Fitch's Paradox and Level-Bridging Principles*

Weng Kin San

forthcoming in *The Journal of Philosophy*

Abstract

Fitch's Paradox shows that if every truth is knowable, then every truth is known. Standard diagnoses identify the factivity/negative infallibility of the knowledge operator and Moorean contradictions as the root source of the result. This paper generalises Fitch's result to show that such diagnoses are mistaken. In place of factivity/negative infallibility, the weaker assumption of any 'level-bridging principle' suffices. A consequence is that the result holds for some logics in which the "Moorean contradiction" commonly thought to underlie the result is in fact consistent. This generalised result improves on the current understanding of Fitch's result and widens the range of modalities of philosophical interest to which the result might be fruitfully applied. Along the way, we also consider a semantic explanation for Fitch's result which answers a challenge raised by Kvanvig (2006).

1 Introduction

Consider propositions of the form 'φ and it is not known that φ'. Such propositions are unknowable. For otherwise, it would be known that: φ and it is not known that φ. And so, it would be the case that: φ is known and it is known that φ is not known. And since knowledge is factive, it would be the case that: φ is known and φ is not known. Contradiction. So, even if propositions of the form 'φ and it is not known that φ' are true, they cannot be

*For helpful comments and discussions, I am grateful to Jeremy Goodman, John Hawthorne, Matt Hewson, and an anonymous referee. I am indebted, most of all, to Tim Williamson, who provided guidance and feedback on multiple drafts. Research on this paper was supported by the Ertegun Graduate Scholarship and the USC Dana & Dornsife Graduate School Fellowship.

known. Thus, if every truth is knowable, then there cannot be truths of the form 'φ and it is not known that φ'. In other words, if every truth is knowable, then every truth is known.

This is a rough sketch of the result commonly known as 'Fitch's Paradox' or the 'Paradox of Knowability'.[1] The literature surrounding Fitch's result is extensive and continues to burgeon. However, much of it focuses on its ramifications for the knowability principle and for anti-realism. This is regrettable, for Fitch's result does not fundamentally pertain to knowability. Formulated in a bimodal language, Fitch's result simply trades on certain properties of the knowledge and possibility operators. At the abstract level, the result shows that for a certain class of modal logics, assuming that the modal operators interact in a particular way gives rise to modal collapse.

Given its generality, one might try applying the result to a wider range of modalities. For instance, substituting belief for knowledge, Fitch's result would say that if every truth is possibly believed, then every truth is believed. The problem is that the standard proof of the result makes use of the factivity of knowledge. But many interesting modalities, like belief, are not factive. Some have noted that the requirement of factivity can be relaxed to a requirement of 'negative infallibility'. But still, not many modalities of philosophical interest are negatively infallible in the required sense.

This paper generalises Fitch's result to show that a much weaker and more general requirement than factivity or negative infallibility suffices. In particular, a level-bridging principle of any kind will do. Both factivity and negative infallibility are instances of level-bridging principles. Other examples include the principle that evidence of evidence is evidence, the principle that whatever is normal is normally normal, and so on.

This generalisation has far-ranging implications. Most importantly, it sheds new light on Fitch's result and reveals a widely accepted diagnosis of it to be false. Contrary to received wisdom, neither factivity/negative infallibility nor Moorean contradictions are at the root of the result. Furthermore, the generalised result widens the range of modalities to which Fitch's result might be fruitfully applied. There are potentially interesting implications for various debates involving level-bridging principles and Fitch-like principles in epistemology and beyond. Along the way, we also consider an illuminating semantic explanation for Fitch's result which provides an answer to a challenge raised by Kvanvig (2006).

[1] It was noted in Fitch (1963) but was first discovered by Alonzo Church (see Salerno (2009)).

Name	Axiom	Frame Condition
($K_\Box$)	$\Box(p \to q) \to (\Box p \to \Box q)$	Trivial
($D_\Box$)	$\Box p \to \Diamond p$	Serial
($T_\Box$)	$\Box p \to p$	Reflexive
($B_\Box$)	$p \to \Box\Diamond p$	Symmetric
($C4_\Box$)	$\Box\Box p \to \Box p$	Dense
($4_\Box$)	$\Box p \to \Box\Box p$	Transitive
($5_\Box$)	$\Diamond p \to \Box\Diamond p$	Euclidean
($I_\Box$)	$p \longleftrightarrow \Box p$	Identity

Table 1: Common axioms and their frame conditions

2 Fitch's Collapse

First, some preliminaries. Our language will be the propositional bimodal language with the modal operators, $\Box$ and $\blacksquare$.[2] As usual, $\Diamond$ is defined as $\neg\Box\neg$ and $\blacklozenge$ as $\neg\blacksquare\neg$. $\Box$ and $\Diamond$ (and similarly, $\blacksquare$ and $\blacklozenge$) are each other's *duals*. A *(bi)modal logic* L is a set of sentences containing all truth-functional tautologies and that is closed under modus ponens (MP) and uniform substitution (US). φ is a *theorem* of L ($\vdash_{\mathsf{L}} \varphi$) iff $\varphi \in \mathsf{L}$; otherwise, $\nvdash_{\mathsf{L}} \varphi$. L' is an *extension* of L and L is a *sublogic* of L' iff $\mathsf{L} \subseteq \mathsf{L}'$. Table 1 lists some common axioms and their frame conditions. (The corresponding axioms for $\blacksquare$ can be obtained by substituting each occurrence of $\Box$ with $\blacksquare$ and each occurrence of $\Diamond$ with $\blacklozenge$).

In addition to $T_\Box$, Fitch's result requires two other assumptions. First, according to $DIST_\Box$, $\Box$ distributes over conjunctions:

($DIST_\Box$) $\Box(\varphi \wedge \psi) \to (\Box\varphi \wedge \Box\psi)$.

Second, according to the Necessitation Rule for $\blacksquare$, if φ is a theorem, then so is $\blacksquare\varphi$:

($RN_\blacksquare$) If $\vdash \varphi$, then $\vdash \blacksquare\varphi$.

Then, according to Fitch's result:

Fitch's Collapse Theorem. *Let* L *be closed under* $RN_\blacksquare$ *and let it contain* $DIST_\Box$ *and* $T_\Box$. *If* $\vdash_{\mathsf{L}} p \to \blacklozenge\Box p$, *then* $\vdash_{\mathsf{L}} p \longleftrightarrow \Box p$.

Proof: First, suppose the Moorean sentence $\Box(p \wedge \neg\Box p)$ for a contradiction. By $DIST_\Box$, $\Box p \wedge \Box\neg\Box p$. But by $T_\Box$, $\Box\neg\Box p \to \neg\Box p$. So, $\Box p \wedge \neg\Box p$. Contradiction. So, by *reductio*, $\vdash_{\mathsf{L}} \neg\Box(p \wedge \neg\Box p)$. Thus, by $RN_\blacksquare$: $\vdash_{\mathsf{L}} \blacksquare\neg\Box(p \wedge \neg\Box p)$. Now,

[2] Quotation marks are omitted where no risk of confusing use and mention arises.

suppose $\vdash_L p \to \blacklozenge\Box p$. Contraposing yields $\vdash_L \blacksquare\neg\Box p \to \neg p$. So, by (US), $\vdash_L \blacksquare\neg\Box(p\wedge\neg\Box p) \to \neg(p\wedge\neg\Box p)$. Thus, it follows from $\vdash_L \blacksquare\neg\Box(p\wedge\neg\Box p)$ that $\vdash_L \neg(p\wedge\neg\Box p)$—or equivalently, $\vdash_L p \to \Box p$. Furthermore, by $T_\Box$, $\vdash_L \Box p \to p$. Thus, $\vdash_L p \leftrightarrow \Box p$.

On the usual epistemic-alethic interpretation of $\Box$ and $\blacksquare$, the upshot is that the knowability principle—the principle that for every proposition, if it is true, then it is possible to know it—results in the collapse of the logical distinction between knowledge and truth. Given the knowability principle, whatever is true is known, and vice versa.

But we can abstract away from the epistemic-alethic interpretation. More generally, instead of the knowability principle, we have the 'Fitch Principle':

(FITCH PRINCIPLE) $p \to \blacklozenge\Box p$.

Fitch's result shows that given some weak assumptions, the Fitch Principle gives rise to the collapse of $\Box$. It renders $\Box$ redundant in that it makes it so that for every formula φ, $\Box\varphi$ and φ are logically equivalent.

In principle, $\Box$ and $\blacksquare$ admit of a wide range of interpretations. So, one might try running Fitch's argument for belief, justification, evidence, provability, normality, obligation, determinacy, and so on. The main obstacle to extending Fitch's result in an interesting way to these and other modalities is that unlike knowledge, many modalities of philosophical interest do not satisfy $T_\Box$, the assumption of factivity. Or at any rate, for many modalities, it is controversial whether they do. So, perhaps it is no surprise that Fitch's result has not found many interesting applications beyond cases involving knowledge.

