A paradox of modal knowledge

Seth Yalcin
yalcin@berkeley.edu

July 18, 2023

1

An urn contains a mix of black and white marbles only. A marble is selected at random from the urn and placed under a cup. We don't look under the cup. Concerning the marble under it, these are true:

(1) It might be black.

(2) It might be white.

Now Carl looks under the cup without letting anyone else see, and as a result comes to know what color the marble is. This is true:

(3) Carl knows it's black or Carl knows it's white.

Carl then sits down in front a blackboard where (1) and (2) are written. He reads them and rationally reflects on the truth of each, bringing to bear his knowledge of the color of the marble. This seems like it'd be true:

(4) Carl knows that (1) and (2) are not both true.

Of course, if knowledge is factive, (4) entails:

(5) (1) and (2) are not both true.

It appears we have fallen into contradiction. Where did we go wrong?

2

The main challenge of the case is the apparent joint consistency of (1), (2), and (4). If these are consistent, then since (1), (2), and (5) are obviously not jointly consistent, the inference from (4) to (5) must fail—which appears to mean that factivity fails.

I will discuss three options for dissolving the paradox. The first option says that the inference from (4) to (5) does fail, but that despite appearances, it isn't actually an instance of the factivity inference, for (4) ascribes to Carl knowledge of a proposition that is not the one negated by (5). The idea will be that something about the context-sensitivity of the epistemic modals implicitly involved here generates an equivocation in the move from (4) to (5).

The second option rejects the joint consistency of (1), (2), and (4). It says that (4) gains the spurious appearance of plausibility from the obvious joint consistency of (1), (2), and (3), together with the mistaken idea that (3) entails, given suitable closure assumptions for knowledge, (4). I explicitly stipulated (3) in the case in order to discuss this response.

The third option rejects factivity. At least at first glance, this should strike us as the option of last resort. So let us consider the other two options before that.

3

Take *Factivity* to be the claim that for all ϕ, $K\phi \vDash \phi$.[1] The first idea denies that (4) and (5) are of the form $K\phi$ and ϕ, respectively. What view could make sense of this? Here is a vague idea that has often been suggested to me. It is that when we assess (1) and (2), we naturally take them to be describing something about *our* evidence or information—the relevant modals are evaluated relative to our context—whereas when we assess (4), we take it to be saying something about what Carl knows about *his* (not our) evidence—it is saying that he knows that his evidence, or the

[1] Linguists would call this the claim that 'knows' is *veridical*, and would use *factive* instead to label its presuppositional dimension.

information within the epistemic reach of his context, doesn't leave both black and white possibilities open. So the move from (4) to (5) is indeed invalid, but not because knowledge isn't factive; rather, there is equivocation in the interpretation of the modals implicitly at issue, the difference being to do with how their context-sensitivity is resolved.

Without more bells and whistles, this view would imply that many cases that superficially *look like* a factivity entailment are not. Take for instance:

> Alice knows that it might be raining.
>
> Therefore, it might be raining.

This may appear to be of the form: $K \Diamond \phi \vDash \Diamond\phi$. But the sort of view just described would interpret the premise as ascribing to Alice knowledge of a modal proposition that is generally different than the one that the conclusion expresses—and so in the relevant sense this will not be true instance of factivity. (The form of the argument is thus more like: $K \Diamond_1 \phi \vDash \Diamond_2\phi$.) Indeed this is predicted to be an invalid argument, since given just that that Alice knows that *her* information is compatible with rain, nothing follows about whether rain is compatible with *our* information.

The equivocation view has a basic problem. It is a familiar point that sentences like these are generally marked:

(6) ??It's not raining and Alice knows that it might be raining.

By denying that 'It might be raining' follows from 'Alice knows that it might be raining', the equivocation view gives up just the entailment needed to explain the trouble with (6). For if we have that entailment, then (6) entails:

(7) ??It's not raining and it might be raining.

The problem of explaining the defect in sentences like (6) then reduces, in a natural way, to the problem of explaining the defect in sentences like (7). The equivocation view not only cannot embrace this natural line of explanation; it has trouble making sense the problem with (6) in the first place. After all, this sort of thing is intuitively consistent:

(8) It's not raining and Alice knows that her evidence is compatible with rain.

But the equivocation view likens 'Alice knows that it might be raining' to the second conjunct here; so more would need saying about why (6) and (8) differ in markedness.[2]

The simple point here is that in general, at least for nonmodal ϕ, sentences of the form $K\Diamond\phi$ *do* intuitively imply $\Diamond\phi$.[3] This is hard to understand if the factivity of knowledge isn't what explains it.

There is of course a grain of truth in the equivocation view: epistemic modals involve a relativity to a body of information. The equivocation view assimilates this relativity to context-sensitivity in the sense of Kaplan [1977/1989]: epistemic modals have a character that is variable across contexts in what is contributed to content. But a different idea is that the relativity in question is more like the relativity of the truth of a sentence to a possible world. We don't "equivocate" about the meaning of a sentence as we consider its truth value relative to different worlds; rather, its meaning just is partly a function on worlds. Likewise, we could say, an epistemic modal clause is is the sort of thing that is true (or is accepted, supported, etc.) *relative to* a body of information; its meaning is at least partly a function of information states. This is one informal way of putting the idea, first developed in Veltman [1996], that epistemic modals are *information-sensitive*.[4] That idea is popular, and can be implemented in many different formal settings.[5] Typical developments of it have the result that sentences

[2]It also would not help to hold that knowledge ascriptions embedding epistemic modals have multiple readings, with the modal sometimes sensitive to the ascriber and sometimes to the subject. Where there are two readings and only one is felicitous in the context, we expect the charitable listener to select the felicitous reading.

