098-016-0638-y.

Fraser, Rachel Elizabeth & John Hawthorne. 2015. Cretan deductions. *Philosophical Perspectives* 29(1). 163–178. https://doi.org/10.1111/phpe.12070.

Fritz, Peter. 2013. A logic for epistemic two-dimensional semantics. *Synthese* 190(10). 1753–1770. https://doi.org/10.1007/s11229-013-0260-x.
Fuhrmann, André. 1989. Reflective modalities and theory change. *Synthese* 81(1). 115–134.
Fuhrmann, André & Isaac Levi. 1994. Undercutting and the Ramsey test for conditionals. *Synthese* 101(2). 157–169.
Fusco, Melissa. forthcoming. A two-dimensional logic for diagonalization and the a priori. *Synthese* 1–16. https://doi.org/10.1007/s11229-020-02574-7.
Gazdar, Gerald. 1979. *Pragmatics: Implicature, presupposition and logical form*. Academic Press.
Gerber, William. 1956. Is taylor's puzzle genuine. *Analysis* 17(1). 23–24. https://doi.org/10.1093/analys/17.1.23.
Geurts, Bart & Janneke Huitink. 2006. Modal concord. *Concord phenomena and the syntax semantics interface* 15–20.
Gibbard, Allan. 1981. Two recent theories of conditionals. In William Harper, Robert C. Stalnaker & Glenn Pearce (eds.), *Ifs*, 211–247. Reidel.
Gillies, Anthony S. 2006. What might be the case after a change in view. *Journal of Philosophical Logic* 35(2). 117–145. https://doi.org/10.1007/s10992-005-9006-7.
Gillies, Anthony S. 2007. Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3). 329–360.
Gillies, Anthony S. 2009. On truth-conditions for if (but not quite only if). *Philosophical Review* 118(3). 325–349.
Ginet, Carl. 1970. What must be added to knowing to obtain knowing that one knows? *Synthese* 21(2). 163–186. https://doi.org/10.1007/BF00413544.
Goldman, Alvin I. 1976. Discrimination and perceptual knowledge. *Journal of Philosophy* 73(November). 771–791. https://doi.org/10.2307/2025679.
Goldman, Alvin I. 1986. *Epistemology and cognition*. Harvard University Press.
Goldstein, Simon. 2019. A theory of conditional assertion. *Journal of Philosophy* 116(6). 293–318. https://doi.org/10.5840/jphil2019116620.
Goodman, Jeremy & Bernhard Salow. 2018. Taking a chance on KK. *Philosophical Studies* 175(1). 183–196.
Greco, Daniel. 2014a. Could KK Be OK? *Journal of Philosophy* 111(4). 169–197.
Greco, Daniel. 2014b. Iteration and fragmentation. *Philosophy and Phenomenological Research* 88(1). 656–673.
Greco, Daniel. 2015a. Iteration principles in epistemology I: Arguments for. *Philosophy Compass* 10(11). 754–764.