Some have noted ways of relaxing the assumptions required for Fitch's result. For instance, Williamson (1993) demonstrates how the result can be derived without $DIST_\Box$. But insofar as we are concerned with broadening the application of Fitch's result to other modalities, being able to dispense with $DIST_\Box$ or $RN_\blacksquare$ is not of much interest. It is the assumption of factivity, rather than those assumptions, that severely restricts the modalities to which the result can be fruitfully applied.

A cursory inspection of the proof reveals an obvious way of weakening the requirement of factivity. In showing that $\Box(p\wedge\neg\Box p)$ is a contradiction, we made use of $T_\Box$ to obtain $\Box\neg\Box p \to \neg\Box p$. However, as Mackie (1980) and many others note, that is of course already furnished by the weaker axiom of negative infallibility:

($NI_\Box$) $\Box\neg\Box p \to \neg\Box p$.

$\mathrm{NI}_{\Box}$ is weaker than $\mathrm{T}_{\Box}$ in the sense that any modal logic containing $\mathrm{T}_{\Box}$ contains $\mathrm{NI}_{\Box}$, but not vice versa. So, with $\mathrm{NI}_{\Box}$ in place of $\mathrm{T}_{\Box}$, we can at least show that the Fitch Principle implies that $\vdash_{\mathsf{L}} p \rightarrow \Box p$.[3]

$\mathrm{NI}_{\Box}$ is a requirement of negative infallibility. According to it, $\Box$ is truth-entailing when it comes to sentences of the form $\neg\Box p$. For instance, under the doxastic interpretation, $\mathrm{NI}_{\Box}$ encodes the assumption that beliefs about one's lack of belief are infallible. If one believes that one does not believe something, then one does not believe it.[4]

But weakening the requirement of factivity to one of negative infallibility does not prove to be much of an advance, at least when it comes to being able to apply Fitch's result more widely. Many modalities of philosophical interest are not widely taken to be negatively infallible in the sense required. For instance, consider the following counterexample to the negative infallibility of belief:

> THE IMPLICIT SEXIST: John is an avowed anti-sexist. In particular, he is prepared to defend vigorously the equality of the sexes in intelligence. Yet, in a variety of contexts, John's behavior and judgments are systematically sexist. Concerning the individual women he knows, John rarely thinks they're as intelligent as the men he knows, even when John has ample evidence of their intelligence. In group discussions, John is systematically less likely to pay attention to and take seriously the contributions of women. On the rare occasions when he does judge a woman to have expressed a novel, interesting idea, he is much more surprised than he would have been if a man had expressed the same idea. Still, John is unaware of these dispositions, and he would deny that he had them if asked (Greco, 2015, 657-658).[5]

It appears that John both believes that women are inferior and believes that he does not believe that women are inferior. This contradicts $\mathrm{NI}_{\Box}$, according to which John's beliefs about what he does not believe cannot be mistaken.

No doubt there are ways to explain away purported counterexamples like the ones above. Indeed, the negative infallibility of belief has its defenders.

[3]It is not so obvious whether $\vdash_{\mathsf{L}} \Box p \rightarrow p$ also holds. As we will later see, the General Collapse Theorem implies that it does.

[4]$\mathrm{NI}_{\Box}$ has been discussed by a few authors. It is axiom (5c) in Rieger's doxastic logic (2015), which he calls 'negative belief infallibility'. Smullyan (1988, 81) calls agents who obey $\mathrm{NI}_{\Box}$ under the doxastic interpretation of $\Box$ 'stable' and those who violate it 'peculiar'. Also see system $\mathsf{K4!}$ in Chellas (1980, 142).

[5]Greco cites this example as *prima facie* evidence that belief does not iterate. His example is repurposed here to serve as a counterexample to the negative infallibility of belief.

And it is validated in the standard KD45 logic for belief. But the mere existence of disagreement surrounding the negative infallibility of belief thwarts our ability to apply Fitch's result to belief with anywhere near the same dialectical force as in the case of knowledge. So, it would be desirable to be able to relax the assumption of factivity further.

However, the prevailing consensus is that this cannot be done. Many identify the factivity/negative infallibility of the $\Box$ operator as the fundamental source of the modal collapse noted in Fitch's result. For instance, Mackie traces the reason for Fitch's result to the factivity of $\Box$:

> [Fitch's result] is initially surprising. But only initially. It comes about because truth-entailing operators can be used to construct self-refuting expressions (1980, 90).

Similarly, Kvanvig argues that the result and its paradoxicality are intimately tied to the factivity of $\Box$:

> it does not appear plausible in the least to suppose that consideration of other non-factive epistemic operators will generate paradoxical results of the sort found in the knowability paradox (2006, 25).

Stjernberg even goes as far as to take the role that factivity plays in Fitch's result as grounds for rejecting the factivity of knowledge:

> Fitch's paradox of knowability also uses the factiveness of knowledge in arriving at the unexpected conclusion that if all truths are knowable, all truths are known... Since the derivation of the troubling result rests on using the principle that knowledge is factive, the Fitch result can also be avoided by giving up factiveness (2009, 31-32).

If the prevailing consensus is anything to go by, the prospect of weakening the requirement of factivity to anything less than negative infallibility seems dire. This would appear to be further vindicated by the fact that Fitch's proof involves showing that $\Box(p \wedge \neg\Box p)$ is a contradiction. But that is a contradiction if and only if $\mathrm{NI}_{\Box}$ is a theorem. So, negative infallibility appears to be essential to Fitch's result.

But appearances are misleading. Not only is it possible to weaken the requirement of negative infallibility, it is possible to weaken it dramatically. §4 shows that instead of factivity or negative infallibility, any level-bridging principle will do. But before that, the next section takes a detour into the semantic side of things. Although this detour is illuminating and provides intuitive motivation for the generalised result proved later, the impatient reader may skip ahead or skim through it without much loss in continuity.

3 A Semantic Perspective

Kvanvig writes:

> I am not claiming here that the paradox is unresolvable, that there is no acceptable way to explain how such a collapse can be tolerated. After all, we know of other contexts in which such a collapse occurs: in one quite popular modal system, there is no logical distinction between actual necessity and possible necessity, for example. It is worth noticing, however, that the loss of this distinction is made palatable by a semantic explanation of the loss, in terms of possible worlds and accessibility relations among them. Perhaps something similar can be done with the knowability [principle], or perhaps there is some other way to relieve our discomfort at the lost distinction resulting from Fitch's proof. The point I want to insist on is the need for some such explanation, a need arising from the fact that the result above is not merely surprising but is, instead, paradoxical (2006, 53-54).

In this section, we sketch a semantic explanation of the kind that Kvanvig demands. Semantic approaches to Fitch's result are relatively uncommon.[6] This is regrettable. The semantic explanations we sketch below provide a useful alternate perspective on why modal collapse results from the Fitch Principle. But more importantly for our purposes, they predict that the threat of modal collapse arises in a much larger class of modal logics than Fitch's original proof suggests. This prediction will inform our conjecture generalising Fitch's result, which we will prove in §4.

We use a standard Kripke possible-worlds semantics. A bimodal Kripke frame $F =< W, R_\Box, R_\blacksquare >$ has a set W as its domain (whose elements we refer to as 'worlds') and two accessibility relations, $R_\Box$ and $R_\blacksquare$, relating worlds. If $wR_\Box v$, we say that w $\Box$-accesses v or that v is an $R_\Box$-successor of w. Similarly, if $wR_\blacksquare v$, we say that w $\blacksquare$-accesses v or that v is an $R_\blacksquare$-successor of w. As is well-known, many common modal axioms have simple corresponding

[6] Some exceptions include Costa-Leite (2004, 2006), Maffezioli *et al.* (2012), Fischer (2013), and Artemov & Protopopescu (2013).

frame conditions. For instance, $T_\Box$ corresponds to the reflexivity of $R_\Box$, $4_\Box$ corresponds to the transitivity of $R_\Box$, and so on (see Table 1).

It is easy to show that the Fitch Principle, $p \to \blacklozenge\Box p$, also has a corresponding frame condition:

Frame Lemma.

$$F \Vdash p \to \blacklozenge\Box p \quad \textit{iff} \quad F \vDash \forall x \exists y (x R_\blacksquare y \land \forall z (y R_\Box z \to x = z)).$$

In words, the frame condition for the Fitch Principle is the condition that every world w (in the domain of the frame) has an $R_\blacksquare$-successor that either: (i) has no $R_\Box$-successor or (ii) has w as its sole $R_\Box$-successor. Pictorially (dashed arrows represent $R_\blacksquare$ and solid ones represent $R_\Box$):

That is, for each world w, there must be a world v such that either: (i) the situation on the left occurs (w $\blacksquare$-accesses v and v does not $\Box$-access any world), or; (ii) the one on the right occurs (w $\blacksquare$-accesses v and v $\Box$-access only w).