[3]As has been assumed in the literature: see, e.g., Yalcin [2012b], Moss [2013, 2018]. Beddor and Goldstein [2021] provide additional motivating data.

[4]The term is from Kolodny and MacFarlane [2010]. See Rothschild and Yalcin [2017] for one abstract discussion.

[5]See for instance Groenendijk et al. [1996], Beaver [2001], Gillies [2004, 2010, 2018], Yalcin [2007, 2012a, 2015], Kolodny and MacFarlane [2010], Klinedinst and Rothschild [2012], Willer [2013], Starr [2014], Moss [2015, 2018], Goldstein [2019], Dever and Schiller [2021], Hawke and Steinert-Threlkeld [2021], Ninan [2021], Santorio [2022]. (Note that one can merge information-sensitivity with a variety of contextualism; this would a "shiftable"

like (6) are a variety of contradiction—they are *epistemic* contradictions. This idea will play a prominent role in our discussion of the second option, which we turn to now.

4

Recall that the second option doesn't question the validity of inference from (4) to (5). Instead it rejects the joint consistency of (1), (2), and (4). How then does this view explain the appeal of (4) in the example?

To bring the challenge in focus, let me draw out (4)'s appeal. We do not know exactly what Carl knows, but we know (3). So we appear to know that whatever Carl knows, it would be in tension—in epistemic contradiction—with one or the other of (1) and (2)—and that Carl can see that. If Carl could know (1) and (2) are both true, one of these would have to be true:

(9) ??Carl knows that it's white and might be black.

(10) ??Carl knows that it's black and might be white.

But these don't make sense. Note that things are not improved if the knowledge operator is distributed across the conjuncts:

(11) ??Carl knows it might be black, and he knows it's white.

(12) ??Carl knows it might be white, and he knows it's black.

These facts are obviously tied to the familiar point that it would be weird for Carl to say something like:

(13) ??I know it's black, but it might be white.[6]

---

or "parametric" contextualism, rather than an indexical contextualism (on the distinction, see Ninan [2010], Yalcin [2022]). But such views also face the problem described above, about capturing the inference from $K \Diamond \phi$ to $\Diamond\phi$.)

[6] "To speak of... knowledge despite uneliminated possibility of error, just *sounds* contradictory", as Lewis [1996] noted. Discussion of these "concessive knowledge attributions" in epistemology goes back to Unger [1975] and Dretske [1981]. Note that in a formal context incorporating epistemic contradiction and the factivity of knowledge, we can say that these constructions sound contradictory because they are.

What this brings out is that if (4) is false, it's not because this is true:

(14) Carl knows that (1) and (2) are both true.

But to reject (4) while also rejecting (14) is a odd position. (1) and (2) are stipulated to be true in the example. What is it that they say that Carl doesn't know? More worryingly, it appears that whatever they say, Carl could never know both of them while retaining the knowledge he now has—else one of (9) or (10) (and one of (11) or (12)) would be true. Considering the matter from Carl's perspective, it is hard to see room for the sort of agnosticism this view implies should be available. ("I know that it's black, but a question remains: might it be white?" That doesn't sound coherent.)

These observations suggest that at least for nonmodal ϕ, pairs of the form $\{K\phi,\ K\Diamond\neg\phi\}$ and $\{K\neg\phi,\ K\Diamond\phi\}$ are inconsistent. Among the attitude verbs that permit general embedding of epistemic modals, it is convenient to have a label for those that give rise to this kind of inconsistency:

> An attitude operator A is *strong* just in case for any nonmodal ϕ, $\{A\phi,\ A\Diamond\neg\phi\}$ is inconsistent; otherwise it is *weak*.

This terminology is inspired by (though not identical to) that of Hawthorne et al. [2016], Rothschild [2020], who along with Beddor and Goldstein [2018] motivate the thesis that belief operators are weak in the above defined sense (i.e., $B\phi$ may be compatible with $B\Diamond\neg\phi$). The evidence reviewed so far suggests that knowledge operators are strong.[7]

Indeed, the evidence plausibly suggests something stronger, viz., that if one knows that it's white (black), one is thereby positioned to know that it's not the case it might be black (white). This is the simple thought

[7]For further examples of attitudes that are plausibly strong, consider supposition, imagination, and the attitude we strike towards the things taken for granted in conversation—what we might label *presupposition* (Stalnaker [1974]) or *acceptance* (Stalnaker [2014]). (E.g., it doesn't seem possible to suppose/imagine/conversationally presuppose that it's raining, while also supposing/imagining/presupposing that it might not be raining.) Note that these are nonfactive attitudes; thus strength doesn't imply factivity. But given that sets like $\{\phi,\ \Diamond\neg\phi\}$ and $\{\neg\phi,\ \Diamond\phi\}$ are inconsistent, the factivity of an attitude will imply that it is strong.