Greco, Daniel. 2015b. Iteration principles in epistemology II: Arguments against. *Philosophy Compass* 10(11). 765–771.
Greco, Daniel. 2015c. Verbal debates in epistemology. *American Philosophical Quarterly* 52(1). 41–55.
Greco, Daniel. 2017. Cognitive mobile homes. *Mind* 126(501). 93–121.
Greco, Daniel. 2019. Is epistemology autonomous? In John McHugh, Jonathan Way & Daniel Whiting (eds.), *Metaepistemology*, Oxford University Press.
Groenendijk, Jeroen, Martin Stokhof & Frank Veltman. 1996. Coreference and modality. In Shalom Lappin (ed.), *Handbook of contemporary semantic theory*, 179–216. Blackwell.
Grove, Adam. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 17(2). 157–170. https://doi.org/10.1007/bf00247909.
Grundmann, Thomas. 2011. Defeasibility theory. In Sven Bernecker & Duncan Pritchard (eds.), *The Routledge companion to epistemology*, 156–166. Routledge.
Gärdenfors, Peter. 1986. Belief revisions and the Ramsey test for conditionals. *Philosophical Review* 95(1). 81–93.
Gärdenfors, Peter. 1988. *Knowledge in flux: Modeling the dynamics of the epistemic states*. Cambridge, MA: MIT Press.
Hájek, Alan. 2003. What conditional probability could not be. *Synthese* 137(3). 273–323. https://doi.org/10.1023/b:synt.0000004904.91112.16.
Hájek, Alan. 2007. My philosophical position says $\ulcorner p \urcorner$ and i don't believe $\ulcorner p \urcorner$. In Mitchell S. Green & John N. Williams (eds.), *Moore's paradox: New essays on belief, rationality, and the first person*, Oxford University Press.
Hall, Michael. 1976. Scepticism and knowing that one knows. *Canadian Journal of Philosophy* 6(4). 655–663. https://doi.org/10.1080/00455091.1976.10716991.
Hanson, William H. 2006. Actuality, necessity, and logical truth. *Philosophical Studies* 130(3). 437–459. https://doi.org/10.1007/s11098-004-5750-8.
Harker, Jay E. 1980. A note on believing that one knows and lehrer's proof that knowledge entails belief. *Philosophical Studies* 37(3). 321–324. https://doi.org/10.1007/BF00372453.
Hawthorne, John. 2003. *Knowledge and lotteries*. OUP.
Hawthorne, John. 2005. The case for closure. In Matthias Steup & Ernest Sosa (eds.), *Contemporary debates in epistemology*, 26–43. Blackwell.
Hawthorne, John & Maria Lasonen-Aarnio. 2009. Knowledge and objective chance. In Patrick Greenough & Duncan Pritchard (eds.), *Williamson on knowledge*, 92–108. Oxford University Press.

Hawthorne, John & Ofra Magidor. 2009. Assertion, context, and epistemic accessibility. *Mind* 118(470). 377–397.
Hawthorne, John & Ofra Magidor. 2010. Assertion and epistemic opacity. *Mind* 119(476). 1087–1105.
Hawthorne, John, Daniel Rothschild & Levi Spectre. 2016. Belief is weak. *Philosophical Studies* 173(5). 1393–1404.
Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9(3). 183–221. https://doi.org/10.1093/jos/9.3.183.
Heylen, Jan. 2015. Closure of a priori knowability under a priori knowable material implication. *Erkenntnis* 80(2). 359–380. https://doi.org/10.1007/s10670-014-9647-4.
Heylen, Jan. 2016a. Being in a position to know and closure. *Thought: A Journal of Philosophy* 5(1). 63–67. https://doi.org/10.1002/tht3.194.
Heylen, Jan. 2016b. Counterfactual theories of knowledge and the notion of actuality. *Philosophical Studies* 173(6). 1647–1673. https://doi.org/10.1007/s11098-015-0573-3.
Hilpinen, Risto. 1970. Knowing that one knows and the classical definition of knowledge. *Synthese* 21(2). 109–132. https://doi.org/10.1007/BF00413541.
Hilpinen, Risto. 1973. Review: E. j. lemmon, if i know, do i know that i know? *Journal of Symbolic Logic* 38(4). 662–662.
Hintikka, Jaakko. 1962. *Knowledge and belief*. Ithaca: Cornell University Press.
Hintikka, Jaakko. 1970. 'knowing that one knows' reviewed. *Synthese* 21(2). 141–162. https://doi.org/10.1007/BF00413543.
Holguín, Ben. 2019. Indicative conditionals and iterative epistemology. *Noûs* Forthcoming.
Holguín, Ben. 2019. Strange knowledge.
Holguín, Ben. ms. Thinking, guessing, and believing.
Holliday, Wesley H. 2015. Epistemic closure and epistemic logic i: Relevant alternatives and subjunctivism. *Journal of Philosophical Logic* 44(1). 1–62. https://doi.org/10.1007/s10992-014-9338-2.
Holliday, Wesley H. forthcoming. Epistemic logic and epistemology. In Sven Ove Hansson Vincent F. Hendricks (ed.), *Handbook of formal philosophy*, Springer.
Holliday, Wesley H. & John Perry. 2014. Roles, rigidity, and quantification in epistemic logic. In Alexandru Baltag & Sonja Smets (eds.), *Johan van benthem on logic and information dynamics*, 591–629. Springer.
Horowitz, Sophie. 2014. Epistemic akrasia. *Noûs* 48(4). 718–744. https://doi.org/10.1111/nous.12026.