For the rest of this section, we restrict our attention to normal modal logics, since many such logics are complete with respect to some class of Kripke frames. A *normal* modal logic is a modal logic that contains ($K_\Box$) and ($K_\blacksquare$) and that is closed under the Necessitation Rules: ($RN_\Box$) ($\varphi/\Box\varphi$) and $RN_\blacksquare$ ($\varphi/\blacksquare\varphi$). $\mathsf{K}_\Box \oplus \mathsf{K}_\blacksquare$ denotes the smallest normal modal logic. $\mathsf{KA}^1 \ldots \mathsf{A}^n{}_\Box \oplus \mathsf{KA}^1 \ldots \mathsf{A}^m{}_\blacksquare$ denotes the smallest normal modal logic containing $A^1_\Box, \ldots, A^n_\Box$ and $A^1_\blacksquare, \ldots, A^m_\blacksquare$.[7] For instance, $\mathsf{KT}_\Box \oplus \mathsf{KD}_\blacksquare$ is the smallest normal modal logic containing $T_\Box$ and $D_\blacksquare$. In certain cases, we use more common notation—for instance, $\mathsf{Triv} = \mathsf{KI}$.

Since all normal modal logics contain $DIST_\Box$ and are closed under $RN_\blacksquare$, a normal modal logic which satisfies the assumptions in Fitch's Collapse Theorem is the smallest one containing $T_\Box$, i.e. $\mathsf{KT}_\Box \oplus \mathsf{K}_\blacksquare$. $\mathsf{KT}_\Box \oplus \mathsf{K}_\blacksquare$ is complete with respect to frames where $R_\Box$ is reflexive. Consider what happens when we assume the Fitch Principle and impose its frame condition on such frames. According to the Frame Lemma, the condition that corresponds to the Fitch Principle is that each world w must have an $R_\blacksquare$-successor that either: (i) has no $R_\Box$-successor or (ii) has w as its sole $R_\Box$-successor. But given

[7] Where $A^1_\Box, \ldots, A^n_\Box$ are axioms containing occurrences of $\Box$ but not $\blacksquare$ and $A^1_\blacksquare, \ldots, A^m_\blacksquare$ are axioms containing occurrences of $\blacksquare$ but not $\Box$.

the assumption that $R_\Box$ is reflexive, (i) cannot obtain. So, it must be that (ii): each world is the sole $R_\Box$-successor of one of its $R_\blacksquare$-successors.

Now, consider an arbitrary world w and let the role of having w as its sole $R_\Box$-successor be played by v. w and v must be identical. For, by reflexivity, v $\Box$-accesses itself. So, if w and v were distinct, that would contradict the assumption that v has w as its sole $R_\Box$-successor. Thus, the role of having w as its sole $R_\Box$-successor must be played by w itself. Since w was arbitrary, this reasoning generalises. Thus, given the reflexivity of $R_\Box$, the frame condition for the Fitch Principle forces each world in the frame to $\Box$-access only itself. Such frames validate $p \longleftrightarrow \Box p$. This would explain why, given $\mathsf{KT}_\Box \oplus \mathsf{K}_\blacksquare$, the Fitch Principle gives rise to the collapse of $\Box$.

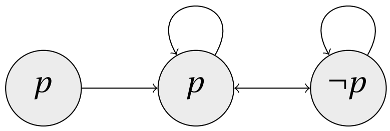
In a similar fashion, we can explain why negative infallibility, in place of factivity, also suffices. For instance, a normal modal logic which validates $\mathrm{NI}_\Box$ but not $\mathrm{T}_\Box$ is $\mathsf{KD4}_\Box \oplus \mathsf{K}_\blacksquare$, which is complete with respect to frames in which $R_\Box$ is serial and transitive. By a similar reasoning, we can show that given the seriality and transitivity of $R_\Box$, the frame condition for the Fitch Principle also forces each world to $\Box$-access only itself.

The semantic reasoning has not thus far revealed anything we did not already know. We already knew that Fitch's result holds for systems that validate $\mathrm{RN}_\blacksquare$, $\mathrm{DIST}_\Box$, and $\mathrm{T}_\Box/\mathrm{NI}_\Box$. But now consider $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$, which is complete with respect to frames where $R_\Box$ is serial and dense. Consider what happens when the frame condition for the Fitch Principle is imposed on such frames. Each world w must have an $R_\blacksquare$-successor that either: (i) has no $R_\Box$-successor or (ii) has w as its sole $R_\Box$-successor. The seriality of $R_\Box$ rules (i) out. So, it must be that (ii): each world is the sole $R_\Box$-successor of one of its $R_\blacksquare$-successors.

Now, consider an arbitrary world w and let the role of having w as its sole $R_\Box$-successor be played by v. In turn, let the role of having v as its sole $R_\Box$-successor be played by u. w and v must be identical. For, by density, it follows from $uR_\Box v$ that there is a world t such that $uR_\Box t$ and $tR_\Box v$. However, by assumption, u $\Box$-accesses only v, and so t and v must be identical, and so $vR_\Box v$. And since v $\Box$-accesses itself, v and w cannot be distinct. For, if they were, that would contradict the assumption that v $\Box$-accesses only w. Thus, the role of having w as its sole $R_\Box$-successor must be played by w itself. And since w was arbitrary, this generalises to every world. Thus, given the seriality and density of $R_\Box$, the frame condition for the Fitch Principle again forces each world to $\Box$-access only itself. If this reasoning is sound, then the Fitch Principle should also give rise to the collapse of $\Box$, given $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$.

And if Fitch's result holds for $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$, then it also holds for $\mathsf{KD5}_\Box \oplus$

$\mathsf{K}_\blacksquare$, which extends $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$. But this would mean that neither factivity nor negative infallibility is required for Fitch's result, since neither $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$ nor $\mathsf{KD5}_\Box \oplus \mathsf{K}_\blacksquare$ contains $\mathrm{T}_\Box$ or $\mathrm{NI}_\Box$. For, consider:



$\Box(p \wedge \neg\Box p)$

Since the accessibility relation for $\Box$ is serial and dense, this is a model of $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$. It is not difficult to check that it is also a model of $\mathsf{KD5}_\Box \oplus \mathsf{K}_\blacksquare$. Thus, $\Box(p \wedge \neg\Box p)$ is consistent in both $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$ and $\mathsf{KD5}_\Box \oplus \mathsf{K}_\blacksquare$, since it is satisfied at the leftmost point of the model. And we have shown that if $\Box(p \wedge \neg\Box p)$ is consistent, then $\mathrm{NI}_\Box$ is not a theorem and thus neither is $\mathrm{T}_\Box$. So, it would appear that Fitch's result holds for a larger class of logics than is commonly recognised. In particular, it appears to hold even for logics where $\Box$ is neither factive nor negatively infallible.

Before we consider a significant caveat to this, we should note another interesting prediction that the semantic reasoning makes. Consider $\mathsf{KDB}_\Box \oplus \mathsf{K}_\blacksquare$, which is complete with respect to frames where $R_\Box$ is serial and symmetric. Again, if we impose the frame condition for the Fitch Principle on such frames, then each world w must have an $R_\blacksquare$-successor that either: (i) has no $R_\Box$-successor or (ii) has w as its sole $R_\Box$-successor. Again, the seriality of $R_\Box$ rules (i) out. So, by (ii), each world is the sole $R_\Box$-successor of one of its $R_\blacksquare$-successors.

Again, consider an arbitrary world w and let the role of having w as its sole $R_\Box$-successor be played by v. In turn, let the role of having v as its sole $R_\Box$-successor be played by u. u and w must be identical. For, by symmetry, it follows from $uR_\Box v$ that $vR_\Box u$. And so, if u and w were distinct, this would contradict the assumption that v $\Box$-accesses only w. Thus, u and w are identical (and u and v are either distinct or identical). Thus, given the seriality and symmetry of $R_\Box$, the frame condition for the Fitch Principle forces the frame to contain only non-reflexive, symmetric $\Box$-pairs of worlds (if $u \neq v$) and/or worlds that $\Box$-access only themselves (if $u = v$):

Such frames validate $p \longleftrightarrow \Box\Box p$. This suggests an extension of Fitch's result. With the logics considered previously, the Fitch Principle results in the collapse of $\Box$ (i.e. φ and $\Box\varphi$ are logically equivalent). It would seem that with

$\mathsf{KDB}_\Box \oplus \mathsf{K}_\blacksquare$, the Fitch Principle results in the collapse of $\Box\Box$ (i.e. φ and $\Box\Box\varphi$ are logically equivalent).

At this juncture, it is important to note a caveat. The kind of semantic reasoning engaged in throughout this section does not have the force of proper proofs. And that is because to reason as we did, we require certain completeness results which are not readily available. For instance, one result that is missing would show that the smallest normal extension of $\mathsf{KT}_\Box \oplus \mathsf{K}_\blacksquare$ containing the Fitch Principle is complete with respect to the class of frames where $R_\Box$ is reflexive and the frame condition for the Fitch Principle is satisfied. Another would show that the smallest normal extension of $\mathsf{KD4}_\Box \oplus \mathsf{K}_\blacksquare$ containing the Fitch Principle is complete with respect to the class of frames where $R_\Box$ is serial and transitive, and the frame condition for the Fitch Principle is satisfied. And so on. Without such results, we cannot properly infer what is provable in certain logics containing the Fitch Principle from what holds in certain frames satisfying its corresponding frame condition.

Nevertheless, as we will soon see, the predictions of the semantic reasoning are in fact vindicated. So, that provides some evidence that the required completeness results can be had, though proving them lies beyond the scope of this paper. Nevertheless, that does not render this section all for naught. Taking the semantic explanations as merely suggestive predictions about when the Fitch Principle might generate modal collapse turns out to be highly instructive. Those predictions inform a conjecture generalising Fitch's result, which we can prove via non-semantic means. That is the aim of the next section.