underneath the intuition that (4) is true in the example, I suggest. Defining $\Box\phi := \neg \Diamond \neg\phi$, the relevant generalization is:

Epistemic Strength. For all nonmodal ϕ, $K\phi \vDash K \Box \phi$.[8]

This would follow from

Entailment Strength. For all nonmodal ϕ, $\phi \vDash \Box\phi$.[9]

—which captures a basic part of the intuition of epistemic contradiction; and the assumption that knowledge is suitably closed under entailment ($\vDash$).[10] I will assume the latter throughout.[11]

Granting Entailment Strength, we can derive $\phi \vDash \neg(\Diamond\phi \wedge \Diamond\neg\phi)$, supposing there is no problem with the classically valid inference $\neg\varphi$ to $\neg(\varphi \wedge \psi)$ for any φ, ψ. Moving forward we'll have a need to refer to sentences of the form $\Diamond\phi \wedge \Diamond\neg\phi$ often, so let me introduce the following abbreviation: $\otimes\phi := \Diamond\phi \wedge \Diamond\neg\phi$. Where $\otimes\phi$ is true, we can describe ϕ as *open*.

[8] I emphasize we have *defined* $\Box$ as a notational shorthand for $\neg \Diamond \neg$. Thus Epistemic Strength is a generalization concerning epistemic *might*, motivated here with data involving that modal exclusively. It will imply claim about epistemic *must* with suitable additional assumptions; but as we won't discuss *must*, we can save the matter for another day. Of course we do expect *might* and *must* to turn out to be duals, or something close; but *must* arguably has a distinctive evidential aspect (see, e.g., Stone [1994], von Fintel and Gillies [2010, 2021]), and engaging this aspect here would only bring orthogonal complications into the dialectic.

[9] This is equivalent to what Yalcin [2007] calls "Lukasiewicz's Principle" under the assumption that that ϕ and $\neg\neg\phi$ are semantically equivalent (intersubstitutable) for nonmodal ϕ.

[10] Beddor and Goldstein [2021] defend a position that implies Entailment Strength but which is incompatible with Epistemic Strength (even restricted to ideal knowers). On their view, knowledge is true safe belief, but on their analysis of safety, safely believing $\Box\phi$ is strictly harder than safely believing ϕ. As a result, knowledge is not preserved under $\vDash$, even for relevantly ideal knowers. This disconnect between consequence and knowledge seems unwelcome. (Thanks here to Simon Goldstein for discussion.)

[11] Moss [2018] offers a penetrating discussion of some cases similar to the puzzle case of this paper (see especially ch. 7.2-4). She does not consider dropping Factivity in response to them. Her view implies that (1), (2), and (4) are (on the relevant readings) inconsistent. Based on her discussion, it appears she might offer to explain our case by assimilating it to a traditional skeptical argument that relies on closure under known entailment. I have trouble seeing how such an account would go, but her discussion deserves a fuller assessment than I can provide here. Meanwhile you can read me as offering Moss another option for responding to the apparent threat posed by what she calls "heretical knowledge".

So to restate, Entailment Strength yields $\phi \vDash \neg \diamondplus \phi$ for all nonmodal ϕ. Thus Epistemic Strength and the closure of knowledge under $\vDash$ let us conclude:

(i) For all nonmodal ϕ, $K\phi \vDash K\neg \diamondplus \phi$

Substituting $\neg\phi$ for ϕ and simplifying,[12] this implies:

(ii) For all nonmodal ϕ, $K\neg\phi \vDash K\neg \diamondplus \phi$

Putting (i) and (ii) together, the following appears to obviously follow:

Decisiveness. For all nonmodal ϕ, $K\phi \vee K\neg\phi \vDash K\neg \diamondplus \phi$.[13]

Now for convenience we can view (3) as of the form $K\phi \vee K\neg\phi$ (thinking of 'black' as short for 'non-white'[14]); and view (4) as equivalent to a claim of the form $K\neg \diamondplus \phi$. Thus Decisiveness implies that (3) implies (4). Thus the claim that the triplet of (1), (2), and (4) is inconsistent implies that the triplet of (1), (2), and (3) is also inconsistent. But the latter claim is, intuitively, completely implausible. So I submit that this is a reductio of the view that (4) really is incompatible with (1) and (2).

To sum up: Epistemic Strength is well-movated. It makes good sense of the judgements we have about (9)-(13). Its truth is to be expected if Entailment Strength is true, and to that extent, it is already motivated by the basic data around epistemic contradiction motivating the latter. And Epistemic Strength implies Decisiveness. So besides the direct superficial appearance that the tale of Carl was consistent up to and including the stipulation of (4), we can appeal also to Decisiveness, plus the intuitive

---

[12] The specific assumption here is that $K\neg(\Diamond\neg\phi \wedge \Diamond\neg\neg\phi)$ is equivalent to $K\neg(\Diamond\phi \wedge \Diamond\neg\phi)$.

[13] A more general form of the idea is:

> **Generalized Decisiveness**. For all nonmodal ϕ, ψ, $(K\phi \vee K\psi) \wedge K\neg(\phi \wedge \psi) \vDash K\neg(\Diamond\phi \wedge \Diamond\psi)$.