Huitink, Janneke. 2012. Modal concord: a case study of dutch. *Journal of semantics* 29(3). 403–437.

Immerman, Daniel. forthcoming. Williamson, closure, and kk. *Synthese* .

Jerzak, Ethan. 2019. Two ways to want? *Journal of Philosophy* 116(2). 65–98. https://doi.org/10.5840/jphil201911624.

Joyce, James M. 1999. *The foundations of causal decision theory*. Cambridge University Press.

Kaplan, D. & R. Montague. 1960. A paradox regained. *Notre Dame Journal of Formal Logic* 1(3). 79–90.

Kaplan, David. 1989. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from kaplan*, 481–563. Oxford University Press.

Karttunen, Lauri. 1972. Possible and must. In *Syntax and semantics volume 1*, 1–20. Brill.

Katsuno, Hirofumi & Alberto O Mendelzon. 1991. Propositional knowledge base revision and minimal change. *Artificial Intelligence* 52(3). 263–294.

Kelp, Christoph9 & Nikolaj Jang Lee Linding Pedersen. 2010. Second-order knowledge. In D. Pritchard & S. Bernecker (eds.), *The routledge companion to epistemology*, Routledge.

Khoo, Justin & Matthew Mandelkern. 2019. Triviality results and the relationship between logical and natural languages. *Mind* 128(510). 485–526. https://doi.org/10.1093/mind/fzy006.

Kiparsky, Paul & Carol Kiparsky. 1971. Fact. In M. Bierwisch & K. E. Heidolph (eds.), *Progress in linguistics*, 143–73. The Hague: Mouton.

Klass, Gregory. 2009. A conditional intent to perform. *Legal Theory* 15(2). 107. https://doi.org/10.1017/S1352325209090089.

Koons, Robert. 2017. Defeasible reasoning. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University winter 2017 edn.

Kornblith, Hilary. 2009. Sosa in perspective. *Philosophical Studies* 144(1). 127–136. https://doi.org/10.1007/s11098-009-9377-7.

Kotzen, Matthew. 2019. A formal account of epistemic defeat. In Cherie Braden, Rodrigo Borges & Branden Fitelson (eds.), *Themes from klein*, Springer Verlag.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Rieser (eds.), *Words, worlds, and contexts*, 38–74. De Gruyter.

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

Kratzer, Angelika. 1991a. Conditionals. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 651–656. Berlin: de Gruyter.

Kratzer, Angelika. 1991b. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 639–650. Berlin: de Gruyter.

Kraus, Sarit, Daniel Lehmann & Menachem Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial intelligence* 44(1-2). 167–207.

Kvanvig, Jonathan L. 2006. Closure principles. *Philosophy Compass* 1(3). 256–267. https://doi.org/10.1111/j.1747-9991.2006.00027.x.

Kyburg, Henry Ely. 1961. *Probability and the logic of rational belief*. Dordrecht: Kluwer.

Lasonen-Aarnio, Maria. 2008. Single premise deduction and risk. *Philosophical Studies* 141(2). 157–173. https://doi.org/10.1007/s11098-007-9157-1.

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

Lasonen-Aarnio, Maria. 2014. Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research* 88(2). 314–345. https://doi.org/10.1111/phpr.12090.

Lassiter, Daniel. 2018. Talking about (quasi-)higher-order uncertainty. In *Tokens of meaning: Papers in honor of Lauri Karttunen*, CSLI Publications.

Lederman, Harvey. 2018. Uncommon knowledge. *Mind* 127(508). 1069–1105. https://doi.org/10.1093/mind/fzw072.

Lehmann, Daniel & Menachem Magidor. 1992. What does a conditional knowledge base entail? *Artificial intelligence* 55(1). 1–60.