4 General Collapse

We know from Fitch's result that the Fitch Principle gives rise to the collapse of $\Box$ (i.e. $\vdash p \longleftrightarrow \Box p$) in $\mathsf{KT}_\Box \oplus \mathsf{K}_\blacksquare$ and $\mathsf{KD4}_\Box \oplus \mathsf{K}_\blacksquare$. The semantic explanation in the previous section predicts that this collapse also occurs in $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$ and $\mathsf{KD5}_\Box \oplus \mathsf{K}_\blacksquare$. Furthermore, it also predicts that the Fitch Principle gives rise to the collapse of $\Box\Box$ (i.e. $\vdash p \longleftrightarrow \Box\Box p$) in $\mathsf{KDB}_\Box \oplus \mathsf{K}_\blacksquare$.

An interesting pattern emerges. Let M_1 and M_2 denote arbitrary strings of $\Box$ and $\Diamond$. And let $|M_1 - M_2|$ be the difference in the number of operators in M_1 and M_2. For instance, if $M_1 = \Box$ and $M_2 = \Box\Diamond\Diamond\Box$, then $|M_1 - M_2| = 3$. And if $M_1 = \Box\Diamond\Diamond\Box\Diamond\Diamond\Box$ and $M_2 = \Box\Diamond$, then $|M_1 - M_2| = 5$. Notice that the logics for which the Fitch Principle is expected to give rise to the collapse of $\Box$ all extend some logic of the form $\mathsf{KDX}_\Box \oplus \mathsf{K}_\blacksquare$, where $\mathrm{X}_\Box$ is $M_1 p \to M_2 p$ with $|M_1 - M_2| = 1$. For instance, $\mathrm{T}_\Box$ ($\Box p \to p$) has one operator to the left of the conditional and none to the right, and $\mathsf{KT}_\Box \oplus \mathsf{K}_\blacksquare$ is equivalent to $\mathsf{KDT}_\Box \oplus \mathsf{K}_\blacksquare$, since $\mathrm{T}_\Box$ implies $\mathrm{D}_\Box$. Similarly, $4_\Box$ ($\Box p \rightarrow \Box\Box p$), $5_\Box$ ($\Diamond p \rightarrow \Box\Diamond p$), and $\mathrm{C}4_\Box$ ($\Box\Box p \rightarrow \Box p$) each has two operators on one side of the conditional and one on the other. Notice furthermore that $\mathrm{B}_\Box$ ($p \rightarrow \Box\Diamond p$), on the other hand, has two operators on the right of the conditional and none on the other. Thus, the logic for which the Fitch Principle is expected to give rise to the collapse of $\Box\Box$ is of the form $\mathsf{KDX}_\Box \oplus \mathsf{K}_\blacksquare$, where $\mathrm{X}_\Box$ is $M_1 p \rightarrow M_2 p$ with $|M_1 - M_2| = 2$.

A natural conjecture is thus that the 'degree' of modal collapse caused by the Fitch Principle in logics extending $\mathsf{KDX}_\Box \oplus \mathsf{K}_\blacksquare$, where $\mathrm{X}_\Box$ is of the form $M_1 p \rightarrow M_2 p$, is a function of the difference in the number of operators in M_1 and M_2. More specifically, when $|M_1 - M_2| = n$, the Fitch Principle gives rise to the collapse of $\Box^n$ (where $\Box^n$ abbreviates $n \geq 0$ iterations of $\Box$ and $\Box^0$ is the empty string). Or to put things differently, let us call a principle of the form $M_1 p \rightarrow M_2 p$ with $|M_1 - M_2| = n$ an 'n-level bridging principle'. So, for instance, $\mathrm{T}_\Box$ ($\Box p \rightarrow p$) and $4_\Box$ ($\Box p \rightarrow \Box\Box p$) are both one-level bridging principles, whereas $\mathrm{B}_\Box$ ($p \rightarrow \Box\Diamond p$) is a two-level bridging principle. And call $p \longleftrightarrow \Box^n p$ 'nth-degree modal collapse'. Then, the conjecture is that given an n-level bridging principle, the Fitch Principle gives rise to nth-degree modal collapse.

The aim of this section is to prove this conjecture. More precisely, we prove the following generalisation of Fitch's result:

General Collapse Theorem. *Let* L *be closed under* $\mathrm{RN}_\blacksquare$ *and* $\mathrm{RM}_\Box$*, and let it contain* $\mathrm{D}_\Box$ *and some n-level bridging principle. If* $\vdash_{\mathsf{L}} p \rightarrow \blacklozenge\Box p$*, then* $\vdash_{\mathsf{L}} p \longleftrightarrow \Box^n p$.[8]

$\mathrm{RM}_\Box$ is the Monotonicity Rule for $\Box$. According to it, if $\varphi \rightarrow \psi$ is a theorem, then so is $\Box\varphi \rightarrow \Box\psi$:

($\mathrm{RM}_\Box$) If $\vdash \varphi \rightarrow \psi$, then $\vdash \Box\varphi \rightarrow \Box\psi$.

We call modal logics closed under $\mathrm{RM}_\Box$ '$\Box$-monotonic'. A fact about $\Box$-monotonic logics is that they validate Duality:

(Duality) If $\vdash M_1 p \rightarrow M_2 p$, then $\vdash \tilde{M}_2 p \rightarrow \tilde{M}_1 p$,

[8]The General Collapse Theorem is a result in bimodal logic. But since nothing in the proof depends on $\blacksquare$ and $\Box$ being distinct operators, the following is an immediate consequence for monomodal logics:

Corollary 1. *Let* L *be closed under* $\mathrm{RN}_\Box$ *and* $\mathrm{RM}_\Box$*, and let it contain* $\mathrm{D}_\Box$ *and some n-level bridging principle. If* $\vdash_{\mathsf{L}} p \rightarrow \Diamond\Box p$*, then* $\vdash_{\mathsf{L}} p \longleftrightarrow \Box^n p$.

where, in general, $\tilde{M}$ is the *dual string* of M (e.g. if $M = \Box\Diamond\Diamond\Box$, then $\tilde{M} = \Diamond\Box\Box\Diamond$). It is well known that normal modal logics validate Duality. For instance, it is a familiar fact that given normality, $\mathrm{T}_\Box$ ($\Box p \to p$) can be equivalently formulated as $p \to \Diamond p$ and $4_\Box$ ($\Box p \to \Box\Box p$) as $\Diamond\Diamond p \to \Diamond p$, and so on. $\Box$-monotonicity suffices for this to hold. Furthermore, as in normal modal logics, $\Box$ also distributes over conjunctions in $\Box$-monotonic logics. That is, all $\Box$-monotonic logics contain $\mathrm{DIST}_\Box$.

Given these facts, we can prove the General Collapse Theorem in three simple steps. The first step is to show that given $\mathrm{RM}_\Box$ and $\mathrm{D}_\Box$, any n-level bridging principle implies $\Box^m p \to \Diamond^{m+n} p$, for some m. The second step is to show that given $\mathrm{RM}_\Box$, $\mathrm{RN}_\blacksquare$, and the Fitch Principle, $\Box^m p \to \Diamond^{m+n} p$ implies $p \to \Diamond^n p$, for any m, n. The third and final step is to show that given $\mathrm{RM}_\Box$, $\mathrm{RN}_\blacksquare$, and the Fitch Principle, $p \to \Diamond^n p$ implies $p \longleftrightarrow \Box^n p$, for any n.

Step 1. First, we show:

Lemma 1. *Let* L *be closed under* $\mathrm{RM}_\Box$ *and let it contain* $\mathrm{D}_\Box$*. If* L *contains an* n*-level bridging principle, then* L *contains* $\Box^m p \to \Diamond^{m+n} p$*, for some* m*.*

This is fairly intuitive. To illustrate, consider the following two-level bridging principle: $\Box\Diamond p \to \Box\Box\Diamond\Box p$. Now, according to $\mathrm{D}_\Box$, $\Box p \to \Diamond p$. So, intuitively, $\Box\Box p \to \Box\Diamond p$ and $\Box\Box\Diamond\Box p \to \Diamond\Diamond\Diamond\Diamond p$. And so, it follows from $\Box\Diamond p \to \Box\Box\Diamond\Box p$ that $\Box\Box p \to \Diamond\Diamond\Diamond\Diamond p$, i.e. $\Box^2 p \to \Diamond^4 p$. So, Lemma 1 holds in this instance. Alternatively, consider a two-level bridging principle with more operators on the left-hand side: e.g. $\Box\Box\Diamond\Box p \to \Box\Diamond p$. By Duality, it is equivalent to one with more operators on the right hand side, i.e. $\Diamond\Box p \to \Diamond\Diamond\Box\Diamond p$. Thus, in the same way as before, we can use $\mathrm{D}_\Box$ to derive $\Box^2 p \to \Diamond^4 p$.