This principle can be motivated by the same considerations favoring Decisiveness. I fix on the latter in order to simplify the dialectic.

[14] That this is merely a convenience will, I hope, become clearer below. See also fn. 13.

consistency of the triplet (1), (2), (3), in defense of the consistency of (4) with (1) and (2).

I am about to conclude that the option of denying that (1), (2) and (4) are jointly consistent is too costly. But first we must consider an important sort of challenge to the reasoning above. The challenge holds that from (i) and (ii), Decisiveness does not follow. In defense of this view, the challenger observes that in a closely related context, the analogous inference clearly does fail (cf. Moss [2018, 7.4]). Consider in particular the inference from the general validity (for nonmodal ϕ) of these inferences:

$$\phi \vDash \neg \diamondtimes \phi \qquad \neg\phi \vDash \neg \diamondtimes \phi$$

to the validity of what we could call

Decisive Disjunction. $\phi \vee \neg\phi \vDash \neg \diamondtimes \phi$

We observed that the truth of $\phi \vDash \neg \diamondtimes \phi$ (and by extension $\neg\phi \vDash \neg \diamondtimes \phi$) is just a short step from Entailment Strength. But the inference from these to Decisive Disjunction must be blocked, as the latter is subject to clear counterexamples. From:

(15) It's black or it's white.

we obviously cannot infer the negation of

(16) It might be black and it might be white.

Indeed, the conjunction of (15) with (16) is plainly consistent, as Holliday and Mandelkern [2022] stress. The observation is continuous with the various failures of constructive dilemma, modus tollens, De Morgan's Laws, and distributivity that epistemic modals have been argued to give rise to in the literature.[15]

The invalidity of Decisive Disjunction is what explains the invalidity of:

[15] For discussion of these and other failures of classical principles, see for instance Gillies [2004], Cantwell [2008], Kolodny and MacFarlane [2010], Yalcin [2012b, 2015], Moss [2015, 2018], Mandelkern [2019, 2020], Holliday and Mandelkern [2022].

Wide Decisiveness. $K(\phi \vee \neg\phi) \vDash K\neg \diamondtimes \phi$

But the difference between Decisiveness and Wide Decisiveness is the key to the matter. Intuitively, there are three logically possible ways for $K(\phi \vee \neg\phi)$ to be true. It can be true while $K\phi$ is true; it can be true while $K\neg\phi$ is true; or it can be true although neither $K\phi$ nor $K\neg\phi$ is true. The third way of knowing the disjunction is intuitively compatible with the truth of $K \diamondtimes \phi$. That is what Wide Decisiveness misses, and why it is invalid. But the first two ways of knowing $\phi \vee \neg\phi$ intuitively *are* incompatible with knowing $\diamondtimes\phi$, as already stressed above, drawing on Epistemic Strength. And when we say $K\phi \vee K\neg\phi$, what are we saying if not that the agent in question is one of these two ways? We are excluding the third way—the only way compatible with $K \diamondtimes \phi$. This is why sentences that are tantamount to $K \diamondtimes \phi \wedge (K\phi \vee K\neg\phi)$ sound weird:

(17) Bill knows that it might be heads and it might be tails. ??He also knows whether it's heads.

If we think of $\vDash$ in broadly informational terms,[16] as would be natural in a context recognizing epistemic contradictions as contradictions, this problem with Wide Decisiveness just dramatizes the problem with Decisive Disjunction. The latter would erase the third, "indecisive" way for a state of information to incorporate disjunctive information.

If this is why Decisive Disjunction and Wide Decisiveness fail, their failure doesn't impugn Decisiveness. That is good news, since for the reasons already reviewed, it seems hard to square the failure of Decisiveness with Epistemic Strength. There of course is a real problem here about finding a semantics for $\vee$ and $\diamond$ that invalidates Decisive Disjunction while making sense of all of the intuitively valid instances of constructive dilemma and distributivity.[17] My aim here hasn't been to give such a theory, but rather

[16] That is, in terms of some version of the idea that $\Gamma \vDash \phi$ just in case any information state that accepts (supports, incorporates, etc.) all the sentences in Γ must also accept ϕ. See Bledin [2014] for one general discussion.

[17] On this challenge, see Moss [2018], Mandelkern [2019], Hawke and Steinert-Threlkeld [2021], Holliday and Mandelkern [2022].

to bring into focus a case for the view that any such theory ought to validate Decisiveness.

Ultimately, the case for Decisiveness is a case against Factivity. We cannot have both, since together they would imply:

Overdecisiveness. For all nonmodal ϕ, $K\phi \vee K\neg\phi \vDash \neg \otimes \phi$

Our main puzzle case already illustrates why this is wrong, but for the sake of variety: suppose I know John is at home or at the office, but I am not sure which; and suppose I know *John* knows where he himself is, wherever that is. Does this already mean I can't know both that he might be home, and that he might be at the office? It seems intuitively obvious I can know both of these, while also knowing that John knows where he himself is. But if we hold Decisiveness fixed, we must give up Factivity in order to say this.

Let me summarize. There are consistent instances of $\otimes\phi \wedge (K\phi \vee K\neg\phi)$. Decisiveness is plausible. So it is plausible that there are consistent instances of $\otimes\phi \wedge K\neg \otimes \phi$—e.g., the conjunction of (1), (2), and (4).