Lehrer, Keith. 1968. Belief and knowledge. *Philosophical Review* 77(4). 491–499. https://doi.org/10.2307/2183013.

Lehrer, Keith. 1970. Believing that one knows. *Synthese* 21(2). 133–140. https://doi.org/10.1007/BF00413542.

Lemmon, E. J. 1959. Is there only one correct system of modal logic? i. *Aristotelian Society Supplementary Volume* 33(1). 23–40. https://doi.org/10.1093/aristoteliansupp/33.1.23.

Lemmon, E. J. 1967. If i know, do i know that i know? In Avrum Stroll (ed.), *Epistemology. new essays in the theory of knowledge*, 54–82. New York: Harper & Row.

Lemmon, E. J. 1977. *An introduction to modal logic: The lemmon notes*. Blackwell.

Lenzen, Wolfgang. 1979. Epistemologische Betrachtungen zu S4, S5. *Erkenntnis* 14(1). 33–56.
Levi, Isaac. 1988. Iteration of conditionals and the Ramsey test. *Synthese* 76(1). 49–81.
Lewis, David. 1975. Adverbs of quantification. In Edward L. Keenan (ed.), *Formal semantics of natural language*, 178–188. Cambridge University Press.
Lewis, David. 1987. Probabilities of conditionals and conditional probabilities. In *Philosophical papers volume ii*, OUP.
Lewis, David K. 1973. *Counterfactuals*. Blackwell.
Lewis, David K. 1996. Elusive knowledge. *Australasian Journal of Philosophy* 74(4). 549–567.
Lewis, Karen S. 2018. Counterfactual discourse in context. *Noûs* 52(3). 481–507.
Lin, Hanti. 2019. Belief revision theory. In Richard Pettigrew & Jonathan Weisberg (eds.), *The open handbook of formal epistemology*, 349–396. PhilPapers Foundation.
Lindström, Sten. 1996. The Ramsey test and the indexicality of conditionals. In André Fuhrmann & Hans Rott (eds.), *Logic, action and information*, Berlin: de Gruyter.
Lindström, Sten & Wlodek Rabinowicz. 1999a. Belief change for introspective agents. In *Spinning ideas, electronic essays dedicated to peter gärdenfors on his fiftieth birthday*, .
Lindström, Sten & Wlodek Rabinowicz. 1999b. Ddl unlimited: Dynamic doxastic logic for introspective agents. *Erkenntnis* 50(2-3). 353–385. https://doi.org/10.1023/a:1005577906029.
Liu, Sebastian. 2020. (un) knowability and knowledge iteration. *Analysis* .
Liu, Sebastian. forthcoming. (Un)knowability and knowledge iteration. *Analysis* .
Liu, Sebastian. ms. On a puzzle for closure and related principles.
Luper, Steven. 2020. Epistemic closure. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University summer 2020 edn.
MacDonald, Margaret. 1937. Symposium: Induction and hypothesis i. *Proceedings of the Aristotelian Society, Supplementary Volumes* 16. 20–35.
MacIver, A. M. 1938. Some questions about "know" and "think". *Analysis* 5(3-4). 43–50. https://doi.org/10.2307/3327014.
Mackie, John Leslie. 1973. *Truth, probability and paradox*. Oxford: OUP.
Mahtani, Anna. 2008. Can vagueness cut out at any order? *Australasian Journal of Philosophy* 86(3). 499–508. https://doi.org/10.1080/00048400802001954.