This reasoning generalises. Any n-level bridging principle either is or is equivalent by Duality to a principle of the form $O_1 \dots O_m p \to O'_1 \dots O'_{m+n} p$, for some m, where each $O_1 \dots O_m, O'_1 \dots O'_{m+n}$ is either $\Box$ or $\Diamond$. Intuitively, it follows from $\mathrm{D}_\Box$ that $\Box^m p \to O_1 \dots O_m p$ and $O'_1 \dots O'_{m+n} p \to \Diamond^{m+n} p$. And so, any n-level bridging principle implies, for some m, $\Box^m p \to \Diamond^{m+n} p$. The idea behind the proof is simple. The only technicality is checking that $\mathrm{RM}_\Box$ is strong enough for regimenting the informal line of reasoning above—that is a tedious but easy exercise.

Step 2. Second, we show:

Lemma 2. *Let* L *be closed under* $\mathrm{RN}_\blacksquare$ *and* $\mathrm{RM}_\Box$*, and let it contain the Fitch Principle. For any* m, n*, if* $\vdash_\mathsf{L} \Box^m p \to \Diamond^{m+n} p$*, then* $\vdash_\mathsf{L} p \to \Diamond^n p$*.*

The proof of this relies on the fact that given the Fitch Principle, we can gradually subtract from $\vdash_{\mathsf{L}} \Box^m p \to \Diamond^{m+n} p$ a $\Box$ on the left of the conditional and a $\Diamond$ on the right. That is, we can show that for any $i \geq 0$:

$$\text{if } \vdash_{\mathsf{L}} \Box^{i+1} p \to \Diamond^{(i+1)+n} p, \text{ then } \vdash_{\mathsf{L}} \Box^{i} p \to \Diamond^{i+n} p. \tag{†}$$

In turn, the proof of this uses exactly the same kind of reasoning that underlies the usual proof of Fitch's result.

First, we identify a 'Moorean' contradiction. Suppose for a contradiction that $\Box(\Box^i p \wedge \neg\Diamond^{i+n} p)$. Distributing $\Box$ over the conjunction yields $\Box^{i+1} p \wedge \Box\neg\Diamond^{i+n} p$. However, contraposing the antecedent of (†) yields $\vdash_{\mathsf{L}} \Box\neg\Diamond^{i+n} p \to \neg\Box^{i+1} p$. Thus, given the antecedent of (†), we derive a contradiction: $\Box^{i+1} p \wedge \neg\Box^{i+1} p$. So, by *reductio*, $\vdash_{\mathsf{L}} \neg\Box(\Box^i p \wedge \neg\Diamond^{i+n} p)$. And so by $\mathrm{RN}_{\blacksquare}$, $\vdash_{\mathsf{L}} \blacksquare\neg\Box(\Box^i p \wedge \neg\Diamond^{i+n} p)$. Now, since L contains the Fitch Principle, i.e. $\vdash_{\mathsf{L}} p \to \blacklozenge\Box p$, by contraposition, $\vdash_{\mathsf{L}} \blacksquare\neg\Box p \to \neg p$. And so, by (US), $\vdash_{\mathsf{L}} \blacksquare\neg\Box(\Box^i p \wedge \neg\Diamond^{i+n} p) \to \neg(\Box^i p \wedge \neg\Diamond^{i+n} p)$. Thus, it follows from what we showed above that $\vdash_{\mathsf{L}} \neg(\Box^i p \wedge \neg\Diamond^{i+n} p)$—or equivalently, $\vdash_{\mathsf{L}} \Box^i p \to \Diamond^{i+n} p$. Thus, (†). And from that, Lemma 2 easily follows since by (†), $\vdash_{\mathsf{L}} \Box^m p \to \Diamond^{m+n} p$ implies $\vdash_{\mathsf{L}} \Box^{m-1} p \to \Diamond^{m+n-1} p$, which in turn implies $\vdash_{\mathsf{L}} \Box^{m-2} p \to \Diamond^{m+n-2} p$, and so on until $\vdash_{\mathsf{L}} \Box^{m-m} p \to \Diamond^{m+n-m} p$.

Step 3. Finally, we show:

Lemma 3. *Let L be closed under* $\mathrm{RN}_{\blacksquare}$ *and* $\mathrm{RM}_{\Box}$*, and let it contain the Fitch Principle. For any n, if $\vdash_{\mathsf{L}} p \to \Diamond^n p$, then $\vdash_{\mathsf{L}} p \leftrightarrow \Box^n p$.*[9]

It is an immediate consequence of Duality that if $\vdash_{\mathsf{L}} p \to \Diamond^n p$, then $\vdash_{\mathsf{L}} \Box^n p \to p$. So, all that is left to show is that if $\vdash_{\mathsf{L}} p \to \Diamond^n p$, then $\vdash_{\mathsf{L}} p \to \Box^n p$. We prove this by showing that given the Fitch Principle, we can turn the $\Diamond$'s into $\Box$'s, one by one, from the inside out. That is, we prove by induction on the number i of operators that for any $i \leq n$:

$$\text{if } \vdash_{\mathsf{L}} p \to \Diamond^n p, \text{ then } \vdash_{\mathsf{L}} p \to \Diamond^{n-i}\Box^i p. \tag{*}$$

The base case where $i = 0$ is trivial. For the inductive step, let the inductive hypothesis be: if $\vdash_{\mathsf{L}} p \to \Diamond^n p$, then $\vdash_{\mathsf{L}} p \to \Diamond^{n-i}\Box^i p$. We will show that: if $\vdash_{\mathsf{L}} p \to \Diamond^n p$, then $\vdash_{\mathsf{L}} p \to \Diamond^{n-(i+1)}\Box^{i+1} p$.

Again, the proof of this is similar to the usual proof of Fitch's result. First, we identify a 'Moorean' contradiction. Suppose $\vdash_{\mathsf{L}} p \to \Diamond^n p$. And assume

[9] Note that the case where $n = 1$ is just Fitch's original result, since by Duality, $\mathrm{T}_{\Box}$ is equivalent to $p \to \Diamond^1 p$.

for a contradiction that $\Box(p \wedge \neg\Diamond^{n-(i+1)}\Box^{i+1}p)$. By DIST$_\Box$, $\Box p \wedge \Box\neg\Diamond^{n-(i+1)}$ $\Box^{i+1}p$. However, by the inductive hypothesis: $\vdash_\mathsf{L} p \rightarrow \Diamond^{n-i}\Box^{i}p$. It follows by (US) that $\vdash_\mathsf{L} \Box p \rightarrow \Diamond^{n-i}\Box^{i+1}p$. And so by contraposition, $\vdash_\mathsf{L} \Box\neg\Diamond^{n-(i+1)}\Box^{i+1}p$ $\rightarrow \neg\Box p$. And thus from $\Box p \wedge \Box\neg\Diamond^{n-(i+1)}\Box^{i+1}p$, we can derive a contradiction: $\Box p \wedge \neg\Box p$. So, by *reductio*, $\vdash_\mathsf{L} \neg\Box(p \wedge \neg\Diamond^{n-(i+1)}\Box^{i+1}p)$. And so by RN$_\blacksquare$, $\vdash_\mathsf{L}$ $\blacksquare\neg\Box(p \wedge \neg\Diamond^{n-(i+1)}\Box^{i+1}p)$. Now, assume the Fitch Principle, $\vdash_\mathsf{L} p \rightarrow \blacklozenge\Box p$. By contraposition, $\vdash_\mathsf{L} \blacksquare\neg\Box p \rightarrow \neg p$. And so, by (US), $\vdash_\mathsf{L} \blacksquare\neg\Box(p \wedge \neg\Diamond^{n-(i+1)}\Box^{i+1}p)$ $\rightarrow \neg(p \wedge \neg\Diamond^{n-(i+1)}\Box^{i+1}p)$. Thus, it follows from what we showed above that $\vdash_\mathsf{L} \neg(p \wedge \neg\Diamond^{n-(i+1)}\Box^{i+1}p)$—or equivalently, $\vdash_\mathsf{L} p \rightarrow \Diamond^{n-(i+1)}\Box^{i+1}p$. This completes the inductive step. And thus, (*), from which Lemma 3 easily follows (let $i = n$).

The General Collapse Theorem follows from Lemmas 1-3 by a simple chain of implications. So, we have generalised Fitch's result so that factivity/negative infallibility is not required. In its place is the much more general requirement of a level-bridging principle of any kind. More specifically, we showed that given some weak background assumptions, the presence of an n-level bridging principle means that the Fitch Principle gives rise to nth-degree modal collapse. That is, given $\vdash M_1 p \rightarrow M_2 p$, where $|M_1 - M_2| = n$, if $\vdash p \rightarrow \blacklozenge\Box p$ then $\vdash p \leftrightarrow \Box^n p$. For instance, given a three-level bridging principle like $p \rightarrow \Diamond\Box\Diamond p$, the Fitch Principle gives rise to third-degree modal collapse (i.e. the collapse of $\Box\Box\Box$).