Indeed, our case recommends the idea that where we have distinct knowers x, y, it is possible that $K_x \otimes \phi \wedge K_y \neg \otimes \phi$, and also possible that $K_x(\otimes\phi \wedge (K_y \neg \otimes \phi))$. To draw this out, let us add an agent to our main puzzle case. Suppose Jane knows everything we know about the case, but not more (she's our avatar, as it were). Jane is normal and relevantly logically competent, and knows Carl is the same. By stipulation, Jane knows that (1) and (2) are true; but our case for (4) remains; so we have an instance of $K_j \otimes \phi \wedge K_c \neg \otimes \phi$. By stipulation Jane also knows knows (3). Thus given Decisiveness, it follows that:

(18) Jane knows Carl knows that (1) and (2) aren't both true.

This just says that she knows what we know, when we know (4).[18]

To say any of this consistently, we must drop Factivity. Let's now engage this option.

[18] One might worry that these results predict that Jane can say things like "Carl knows that I'm wrong!" or "Carl disagrees with me, and he's right!" Not so. What the case demands is just that Carl has strictly more relevant information that Jane, and that Jane knows that; but in general, when B has strictly more information than A, it does not follow that B knows that A is wrong about something, or that B disagrees somehow with

5

To see that dropping Factivity is not an *ad hoc* reaction to the facts, we can begin by noticing that putative factive entailment of (4), viz., (5), is something that agents epistemically positioned like us (or Jane) *can do nothing with*. To bring this out, suppose we ask Carl for a hint about the marble's color, and he replies: "I'll tell you this much..." and then he utters (5). Intuitively, this tells us nothing (save that Carl has a bad sense of humor). Not that we don't trust Carl, and not that Carl doesn't speak from knowledge (it would seem weird to retort, "You don't know that!"). Rather, even granting that we trust Carl and that he speaks whereof he knows, still, it would seem like a mistake to respond to Carl's assertion of (5) by ceasing to consider (1) and (2) both true.

Why is this? It seems that it is because we have no rational path for transitioning the state of information we are in to one that accepts (5). To do that, we'd have to give up one of (1) or (2). But this is tantamount to settling on a view about the marble's color.[19] And (5) gives us no new information about the marble's color. So it provides us with no direction for which of (1) or (2) to drop. But we cannot just pick one of them to drop at random. So we have no rational way to update with (5). To transition from the state of information we are in to one that accepts (5), we have to learn the marble's color. This is exactly why Carl's "hint" seems like a tease. Although Carl's knowing (5) is, we know, a substantive feature of

A. Disagreement arises, not when agents have different information, but when the agents don't agree about who has more information; but Jane (and we can stipulate, Carl) are on the same page about how they are informationally ordered. What is unusual about our case is that we see violations of the generalization that for any x and y, if x has strictly more information that y, then $K_y\phi$ only if $K_x\phi$. Though Carl has strictly more information, Jane knows (16), while Carl doesn't. What this implies is that Jane's knowing that isn't knowledge of some fact, some way the world is, that Carl lacks. Instead it must have something more to do with with her ignorance—with what her epistemic state leaves open (cf. Yalcin [2011]).

[19] Is agnosticism not available? Can't we just cease thinking that (1) and (2) are both true, without needing to positively reject one of them? The challenge is to make sense of this. We started out knowing (1) and (2). Then Carl says (5). Given (3) and Decisiveness, we already knew he knew that. Why should his saying (5) demote the conjunction of (1) and (2) in our view? What is to (i) know that it's black or white, (i) not know whether it is black (white), and yet (iii) not know whether it might be black and might be white?

his epistemic state founded on his knowledge of the marble's color, (5) is informationally inert from our perspective: Carl's testifying (5) is not a way for us to get to a knowledge state that supports (5).

Write the *update* of information state i with ϕ as: $i[\phi]$. The preceding suggests a constraint that any definition of the update function should satisfy: for any nontrivial state i and nonmodal ϕ, ψ, if $i \vDash \phi \vee \psi$ and $i \vDash \Diamond\phi \wedge \Diamond\psi$, then $i[\neg(\Diamond\phi \wedge \Diamond\psi)]$ is undefined. The need for this constraint on update in turn motivates the following constraint on information states in general:

> **Disjunctive Liveness.** For any state i and nonmodal ϕ, ψ, if $i \vDash \phi \vee \psi$, then $i \vDash \phi$ or $i \vDash \psi$ or $i \vDash \Diamond\phi \wedge \Diamond\psi$.

The idea is that whatever information states are, there are none that accept $\phi \vee \psi$ while failing to accept any of ϕ, ψ, or $\Diamond\phi \wedge \Diamond\psi$.[20] If Disjunctive Liveness is right, we cannot come to accept (5) while retaining our knowledge that it is black or white except by leaping to a conclusion about the color. But however the update function is defined, it is clear that it ought not sanction such leaps. So we should expect [(5)] to be undefined on the state of information we (and Jane) are in, as it indeed seems to be.[21]

Our inability to update on (5) evidently doesn't threaten our ability to update on (4)—not to mention (3), which (we have argued) implies (4). This is not what Factivity predicts: if Factivity holds, any problem updating i with ϕ is a problem we expect $K\phi$ to inherit. In this way Factivity makes the facts harder to understand, not easier.