Makinson, David & Peter Gärdenfors. 1991. Relations between the logic of theory change and nonmonotonic logic. In *The logic of theory change*, 183–205. Springer.
Makinson, David C. 1965. The paradox of the preface. *Analysis* 25(6). 205. https://doi.org/10.1093/analys/25.6.205.
Malcolm, Norman. 1952. Knowledge and belief. *Mind* 61(242). 178–189. https://doi.org/10.1093/mind/LXI.242.178.
Malcolm, Norman, Georg Henrik Wright & Ludwig Wittgenstein. 2001. *Ludwig wittgenstein: a memoir*. Oxford University Press.
Mandelkern, Matthew. 2018a. The case of the missing 'if': Accessibility relations in Stalnaker's theory of conditionals. *Semantics and Pragmatics* .
Mandelkern, Matthew. 2018b. Talking about worlds. *Philosophical Perspectives* 32(1). 298–325. https://doi.org/10.1111/phpe.12112.
Mandelkern, Matthew. 2019a. Bounded modality. *Philosophical Review* 128(1). 1–61. https://doi.org/10.1215/00318108-7213001.
Mandelkern, Matthew. 2019b. Practical moore sentences. *Noûs* https://doi.org/10.1111/nous.12287.
Mandelkern, Matthew. forthcoming. A counterexample to modus ponenses. *Journal of Philosophy* .
Mandelkern, Matthew. ms. If p, then p!
Mandelkern, Matthew & Justin Khoo. 2019. Against preservation. *Analysis* 79(3). 424–436.
Marušić, Berislav. 2013. The self-knowledge gambit. *Synthese* 190(12). 1977–1999.
McGee, Vann. 1985. A counterexample to modus ponens. *Journal of Philosophy* 82(9). 462–471. https://doi.org/jphil198582937.
McHugh, Conor. 2010. Self-knowledge and the kk principle. *Synthese* 173(3). 231–257. https://doi.org/10.1007/s11229-008-9404-9.
Meyer, J-J Ch & Wiebe Van Der Hoek. 2004. *Epistemic logic for ai and computer science*, vol. 41. Cambridge University Press.
Montague, Richard. 1970. Universal grammar. *Theoria* 36(3). 373–398. https://doi.org/10.1111/j.1755-2567.1970.tb00434.x.
Montminy, Martin. 2013. Explaining dubious assertions. *Philosophical Studies* 165(3). 825–830.
Moore, George Edward. 1962. *Commonplace book: 1919-1953*. New York: Routledge.
Moretti, Luca & Tommaso Piazza. 2018. Defeaters in current epistemology: Introduction to the special issue. *Synthese* 195(7). 2845–2854.

Moss, Sarah. 2012. On the pragmatics of counterfactuals. *Noûs* 46(3). 561–586.
Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics* 8. 5–1.
Neta, Ram. 2004. Perceptual evidence and the new dogmatism. *Philosophical Studies* 119(1-2). 199–214.
Nozick, Robert. 1981. *Philosophical explanations*. Harvard University Press.
O'Connor, D. J. 1948. Pragmatic paradoxes. *Mind* 57(227). 358–359. https://doi.org/10.1093/mind/LVII.227.358.
Over, David E. 1987. Assumptions and the supposed counterexamples to modus ponens. *Analysis* 47(3). 142–146.
Pacuit, Eric. 2013a. Dynamic epistemic logic i: Modeling knowledge and belief. *Philosophy Compass* 8(9). 798–814. https://doi.org/10.1111/phc3.12059.
Pacuit, Eric. 2013b. Dynamic epistemic logic ii: Logics of information change. *Philosophy Compass* 8(9). 815–833. https://doi.org/10.1111/phc3.12060.
Pollock, John. 1986. *Contemporary theories of knowledge*. Rowman & Littlefield.
Pollock, John & Joe Cruz. 1999. *Contemporary theories of knowledge, 2nd edition*. Rowman & Littlefield.
Popper, Karl. 1959. *The logic of scientific discovery*. Routledge.
Prichard, H. A. 1950. *Knowledge and perception*. Oxford University Press.
Quine, W. V. 1953. On a so-called paradox. *Mind* 62(245). 65–67. https://doi.org/10.1093/mind/LXII.245.65.
Rabinowicz, Wlodek & Krister Segerberg. 1994. Actual truth, possible knowledge. *Topoi* 13(2). 101–115. https://doi.org/10.1007/BF00763509.
Rabinowicz, Wlodzimierz. 1996. Stable revision, or is preservation worth preserving? In Andre Fuhrmann & Hans Rott (eds.), *Logic, action and information*, 101–28. Berlin: de Gruyter.
Radford, Colin. 1966. Knowledge—by examples. *Analysis* 27(1). 1–11. https://doi.org/10.1093/analys/27.1.1.
Ramsey, Frank Plumpton. 1931. *The foundations of mathematics and other logical essays*. Routledge and Kegan Paul.
Rényi, Alfred. 1970. *Foundations of probability*. San Francisco: Holden-Day.
Rieger, Adam. 2015. Moore's paradox, introspection and doxastic logic. *Thought: A Journal of Philosophy* 4(4). 215–227. https://doi.org/10.1002/tht3.181.
Rosenkranz, Sven. 2018. The structure of justification. *Mind* 127(506). 629–629. https://doi.org/10.1093/mind/fzx039.
Rothschild, Daniel. forthcoming. What it takes to believe. *Philosophical Studies* .
Rothschild, Daniel & Levi Spectre. 2018a. At the threshold of knowledge. *Philosophical Studies* 175(2). 449–460.