But for all we have shown, it is possible that a stronger result holds. Perhaps the three-level bridging principle above also suffices for the Fitch Principle to give rise to second-degree modal collapse. Or, perhaps given the Fitch Principle, any level-bridging principle whatsoever suffices for the full collapse of $\Box$. That is not ruled out by the General Collapse Theorem. But that is not the case. The modal collapse identified in the General Collapse Theorem is, in a sense, the full extent of the collapse. The degree of the modal collapse that occurs really is essentially a function of the difference in the number of operators in the level-bridging principle. For a precise statement and proof of this, the reader is referred to the Appendix.

5 Diagnosing Fitch's Result

The General Collapse Theorem informs our understanding of Fitch's result in important ways. In particular, it invalidates standard diagnoses of the result. First, factivity or negative infallibility is often thought to be the ultimate source of the modal collapse identified in Fitch's result. The quotes from Mackie, Kvanvig, and Stjernberg from earlier are illustrative. However, the General Collapse Theorem shows that what underlies the result is, more generally, a level-bridging principle of any kind. To be sure, the principles of factivity and negative infallibility are instances of level-bridging principles. But to identify factivity or negative infallibility as the root of Fitch's result would be to falsely diagnose as a cause what is in fact merely a symptom of a more general pathology.

Another common diagnosis takes 'Moorean' contradictions of the form $\Box(p \wedge \neg\Box p)$ to be at the base of Fitch's result. A key step of Fitch's original proof involves showing that $\Box(p \wedge \neg\Box p)$ is a contradiction. This has led many to believe that Fitch's result is inextricably linked to Moorean sentences. For instance: "the Moore sentence is at the root of Fitch's famous "paradox of knowability"" (Holliday & Icard III, 2010); "the Fitch paradox... can be seen as derivative from Moore's paradox" (Bonnay & Égré, 2011, 105); "the epistemic Moore sentence $p \wedge \neg K_a p$ lies... at the bottom of the Fitch paradox" (Égré, 2014, 538); "Fitch's paradox is intimately tied to [the] so-called "Moore's paradox"" (Cresto, 2017).

While contradictions that might very loosely be called 'Moorean' do play a role (the proofs of Lemmas 2 and 3 make use of such contradictions), the connection between Moorean phenomena and the Fitchean phenomena of modal collapse is more complicated than commonly thought. The General Collapse Theorem implies that the Fitch Principle can give rise to the collapse of $\Box$ even for logics in which $\Box(p \wedge \neg\Box p)$ is consistent. In particular, the General Collapse Theorem applies to $\mathsf{KDC4}_\Box \oplus \mathsf{K}_\blacksquare$ and $\mathsf{KD5}_\Box \oplus \mathsf{K}_\blacksquare$. As we showed in §3, $\Box(p \wedge \neg\Box p)$ is consistent in those logics.[10]

These improvements on our understanding of the source of Fitch's result and its extent have important downstream consequences. In particular, they help in evaluating various responses to the result. For instance, consider those who cite Fitch's result as a reason to reject the principle that knowledge is factive:

> In discussions of Fitch's paradox, it is usually assumed without further argument that knowledge is factive, that if a subject knows that p, then p is true. It is argued that this common assumption is not as well-founded as it should be, and that there in fact

[10]Instead of those of the form $\Box(p \wedge \neg\Box p)$, one might characterise Moorean sentences (for a logic L) more broadly as sentences of the form $\Box\varphi$ such that φ is consistent but $\Box\varphi$ is not (i.e. $\nvdash_{\mathsf{L}} \neg\varphi$ and $\vdash_{\mathsf{L}} \neg\Box\varphi$). Sentences φ with this feature are sometimes called '(logical) blindspots' (Sorensen, 1988). But doing this will not help restore the connection between Fitch's result and Moorean phenomena—for, it is obvious that Fitch's result holds for $\mathsf{Triv}_\Box \oplus \mathsf{K}_\blacksquare$, for which there are no 'blindspots'.

> are certain reasons to be suspicious of the unrestricted version of the factiveness claim... Since the derivation of the troubling result rests on using the principle that knowledge is factive, the Fitch result can also be avoided by giving up factiveness (Stjernberg, 2009, 29-32).

The General Collapse Theorem deprives this kind of response of any force. *Contra* Stjernberg, Fitch's result cannot be avoided by simply giving up the factivity of knowledge. In fact, every single level-bridging principle for knowledge must be given up.[11]

Another upshot concerns analogues of the knowability principle with non-factive operators in place of knowledge. For instance, Mackie (1980) and Chase & Rush (2018) discuss analogues of the knowability principle involving justification and verification, Edgington (1985, 2010) and Williamson (forthcoming) discuss analogues involving probability, predictability, and non-factive evidence, and Chalmers (2012) considers an analogue involving scrutability. The General Collapse Theorem shows that whether the operator in question is factive/negatively infallible does not go to the heart of the matter. The threat of modal collapse looms so long as the operator obeys some level-bridging principle. For instance, if having been verified to be verified implies having been verified, then the verifiability principle is just as vulnerable to Fitch's result as the knowability principle.

The General Collapse Theorem also brings into question other uses to which Fitch's result has been put. For instance, Chase & Rush (2018) argue that the result poses special problems for internalists about justification. They recognise that, in place of $T_\Box$, $D_\Box$ and $4_\Box$ suffice for Fitch's result. Interpreting $\Box$ in terms of justification, they surmise that since internalists are inclined to accept $4_\Box$, they are forced to reject the justificatory analogue of the knowability principle, on pain of losing the logical distinction between truth and justification. But given the General Collapse Theorem, even if this is a problem, it is not a problem specific to internalists. For, it is not only internalists who accept $4_\Box$ that have to reject the justifiability principle. Anyone who accepts any level-bridging principle for justification must too. And it is not clear that the tenability of level-bridging principles like 'if there is justification for there being justification for p, then there is not justification for not-p' is a partisan issue divided along internalist-externalist lines. So,

[11]That is, every *non-trivial* level-bridging principle, i.e. n-level bridging principles where $n \neq 0$. For convenience, I will often write 'level-bridging principles' when I mean 'non-trivial level-bridging principles'. Context should make it clear when the trivial case is meant to be excluded.

Fitch's result cannot be mobilised in the way that Chase and Rush do to adjudicate on the internalist-externalist debate.

6 The Fitch Principle vs Level-Bridging Principles

The General Collapse Theorem has other important implications. It exposes a tension between the Fitch Principle, on the one hand, and level-bridging principles for $\Box$, on the other. Together, they give rise to modal collapse. And since modal collapse of any degree is undesirable for most modalities of philosophical interest, either the Fitch Principle must be rejected or all level-bridging principles must be rejected. This has potentially interesting implications for issues spanning many areas for philosophy.

6.1 The Fitch Principle

First, consider the Fitch Principle. The doxastic-deontic interpretation of $\Box$ and $\blacksquare$ is an interesting case. On such an interpretation, the Fitch Principle ($p \to \blacklozenge\Box p$) formalises a weak, narrow-scope truth norm of belief:

(TRUTH) If p is true, then it is permissible to believe p.

Whiting (2013) and others defend such a norm. TRUTH also follows from various other norms, such as: the strong narrow-scope truth norm according to which, if p is true, then it is obligatory to believe p; or (assuming that the permissibility of being certain that p implies the permissibility of believing p) the norm according to which if p is true, then it is permissible to be certain that p; or (assuming that the permissibility of asserting p implies the permissibility of believing p), the norm according to which if p is true, then it is permissible to assert that p; and so on.

The background assumptions of the General Collapse Theorem (i.e. $RM_\Box$ and $RN_\blacksquare$) are plausible on the doxastic-deontic interpretation. In fact, both doxastic and deontic logic are standardly taken to satisfy KD45. The upshot is that given any level-bridging principle for belief, TRUTH gives rise to modal collapse. In particular, given any one-level bridging principle like 'believing implies believing that one believes', TRUTH results in the loss of any logical distinction between truth and belief—everything true is believed and everything believed is true. Similarly, given any two-level bridging principle like 'believing that one believes that one believes p implies that p is consistent with one's beliefs', TRUTH results in the loss of any logical distinction between truth and second-order belief—everything true is believed to be believed and vice versa.

Modal collapse of any degree is undesirable, since truth is not equivalent to belief of any order. Thus, proponents of TRUTH must reject every level-bridging principle for belief. Consequently, so must proponents of norms that entail TRUTH, like the ones previously listed. To be sure, this is not a decisive refutation of TRUTH or those stronger norms. But the fact that their proponents are so severely restricted in which principles governing belief they can endorse might diminish the appeal of those norms. Notably, competing norms for belief, such as wide-scope truth norms or knowledge norms, do not obviously face similar problems, since unlike TRUTH, their logical forms do not map onto that of the Fitch Principle.

There might be lingering suspicions that this problem is simply derivative of the problem that Moorean sentences are thought to pose for truth norms of belief (see Bykvist & Hattiangadi (2007, 2013)). Such concerns about the novelty of the problem raised here are easily dispelled. We have already seen that, contrary to conventional wisdom, there is no straightforward connection between Moorean phenomena and the phenomenon of modal collapse identified in the General Collapse Theorem.