Factivity also stand in the way of what otherwise looks to be straightforward story about what it is for Jane to know (1)-(4). Though there isn't space here to motivate a detailed model of mighty knowledge, one might

[20]For an example of a model of information states that would allow for states i that accept $\phi \vee \psi$ without accepting any of ϕ, ψ, or $\Diamond\phi \wedge \Diamond\psi$, see for instance Willer [2013].

[21]To unpack this a bit more: I assume a theory of update will at least implicitly define a notion of "informational nearness" such that in general for any ϕ and i, $[\phi]$ maps i to the nearest i' to i such that $i' \vDash \phi$, if such there be, and otherwise $[\phi]$ is undefined. The claim then is that for nonmodal ϕ, ψ, if $i \vDash \phi \vee \psi$, $i \nvDash \phi$, $i \nvDash \psi$ and $i[\phi] \neq i[\psi]$, then given Disjunctive Liveness, $i[\phi]$ and $i[\psi]$ are the two most plausible candidates for $i[\neg(\Diamond\phi \wedge \Diamond\psi)]$. But (I claim) no legitimate formalization of nearness will reckon one of $i[\phi]$ or $i[\psi]$ informationally closer to i. So we expect $i[\neg(\Diamond\phi \wedge \Diamond\psi)]$ not to be defined.

expect a reasonable model to imply at least the following. First, since Jane knows that it's black or white but not which, her epistemic state leaves open live alternatives of both sorts. Second, since Jane knows that Carl knows what color it is, this means that at each of Jane's live epistemic alternatives, Carl's epistemic state does *not* leave open alternatives of both sorts: at Jane's black (white) alternatives, Carl knows it's black (white). What Decisiveness recommends is the idea that Jane's knowledge of (4) is basically grounded in this latter fact about her epistemic alternatives. If so, then while there are are real challenges (reviewed in Beddor and Goldstein [2021]) about explaining exactly what sort of "leaving open" or "liveness" makes for mighty knowledge, there would seem to be nothing fundamentally mysterious about Jane's overall epistemic state, when she knows (1)-(4); the state is perfectly ordinary and consistent. Factivity throws a wrench into the story, by requiring that Jane's mighty knowledge be compatible with what she takes Carl to know. But this demand is unmotivated if mighty knowledge is the sort of thing whose loss is compatible with strict information gain. And that is what it seems to be, naively: for one can go from a state of knowing both (1) and (2) to a state of *not* knowing both just by learning the marble's color.

The failure of (4) to imply (5) would certainly be mystifying if the latter described a way the world could be (i.e., it had an ordinary possible worlds truth-condition), and the former reported that Carl knows the world is that way. But evidently, that is not the proper understanding of either sentence. (5), like (1) and (2), is information-sensitive. And given Decisiveness, (4) does not say more than (3) does. To be sure, (4) characterizes Carl's epistemic state, but not by naming one of the propositions he knows. If (4) does not say more than (3) does, then whatever exactly it says about Carl's state of knowledge, surely it must be compatible with what we know, when we know (1) and (2), since again, (1), (2), and (3) are jointly compatible.

Our case shows that Factivity overgenerates. But I do not deny that Factivity makes desirable predictions most everywhere else. So there must be another generalization nearby that is true. An obvious one is Nonmodal Factivity: the claim that for all nonmodal ϕ, $K\phi \vDash \phi$. We have seen no reason to doubt this claim, and nothing we have said stands in the way of accepting it. But whereas Factivity overgenerates, Nonmodal Factivity undergenerates: we have already observed that we want $K \Diamond \phi \vDash \Diamond\phi$ for nonmodal ϕ, for instance. So although our counterexamples to factivity involve $\Diamond$, there is no easy solution that merely excludes all modal talk from factivity inferences. We need some more subtle weakening of Factivity, one that suitably weeds out curiosities like $\neg \otimes \phi$ and its kin.

6

In looking for the generalization that does better than Factivity, we need a better understanding of these curiosities. Moss [2018] identifies the key notion we need. Sentences like $\otimes\phi$ are plausibly what she calls *heretical*. In the context of Moss's theory, 'heretical' labels a certain variety of probabilistic content, but (I think Moss would agree) the basic idea of heretical content can be generalized beyond a probabilistic setting. In essence, heretical content is content that is incompatible with full information—with any proposition that completely characterizes a state of the world (actual or possible). Contradictions are trivially heretical, but the idea is that there is a nontrivial variety. The conjunction of (1) and (2) (viz., (16)) is plausibly heretical because any full state of information will settle the marble's color, and thereby rule out one of the conjuncts; thus it is not the sort of thing that could be accepted by God's state of information. The more general case is captured by sentences like $(\Diamond\phi \wedge \Diamond\psi) \wedge \neg(\phi \wedge \psi)$. For instance:

(19) Alice might be tall and she might be rich, but she's not both.

Though the relevant ϕ and ψ here are not logically incompatible, it is enough that the acceptability of (19) precludes their joint truth. As a result, whatever Alice is like, there is no possible way to be fully informed about Alice's properties and also know (19).