Rothschild, Daniel & Levi Spectre. 2018b. A puzzle about knowing conditionals. *Noûs* 52(2). 473–478.
Rott, Hans. 1989. Conditionals and theory change: Revisions, expansions, and additions. *Synthese* 81(1). 91–113.
Rott, Hans. 2011. Reapproaching Ramsey: Conditionals and iterated belief change in the spirit of AGM. *Journal of Philosophical Logic* 40(2). 155–191.
Rott, Hans. 2017. Preservation and postulation: Lessons from the new debate on the Ramsey test. *Mind* 126(502). 609–626.
Russell, Bertrand. 1912. *The problems of philosophy*. OUP Oxford.
Salow, Bernhard. 2019. Elusive externalism. *Mind* 128(510). 397–427.
Samet, Dov. 1997. On the Triviality of High-Order Probabilistic Beliefs. Tech. Rep. 9705001 University Library of Munich, Germany. https://ideas.repec.org/p/wpa/wuwpga/9705001.html.
Samet, Dov. 1998. Iterated Expectations and Common Priors. *Games and Economic Behavior* 24(1-2). 131–141. https://ideas.repec.org/a/eee/gamebe/v24y1998i1-2p131-141.html.
San, Weng. 2019. Disappearing diamonds: Fitch-like results in bimodal logic. *Journal of Philosophical Logic* 48(6). 1003–1016. https://doi.org/10.1007/s10992-019-09504-0.
San, Weng Kin. ms. KK, knowledge, knowability.
Santorio, Paolo. 2018. Credence for epistemic discourse. Ms.
Schechter, Joshua. 2013. Rational self-doubt and the failure of closure. *Philosophical Studies* 163(2). 428–452. https://doi.org/10.1007/s11098-011-9823-1.
Schlenker, Philippe. 2002. A plea for monsters. *Linguistics and Philosophy* 26(1). 29–120. https://doi.org/10.1023/A:1022225203544.
Schoenfield, Miriam. forthcoming. Meditations on beliefs formed arbitrarily. In John Hawthorne Tamar Szabó Gendler (ed.), *Oxford studies in epistemology 7*, Oxford: OUP.
Segerberg, Krister. 2006. Moore problems in full dynamic doxastic logic. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 91(1). 95–110.
Shoemaker, Sydney. 1994. Self-knowledge and "inner sense" lecture ii: The broad perceptual model. *Philosophy and Phenomenological Research* 54(2). 271–290. https://doi.org/ppr199454299.
Shoemaker, Sydney. 1995. Moore's paradox and self-knowledge. *Philosophical Studies* 77(2-3). 211–28. https://doi.org/10.1007/bf00989570.
Silk, Alex. 2017. How to embed an epistemic modal: Attitude problems and other defects of character. *Philosophical Studies* 174(7). 1773–1799. https://doi.org/10.1007/s11098-016-0827-8.