Besides the doxastic-deontic interpretation, there are various other interesting interpretations of the Fitch Principle to consider. To note just one, interpret $\Box$ as 'God wills it to be the case that' and $\blacksquare$ as 'it is logically necessary that'. Then, the Fitch Principle is the principle according to which every truth is such that it is logically possible that God wills it to be the case. This seems undeniable for an omnipotent being. However, the upshot of the General Collapse Theorem is that if $\Box$ obeys any level-bridging principle at all (for instance, the one-level bridging principle according to which God's will is 'non-akratic', in the sense that if God wills that God wills that p, then God wills that p) then modal collapse occurs—whatever is true is God's will and vice versa.[12] This is an undesirable result for theists.

6.2 Level-Bridging Principles

Besides principles that have the form of the Fitch Principle, the General Collapse Theorem also sheds light on various level-bridging principles, many of which are the subject of vigorous debate. For instance, externalists and internalists are divided on level-bridging principles for justification, such as: if there is justification for believing something, then there is justification for believing that there is justification for believing it ($Jp \to JJp$); its con-

[12]A background assumption is $D_\Box$, the assumption that God's will is consistent (i.e. if God wills that p, then God does not will that not-p).

verse ($JJp \to Jp$); if there is not justification for believing something, then there is justification for believing that there is not justification for believing it ($\neg Jp \to J\neg Jp$); and its converse ($J\neg Jp \to \neg Jp$) (see Smithies (2012) for a defence of all four principles). The upshot of the General Collapse Theorem is that accepting any such principle requires rejecting $p \to \blacklozenge Jp$, for any interpretation of $\blacksquare$ on which $\mathrm{RN}_{\blacksquare}$ is plausible. For instance, it requires denying that every truth is such that there is possibly justification for believing it, every truth is such that there is permissibly justification for believing it, and so on.

Various level-bridging principles for evidence are also matters of dispute. For instance, proponents of the '$E{=}K$'-thesis like Williamson (2000) believe that evidence is factive ($Ep \to p$) while others disagree. There is also disagreement about whether evidence of evidence is evidence ($EEp \to Ep$) (see Feldman (2014) and Fitelson (2012)). If either of those principles hold, then on pain of modal collapse, $p \to \blacklozenge Ep$ must fail for any interpretation of $\blacksquare$ on which $\mathrm{RN}_{\blacksquare}$ is plausible. For instance, it cannot be that every truth is such that for all one knows, there is evidence for it or that every truth is such that it is not normally the case that there is not evidence for it.

Similarly, there is debate surrounding level-bridging principles for other interpretations of $\Box$. For instance, Smith's (2006) logic validates the principle that whatever is normally the case is normally, normally the case ($Np \to NNp$), which Carter (forthcoming) challenges. The issue of whether 'ought' iterates ($Op \to OOp$) is also a live one (see Immerman (forthcoming)).

This is far from an exhaustive list of interesting level-bridging principles to consider. But in each case, the General Collapse Theorem shows that accepting a level-bridging principle for $\Box$ requires rejecting the Fitch Principle ($p \to \blacklozenge\Box p$) for any interpretation of $\blacksquare$ on which $\mathrm{RN}_{\blacksquare}$ is plausible, insofar as modal collapse is to be avoided. The hope is that in each case, weighing the relevant level-bridging principles against various principles of the form $p \to \blacklozenge\Box p$ can help shed light on the tenability of those level-bridging principles and help move the debates forward.

But one might be skeptical about whether this kind of application of the General Collapse Theorem can actually inform the debates on level-bridging principles in any interesting way. After all, on the original epistemic-alethic interpretation, when it came down to choosing between the factivity of knowledge and the knowability principle, it is clearly the latter that should be rejected. So, why think that the tension between the level-bridging principles and the Fitch Principle should ever put any pressure on the level-bridging principles? Why shouldn't the tension always be resolved in favour of the level-bridging principles, as in the epistemic-alethic case?

The difference is that it is not always as clear as in the epistemic-alethic case whether to place the blame for the threat of modal collapse on the Fitch Principle or on the level-bridging principles. For one, the various level-bridging principles for justification, evidence, normality, obligation, and so on are nowhere near as uncontentious as the factivity of knowledge. For another, in the case of knowledge, Fitch's proof already furnishes straightforward counterexamples to the knowability principle of the form 'p and it is not known that p'. And while this is also the case with other factive operators, we have already noted that, depending on the level-bridging principle under consideration, there may not always be similarly straightforward counterexamples to the Fitch Principle, since $\Box(p \wedge \neg\Box p)$ may not always be a contradiction.

In fact, there are unlikely to be highly general and abstract considerations that adjudicate between the Fitch Principle and the level-bridging principles. Whether it is the level-bridging principles that should give way to the Fitch Principle or the other way around will likely have to be decided on a case-by-case basis. Each particular interpretation of $\Box$ deserves specialised attention. The task of weighing the Fitch Principle against the various level-bridging principles, for each interpretation of $\Box$, lies beyond the scope of this paper.

7 Conclusion

Hart called Fitch's result "an unjustly neglected logical gem" (1979, 164). This remark of Hart's was made four decades ago, which is more than a decade after the publication of Fitch's paper and more than three decades since Alonzo Church first conveyed the result to Fitch in 1945. The intervening years since Hart's remark have seen a massive upsurge in interest in the result, mostly focusing on its implications for the knowability principle and for anti-realism. Even so, as I hope to have shown in this paper, we have barely scratched the surface of the logical and philosophical gem that is Fitch's result. Up till now, we have failed to appreciate the full extent and generality of the result and its potential applicability to a wide variety of philosophical issues, ranging from norms for belief and whether evidence of evidence is evidence to iteration principles for normality and obligation. Once generalised, the potential of Fitch's result is limitless, constrained only by our ability to think of interesting modalities to apply it to.

But that is not all. Further underscoring Hart's remark concerning the neglected potential of Fitch's result is the fact that the possibility of extending Fitch's result in interesting ways goes beyond what we showed in this paper. In this paper, we focused on the Fitch Principle ($p \to \blacklozenge\Box p$). But we might consider, more generally, for any principle of the form $X_\Box$ $(\varphi \to \psi)$, its modally qualified counterpart, $X_\Box^\blacklozenge$ $(\varphi \to \blacklozenge\psi)$. On the alethic interpretation of $\blacksquare$, while $X_\Box$ is the principle that φ implies ψ, $X_\Box^\blacklozenge$ is the apparently weaker principle that φ implies possibly ψ. Put very generally, the lesson of Fitch's result is that in some cases, $X_\Box^\blacklozenge$ might turn out to be as strong as its unqualified counterpart, $X_\Box$. In particular, the Fitch Principle ($p \to \blacklozenge\Box p$) turns out to be as strong as its unqualified counterpart, $p \to \Box p$.

In San (forthcoming), I show that this insight can be generalised to many other interesting instances of $X_\Box$ and $X_\Box^\blacklozenge$. For example, given some weak assumptions, $4_\Box^\blacklozenge$ ($\Box p \to \blacklozenge\Box\Box p$) and $5_\Box^\blacklozenge$ ($\Diamond p \to \blacklozenge\Box\Diamond p$) turn out to jointly imply $4_\Box$ ($\Box p \to \Box\Box p$) and $5_\Box$ ($\Diamond p \to \Box\Diamond p$). This means, for instance, that interpreting $\Box$ in terms of knowledge and $\blacksquare$ in terms of alethic necessity, the apparently weak variants of KK and K¬K—according to which knowing implies the possibility of knowing that one knows and not knowing implies the possibility of knowing that one does not know—taken together, turn out to have the full logical strength of their unqualified counterparts, KK and K¬K—according to which knowing implies knowing that one knows and not knowing implies knowing that one does not know. This has important epistemological implications, which I explore in another paper.

It remains to be seen what further logical and philosophical insights can be mined from Fitch's result.

A Appendix

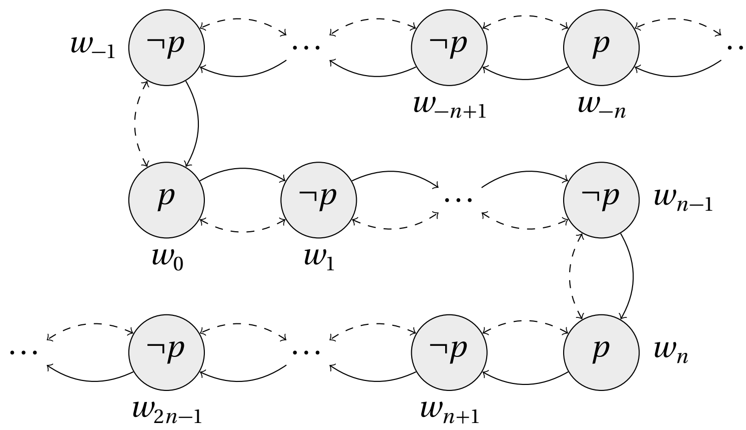
We show that the modal collapse identified in the General Collapse Theorem is, in a sense, the full extent of the collapse. More precisely:

No Collapse Theorem. *Let* $X_\Box$ *be an* n*-level bridging principle and let* L *be the smallest normal extension of* $\mathsf{KDX}_\Box \oplus \mathsf{S5}_\blacksquare$ *containing the Fitch Principle. Then, for every* $i \neq kn$*, where* k *is an integer,* $\nvdash_{\mathsf{L}} p \longleftrightarrow \Box^i p$*. (Thus, also* $\nvdash_{\mathsf{L}'} p \longleftrightarrow \Box^i p$*, for every sublogic* L' *of* L*).*