Still more generally, we may view hereticality as state-relative. While (19) is heretical relative to any state,

(20) Alice might be tall and she might be rich.

is heretical just relative to states that include the information that Alice is not both tall and rich. (For convenience we formalized (16) as $\neg \otimes \phi$, but we could just as well have treated its heredicality as state-relative.)

To get to a definition of this notion, suppose we have found some reasonable notion of *information state* and an affiliate notion of $\vDash$ defined on a language containing $\Diamond, K$, and the connectives with at least the following features. (i) States are partially ordered by informativeness, so that $i \geq j$ just in case i includes at least as much information as j. (ii) There is an *absurd* state $\varnothing$, the maximal element of the ordering; it is the only state accepting inconsistent sentences. (iii) There are "divine" states of omniscience or full information, where i is *divine* just in case $i \neq \varnothing$ and for any j, if $j > i$, then $j = \varnothing$ (adapting Moss's terminology). So far these imply that for any ordinary state i, there is always some divine j such that $j > i$—where i is *ordinary* if it is neither absurd nor divine. (iv) Finally, if i is divine, then for all ϕ, $i \vDash \phi$ or $i \vDash \neg\phi$. I take these to be rather ordinary assumptions, and ones that diverse models might satisfy. In the context of a model of this sort, we can define hereticality like this:

> ϕ is *heretical* at i just in case for every divine $j \geq i$, $j \vDash \neg\phi$.

Say that ϕ is *strongly* heretical just in case ϕ is heretical at all states. For nonabsurd i, we can say that ϕ is *interestingly* heretical at i just it case ϕ is heretical at i and $i \vDash \phi$.

Following Moss before us, what we have been seeing throughout is the case for interestingly heretical content, and for the possibility of heretical knowledge. The idea that there exist states i and sentences ϕ, ψ such that $\Diamond\phi \wedge \Diamond\psi$ is interestingly heretical at i depends on Veltman [1996]'s influential idea that it is possible to go from a state of accepting $\Diamond\phi$ to one that doesn't, *without* losing information—on the idea that $\Diamond\phi$ is not generally persistent (where φ is *persistent* just in case for any i, when $i \vDash \varphi$ and $j \geq i$, then $j \vDash \varphi$). The curious thing about knowing ϕ, where ϕ is heretical, is that it is a foregone conclusion that for any possible way you might strictly gain information and arrive at divine state, you will end up at a divine state of knowing $\neg\phi$. That's the curiosity at the heart of our case: we (and Jane) know (16), while also recognizing that Carl, who is relevantly fully informed, knows its negation. This is in effect to recognize that once *we* learn strictly more (i.e., the marble's color), we'll come to reject a thing we take ourselves to now know.

Say that i is *destined* for ϕ, $i \mapsto \phi$, just in case for all divine $j \geq i$, $j \vDash \phi$; say that ϕ is *delayed* at i just in case $i \mapsto \phi$ but $i \nvDash \phi$; and say that ϕ is *delayable* in case it is delayed at some i. Whenever ϕ is interestingly heretical at i, i is destined for $\neg\phi$, but destiny is delayed. Our whole puzzle is rooted in the possibility of this kind of informational delayed destiny. We (Jane) are in states destined for (5), but we don't yet accept (5)—indeed we accept something equivalent to its negation. Delayability is the unusual property that $\neg \otimes \phi$ and its kin have—the property setting it apart both from nonmodal talk and modal talk of the more ordinary variety.

Here is a way to see things. For any state i, there are the sentences ϕ accepts (i's *acceptance set*) and the sentences i is destined for (i's *divine set*). Restricted to nonmodal ϕ, it is usual to assume these sets are the same, i.e., that $i \vDash \phi$ iff $i \mapsto \phi$. A gap between the sets opens once we recognize non-persistent ϕ. For these, it may be that $i \vDash \phi$, though $i \not\mapsto \phi$: we have cases where what i accepts isn't necessarily part of i's destiny. Non-persistence opens the door to heretical content, and thereby to the possibility of ϕ that are destined but delayed, which corresponds with the reverse case: cases where $i \nvDash \phi$, but $i \mapsto \phi$. If non-persistent things are nontrivially knowable, we are bound allow for the possibility of interestingly heretical knowledge—and for the paradox the paper started with. But the paradox dissolves if factivity gives out in the case, and I have suggested that is indeed what happens. For we and Jane, (5) is delayed, but that's squareable with our also knowing that it isn't delayed for Carl, since we know Carl knows more.

Knowledge isn't factive, but it's close. What looks true, for all we have seen, is that $i \vDash K\phi$ implies $i \vDash \phi$ at least when ϕ isn't delayed at i. Thus knowledge seems at least to be what we could call *parafactive*—where that is the idea that if $i \vDash K\phi$, then $i \vDash \phi$ or $i \mapsto \phi$.[22]

[22] Note that in the context of the assumption that for nonmodal ϕ, $i \mapsto \phi$ implies $i \vDash \phi$, the claim that K is parafactive would imply Nonmodal Factivity. In the context of the

Whether the parafactivity of knowledge (or something like it) could suffice to capture all we wanted from Factivity is a question to explore elsewhere. The main aim of this section was to bring some formal notions relevant for understanding our puzzle into focus. It seems reasonable to expect that they may be relevant to framing whatever generalization succeeds Factivity. Meanwhile, the claim of the paper was just that the search for a replacement generalization is motivated. Factivity leads to paradox.