Smithies, Declan. 2012. Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research* 85(2). 273–300.
Smullyan, Raymond M. 1987. *Forever undecided: A puzzle guide to gödel*. Oxford University Press.
Sobel, J. Howard. 1970. Utilitarianisms: Simple and general. *Inquiry* 13(1-4). 394–449. https://doi.org/10.1080/00201747008601599.
Sorensen, Roy. 2000. Moore's problem with iterated belief. *Philosophical Quarterly* 50(198). 28–43. https://doi.org/10.1111/1467-9213.00165.
Sorensen, Roy A. 1984. Conditional blindspots and the knowledge squeeze: A solution to the prediction paradox. *Australasian Journal of Philosophy* 62(2). 126–135. https://doi.org/10.1080/00048408412341321.
Sorensen, Roy A. 1988. *Blindspots*. Oxford: Oxford University Press.
Sosa, David. 2009a. Dubious assertions. *Philosophical Studies* 146(2). 269–272.
Sosa, Ernest. 2009b. *A virtue epistemology: Apt belief and reflective knowledge, volume i*. Oxford University Press.
Spencer, Jack. 2017. Able to do the impossible. *Mind* 126(502). 466–497. https://doi.org/10.1093/mind/fzv183.
Stalnaker, Robert. 1968. A Theory of Conditionals. In W. L. Harper, R. Stalnaker & G. Pearce (eds.), *Studies in logical theory, american philosophical quarterly*, 98–112. Blackwell.
Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5(3). 269–286.
Stalnaker, Robert. 1984. *Inquiry*. Cambridge University Press.
Stalnaker, Robert. 1994. What is a nonmonotonic consequence relation? *Fundamenta Informaticae* 21(1, 2). 7–21.
Stalnaker, Robert. 1999. *Context and content*. Oxford: OUP.
Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25(5-6). 701–721. https://doi.org/10.1023/a:1020867916902.
Stalnaker, Robert. 2006. On logics of knowledge and belief. *Philosophical Studies* 128(1). 169–199.
Stalnaker, Robert. 2009a. Conditional propositions and conditional assertions. In Andy Egan & B. Weatherson (eds.), *Epistemic modality*, Oxford University Press.
Stalnaker, Robert. 2009b. Iterated belief revision. *Erkenntnis* 70(2). 189–209.
Stalnaker, Robert. 2009c. On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility. *Mind* 118(470). 399–409.
Stalnaker, Robert. 2015. Luminosity and the kk thesis. In Sanford Goldberg (ed.), *Externalism, self-knowledge, and skepticism: New essays*, 19–40. Cambridge University Press.

Stanley, Jason. 2008. Knowledge and certainty. *Philosophical Issues* 18(1). 35–57. https://doi.org/10.1111/j.1533-6077.2008.00136.x.
Stine, G. C. 1976. Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies* 29(4). 249–261.
Stojnić, Una. 2017. One's modus ponens: Modality, coherence and logic. *Philosophy and Phenomenological Research* 95(1). 167–214. https://doi.org/10.1111/phpr.12307.
Stroud, Barry. 1984. *The significance of philosophical scepticism*. Oxford University Press.
Sturgeon, Scott. 2014. Pollock on defeasible reasons. *Philosophical Studies* 169(1). 105–118. https://doi.org/10.1007/s11098-012-9891-x.
Tang, Weng. 2018. In defence of single-premise closure. *Philosophical Studies* 175(8). 1887–1900. https://doi.org/10.1007/s11098-017-0938-x.
Taylor, Richard. 1955. Knowing what one knows. *Analysis* 16(3). 63–65. https://doi.org/10.1093/analys/16.3.63.
Unger, Peter. 1975. *Ignorance: A case for scepticism*. Oxford University Press.
Vetter, Barbara. 2015. *Potentiality: From dispositions to modality*. Oxford University Press.
Voorbraak, F. 1993. *As far as i know. epistemic logic and uncertainty*: Universiteit Utrecht Utrecht dissertation.
Voorbraak, Frans. 1990. The logic of objective knowledge and rational belief. In *European workshop on logics in artificial intelligence*, 499–515. Springer.
Willard-Kyle, Christopher. forthcoming. Being in a position to know is the norm of assertion. *Pacific Philosophical Quarterly* .
Willer, Malte. 2010. New surprises for the Ramsey test. *Synthese* 176(2). 291–309.
Willer, Malte. forthcoming. Lessons from sobel sequences. *Semantics and Pragmatics* .
Williams, John N. 2007. Moore's paradoxes and iterated belief. *Journal of Philosophical Research* 32. 145–168. https://doi.org/10.5840/jpr20073236.
Williams, John N. 2015a. Moore's paradox in thought: A critical survey. *Philosophy Compass* 10(1). 24–37. https://doi.org/10.1111/phc3.12187.
Williams, John N. 2015b. Moore's paradox in speech: A critical survey. *Philosophy Compass* 10(1). 10–23. https://doi.org/10.1111/phc3.12188.
Williamson, Timothy. 1996. Knowing and asserting. *Philosophical Review* 105(4). 489.
Williamson, Timothy. 1999. On the structure of higher-order vagueness. *Mind* 108(429). 127–143. https://doi.org/10.1093/mind/108.429.127.
Williamson, Timothy. 2000. *Knowledge and its limits*. OUP.