In particular, there is a model of L where for every $i \neq kn$, $p \wedge \neg\Box^i p$ is satisfied somewhere. First, let there be countably infinitely many worlds, with each one indexed by an integer. Then, let $R_\Box$ be such that each world $\Box$-accesses only the world whose index is the integer succeeding it, i.e. $\ldots$, $w_{-2}R_\Box w_{-1}$, $w_{-1}R_\Box w_0$, $w_0 R_\Box w_1$, $w_1 R_\Box w_2, \ldots$. And let $R_\blacksquare$ be the universal accessibility relation, i.e. every world $\blacksquare$-accesses every other world. Finally, let p be true only at w_{kn}, for every integer k. More formally, the model $M =< W, R_{\Box}, R_{\blacksquare}, V >$ is such that:

$$\begin{array}{lcl} W & = & \{w_i \mid i \in \mathbb{Z}\}; \\ R_{\Box} & = & \{< w_i, w_{i+1} > \mid i \in \mathbb{Z}\}; \\ R_{\blacksquare} & = & \{< w_i, w_j > \mid i, j \in \mathbb{Z}\}; \\ V(p) & = & \{w_i \mid i = kn, k \in \mathbb{Z}\}, \text{for every propositional letter } p. \end{array}$$

The resulting model is depicted below with solid arrows representing $R_{\Box}$ and dashed ones representing $R_{\blacksquare}$ (with the reflexive and transitive $R_{\blacksquare}$–arrows omitted):



Clearly, by construction, for every $i \neq kn$, $p \wedge \neg\Box^i p$ is satisfied at w_0. So, all that is left to check is that the model above is a model of L. It is a model of $\mathsf{KD}_{\Box} \oplus \mathsf{S5}_{\blacksquare}$, since $R_{\Box}$ is serial and $R_{\blacksquare}$ is universal. Furthermore, for any m, $\Diamond^m p \to \Box^{m+n} p$ is satisfied throughout the model. (Intuitively, this is because the model is constructed such that things look exactly the same from the perspective of worlds that are n-steps apart). And we can prove that, given $\mathsf{D}_{\Box}$, for any n-level bridging principle, there is some m such that $\Diamond^m p \to \Box^{m+n} p$ is stronger (in a way similar to the proof of Lemma 1 that given $\mathsf{D}_{\Box}$, for any n-level bridging principle, there is some m such that $\Box^m p \to \Diamond^{m+n} p$ is weaker). And so, if $\Diamond^m p \to \Box^{m+n} p$ is satisfied throughout the model for any m, then so is any n-level bridging principle. So, where $\mathsf{X}_{\Box}$ is an n-level bridging principle, the model above is a model of $\mathsf{KDX}_{\Box} \oplus \mathsf{S5}_{\blacksquare}$.

Finally, the model also satisfies the frame condition associated with the Fitch Principle (see the Frame Lemma), since every world $\blacksquare$-accesses some world that $\Box$-accesses only it. In particular, for every integer k, w_k $\blacksquare$-accesses w_{k-1} and w_{k-1} $\Box$-accesses only w_k. Thus, the model above is a model of L, where L is the smallest normal extension of $\mathsf{KDX}_{\Box} \oplus \mathsf{S5}_{\blacksquare}$ containing the Fitch Principle. And so, the No Collapse Theorem follows.

References

Artemov, Sergei, & Protopopescu, Tudor. 2013. Discovering Knowability: A Semantic Analysis. *Synthese*, **190**(16), 3349–3376.

Bell, David, & Hart, William D. 1979. The Epistemology of Abstract Objects: Access and Inference. *Proceedings of the Aristotelian Society*, **53**, 153–165.

Bonnay, Denis, & Égré, Paul. 2011. Knowing One's Limits: An Analysis in Centered Dynamic Epistemic Logic. *In:* Girard, Patrick, Roy, Oliver, & Marion, Mathieu (eds), *Dynamic Formal Epistemology*. Springer.

Bykvist, Krister, & Hattiangadi, Anandi. 2007. Does Thought Imply Ought? *Analysis*, **67**(4), 277–285.

Bykvist, Krister, & Hattiangadi, Anandi. 2013. Belief, Truth, and Blindspots. *In:* Chan, Timothy (ed), *The Aim of Belief*. Oxford University Press.

Carter, Sam. forthcoming. Higher-Order Ignorance Inside the Margins. *Philosophical Studies*.

Chalmers, David. 2012. *Constructing the World*. Oxford University Press.

Chase, James, & Rush, Penelope. 2018. Factivity, Consistency and Knowability. *Synthese*, **195**, 899–918.

Chellas, Brian. 1980. *Modal Logic: An Introduction*. Cambridge University Press.

Costa-Leite, Alexandre. 2004. Combining Possibility and Knowledge. *In:* Carnielli, F., Dionisio, F., & Mateus, P. (eds), *Proceedings of the Workshop on the Combination of Logics: Theory and Applications*. IST-Department of Mathematics, Center for Logic and Computation.

Costa-Leite, Alexandre. 2006. Fusions of Modal Logics and Fitch's Paradox. *Croatian Journal of Philosophy*, **6**(17), 281–290.

Cresto, Eleonora. 2017. Lost in translation: Unknowable Propositions in Probabilistic Frameworks. *Synthese*, **194**, 3955–3977.

Edgington, Dorothy. 1985. The Paradox of Knowability. *Mind*, **94**, 557–568.

Edgington, Dorothy. 2010. Possible Knowledge of Unknown Truth. *Synthese*, **173**(1), 41–52.

Égré, Paul. 2014. Epistemic Logic. *In:* Horsten, Leon, & Pettigrew, Richard (eds), *The Bloomsbury Companion to Philosophical Logic*. Bloomsbury.

Feldman, Richard. 2014. Evidence of Evidence is Evidence. *In:* Matheson, & Vitz (eds), *The Ethics of Belief*. Oxford University Press.

Fischer, Martin. 2013. Some Remarks on Restricting the Knowability Principle. *Synthese*, **190**(1), 63–88.

Fitch, Frederic. 1963. A Logical Analysis of Some Value Concepts. *The Journal of Symbolic Logic*, **28**, 135–142.

Fitelson, Branden. 2012. Evidence of Evidence is not (Necessarily) Evidence. *Analysis*, **72**(1), 85–88.

Greco, Daniel. 2015. Iteration and Fragmentation. *Philosophy and Phenomenological Research*, **91**(3), 656–673.

Holliday, Wesley H., & Icard III, Thomas F. 2010. Moorean Phenomena in Epistemic Logic. *In:* Bekelemishev, L., Goranko, V., & Shehtman, V. (eds), *Advances in Modal Logic*, vol. 8. College Publications.

Immerman, Daniel. forthcoming. Does Ought Imply Ought Ought? *Philosophical Quarterly*.

Kvanvig, Jonathan. 2006. *The Knowability Paradox*. Oxford University Press.

Mackie, John L. 1980. Truth and Knowability. *Analysis*, **40**(2), 90–92.

Maffezioli, Paolo, Naibo, Alberto, & Negri, Sara. 2012. The Church-Fitch Knowability Paradox in the Light of Structural Proof Theory. *Synthese*, **190**(14), 2677–2716.

Rieger, Adam. 2015. Moore's Paradox, Introspection and Doxastic Logic. *Thought: A Journal of Philosophy*, **4**, 215–227.

Salerno, Joe. 2009. Knowability Noir: 1945-1963. *In:* Salerno, Joe (ed), *New Essays on the Knowability Paradox*. Oxford University Press.

San, Weng Kin. forthcoming. Disappearing Diamonds: Fitch-Like Results in Bimodal Logic. *Journal of Philosophical Logic*.

Smith, Martin. 2006. Ceteris Paribus Conditionals and Comparative Normalcy. *Journal of Philosophical Logic*, **36**(1), 97–121.

Smithies, Declan. 2012. Moore's Paradox and the Accessibility of Justification. *Philosophy and Phenomenological Research*, **85**(2), 273–300.

Smullyan, Raymond M. 1988. *Forever Undecided: A Puzzle Guide to Gödel.* Oxford University Press.

Sorensen, Roy. 1988. *Blindspots.* Clarendon Press.

Stjernberg, Fredrik. 2009. Restricting Factiveness. *Philosophical Studies*, **146**, 29–48.

Whiting, Daniel. 2013. Nothing but the Truth: On the Norms and Aims of Belief. *In:* Chan, Timothy (ed), *The Aim of Belief.* Oxford University Press.

Williamson, Timothy. 1993. Verificationism and Non-Distributive Knowledge. *Australasian Journal of Philosophy*, **71**(1), 78–86.

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

Williamson, Timothy. forthcoming. Edgington on Possible Knowledge of Unknown Truth. *In:* Hawthorne, John, & Walters, Lee (eds), *Conditionals, Probability, and Paradox: Themes from the Philosophy of Dorothy Edgington.* Oxford University Press.