References

David Beaver. *Presupposition and Assertion in Dynamic Semantics.* Studies in Logic, Language and Information. CSLI Publications, 2001.

Bob Beddor and Simon Goldstein. Believing epistemic contradictions. *Review of Symbolic Logic*, 11(1):87–114, 2018.

Bob Beddor and Simon Goldstein. Mighty knowledge. *The Journal of Philosophy*, 118(5):229–269, 2021.

Justin Bledin. Logic informed. *Mind*, 130(490):277–316, 2014.

J. Cantwell. Changing the modal context. *Theoria*, 74(4):331–351, 2008.

Josh Dever and Henry Ian Schiller. This paper might change your mind. *Noûs*, 55 (4):863–890, 2021.

Fred Dretske. The pragmatic dimension of knowledge. *Philosophical Studies*, 40: 363–78, 1981.

Kai von Fintel and Anthony Gillies. Must... stay... strong! *Natural Language Semantics*, 18(4):351–383, 2010.

Kai von Fintel and Anthony Gillies. Still going strong. *Natural Language Semantics*, 29:91–113, 2021.

Anthony Gillies. Epistemic conditionals and conditional epistemics. *Noûs*, 38(4): 585–616, 2004.

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

Anthony Gillies. Updating data semantics. *Mind*, 129(513):1–41, 2018.

Simon Goldstein. Generalized update semantics. *Mind*, 128(511):795–835, 2019.

assumption that $i \mapsto \Diamond\phi$ implies $i \vDash \Diamond\phi$, it would also imply the validity of the inference from $K \Diamond \phi$ to $\Diamond\phi$ for nonmodal ϕ.

Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and modality. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–213. Blackwell, 1996.

Peter Hawke and Shane Steinert-Threlkeld. Semantic expressivism for epistemic modals. *Linguistics and Philosophy*, 44(2):475–511, 2021.

John Hawthorne, Daniel Rothschild, and Levi Spectre. Belief is weak. *Philosophical Studies*, 173(5):1393–1404, 2016.

Wesley H Holliday and Matthew Mandelkern. The orthologic of epistemic modals. *arXiv preprint arXiv:2203.02872*, 2022.

David Kaplan. Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford University Press, Oxford, 1977/1989.

Nathan Klinedinst and Daniel Rothschild. Connectives without truth tables. *Natural Language Semantics*, forthcoming, 2012.

Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, CVII (3):115–43, March 2010.

David Lewis. Elusive knowledge. *Australasian Journal of Philosophy*, 74(4):549–567, 1996.

Matthew Mandelkern. Bounded modality. *Philosophical Review*, 128(1):1–61, 2019.

Matthew Mandelkern. Dynamic non-classicality. *Australasian Journal of Philosophy*, 98(2):382–392, 2020.

Sarah Moss. Epistemology formalized. *Philosophical Review*, 122(1):1–43, 2013.

Sarah Moss. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics*, 8(5):1–81, 2015.

Sarah Moss. *Probabilistic Knowledge*. Oxford University Press, 2018.

Dilip Ninan. Semantics and the objects of assertion. *Linguistics and Philosophy*, 33(5):355–380, 2010.

Dilip Ninan. Naming and epistemic necessity. *Noûs*, 55(2):334–362, 2021.

Daniel Rothschild. What it takes to believe. *Philosophical Studies*, 177(5):1345–1362, 2020.

Daniel Rothschild and Seth Yalcin. On the dynamics of conversation. *Nous*, 51(1): 24–48, 2017.

Paolo Santorio. Path semantics for indicative conditionals. *Mind*, 131(521):59–98, 2022.

Robert Stalnaker. Pragmatic presuppositions. In Milton K. Munitz and Peter Unger, editors, *Semantics and Philosophy*. New York University Press, 1974.

Robert Stalnaker. *Context*. Oxford University Press, Oxford, 2014.

William B Starr. A uniform theory of conditionals. *Journal of Philosophical Logic*, 43(6):1019–1064, 2014.

Matthew Stone. The reference argument of epistemic must. In *Proceedings of the International Workshop on Computational Semantics*, pages 181–90. 1994.

Peter Unger. *Ignorance: a case for skepticism*. Oxford University Press, Oxford, 1975.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25 (3):221–61, 1996.

Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122(1):45–92, 2013.

Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

Seth Yalcin. Nonfactualism about epistemic modality. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 295–332. Oxford University Press, Oxford, 2011.

Seth Yalcin. Context probabilism. In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit Weidman Sassoon, Katrin Schulz, and Matthijs Westera, editors, *Proceedings of the Eighteenth Amsterdam Colloquium*, volume 7218 of *Lecture Notes in Computer Science*, pages 12–21. Springer, 2012a.

Seth Yalcin. A counterexample to Modus Tollens. *Journal of Philosophical Logic*, 41(6):1001–1024, 2012b.

Seth Yalcin. Epistemic modality *de re*. *Ergo*, 2(19):475–527, 2015.

Seth Yalcin. Context-sensitivity by way of parameters. In Ernest Lepore and Ernest Sosa, editors, *Oxford Studies in the Philosophy of Language*, volume 2, pages 267–304. Oxford University Press, 2022.