Williamson, Timothy. 2005. Contextualism, subject-sensitive invariantism, and knowledge of knowledge. *Philosophical Quarterly* 55(219). 213–235.
Williamson, Timothy. 2007. How probable is an infinite sequence of heads? *Analysis* 67(3). 173–180. https://doi.org/10.1093/analys/67.3.173.
Williamson, Timothy. 2009a. Probability and danger. In *Amherst lecture in philosophy*, 1–35.
Williamson, Timothy. 2009b. Reply to john hawthorne and maria lasonen-aarnio, 313–29. Oxford: OUP.
Williamson, Timothy. 2011. Improbable knowing. In T. Dougherty (ed.), *Evidentialism and its discontents*, OUP.
Williamson, Timothy. 2013a. Gettier cases in epistemic logic. *Inquiry: An Interdisciplinary Journal of Philosophy* 56(1). 1–14. https://doi.org/10.1080/0020174X.2013.775016.
Williamson, Timothy. 2013b. Response to Cohen, Comesaña, Goodman, Nagel, and Weatherson on Gettier Cases in Epistemic Logic. *Inquiry* 56(1). 77–96.
Williamson, Timothy. 2017. Model-building in philosophy. In R. Blackford & D. Broderick (eds.), *Philosophy's future: The problem of philosophical progress*, Oxford: Wiley.
Williamson, Timothy. 2019. Evidence of evidence in epistemic logic. In Mattias Skipper Rasmussen & Asbjørn Steglich-Petersen (eds.), *Higher-order evidence*, Oxford University Press.
Woozley, A. D. 1952. Knowing and not knowing. *Proceedings of the Aristotelian Society* 53(n/a). 151–172.
Yaffe, Gideon. 2004. Conditional intent and mens rea. *Legal Theory* 10(4). 273–310. https://doi.org/10.1017/s135232520404025x.
Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026. https://doi.org/10.1093/mind/fzm983.
Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–37. https://doi.org/10.1111/j.1747-9991.2010.00360.x.
Yalcin, Seth. 2012. A counterexample to modus tollens. *Journal of Philosophical Logic* 41(6). 1001–1024. https://doi.org/10.1007/s10992-012-9228-4.
Yap, Audrey. 2014. Idealization, epistemic logic, and epistemology. *Synthese* 191(14). 3351–3366. https://doi.org/10.1007/s11229-014-0448-8.
Yli-Vakkuri, Juhani & John Hawthorne. 2020. Modal epistemology. *Philosophical Studies* .
Yli-Vakkuri, Juhani & John Hawthorne. ms. Modal epistemology. Manuscript.
Zeijlstra, Hedde. 2007. Modal concord. In *Semantics and linguistic theory*, vol. 17, 317–332.

Declaration of Originality of Authorship of a thesis submitted for examination in the BPhil in Philosophy in the academic year 2019-2020

I, Richard Roth, affirm that my submitted thesis entitled Iteration and Preservation is my own original work, except where otherwise stated, and that I have only shown drafts of it to my supervisor, Prof Tim Williamson, in approximately eight hours of supervision, and that I have received help in its preparation from Dr/Prof N/A *(please list all Oxford Philosophy Faculty members you received help from and specify the sort of help received, or write "N/A").*

Signed:

Date: .....30th of June.....