Knowledge-first Evidentialism about Rationality*

Julien Dutant

Penultimate draft
Forthcoming in Dorsch & Dutant (eds.), *The New Evil Demon*, Oxford University Press.

Abstract

Knowledge-first evidentialism combines the view that it is rational to believe what is supported by one's evidence with the view that one's evidence is what one knows. While there is much to be said for the view, it is widely perceived to fail in the face of cases of reasonable error—particularly extreme ones like new Evil Demon scenarios (Wedgwood, 2002). One reply has been to say that even in such cases what one knows supports the target rational belief (Lord, 201x, this volume). I spell out two versions of the strategy. The direct one uses what one knows as the input to principles of rationality such as conditionalization, dominance avoidance, etc. I argue that it fails in hybrid cases that are Good with respect to one belief and Bad with respect to another. The indirect strategy uses what one knows to determine a body of *supported* propositions that is in turn the input to principles of rationality. I sketch a simple formal implementation of the indirect strategy and show that it avoids the difficulty. I conclude that the indirect strategy offers the most promising way for knowledge-first evidentialists to deal with the New Evil Demon problem.

*I am grateful to Davide Fassio and Errol Lord for discussions of the ideas of this paper and to Charity Anderson, Brian Weatherson, Jo Wolff, Eliot Michaelson, Amanda Green, Tim Williamson and, again, Davide Fassio, for detailed comments on earlier drafts. Thanks also to Matt Benton, Maria Lasonen–Arnio, Juan Comesaña, Bill Dunaway, Dmitri Gallow, Jeremy Goodman, John Hawthorne, Brian Hedden, Emil Moeller, Igal Kvart, Clayton Littlejohn, Nico Silins, Martin Smith, Wolfgang Schwarz, Christian Straßer, Jacques Vollet, and audiences at Geneva, Oxford and Bochum for helpful comments and suggestions.

Much can be said for the view that being rational is, roughly, doing one's best in view of what one knows. It fits some patterns of ordinary appraisal, cohere with attractive views of evidence and reasons, provides systematic unity in epistemology, and solves a problem in the interpretation of decision theory. The view has few champions, however. For it appears to founder in the face of some cases of *excusable error*—the most dramatic of which involve a Cartesian demon. In such cases one is misled in a way that impugns one's knowledge but seems to leave one's rationality untouched. Granted, some argue that in cases of excusable error one's beliefs and intentions are not *not fully justified* (Williamson, 2010, p. 359), *not appropriate* (John Hawthorne and Stanley, 2008), or that one *ought not* to have them. But it is much harder to claim that these beliefs and intentions are *not rational* (Wedgwood, 2002).[1] Since one may lose knowledge, so to speak, and do so quite drastically, without losing one's rationality, being rational does not seem to be a matter of doing—believing, intending, and so on—what's best in view of what one knows.

As a result the view—call it *knowledge-first evidentialism about rationality*—has barely any defenders. Worse, it is hardly even considered. Rather, dominant accounts of rationality depart from it in either of two ways. Some maintain the evidentalist idea that being rational is doing one's best *in view of* something, but something internal: *what one believes* or *one's experiences*, internally construed. Others replace the evidentialist idea with the procedural or dispositionalist idea that rationality is a matter of having, manifesting, exercising or doing as if one had good dispositions. Some dispositionalists views are externalist and framed in terms of knowledge, such as the view on which being rational is exhibiting a disposition to know. But they reject the idea that what's rational is a matter of what one actually knows. All these alternatives are partly motivated by the need to accomodate cases of excusable error.

There is, however, a loophole in the argument against knowledge-first evidentialism (Williamson, 2000, pp. 199–200; Hornsby, 2008; Lord, 201x). The view takes what is rational to be determined by two factors: what counts as best, which may depend on one's preferences or interests, and what one knows. Keeping the first factor fixed, *there cannot be a difference in rationality without a difference in knowledge*. But that does not entail that *there is no difference in knowledge without a difference in rationality*.

[1] Sutton (2007, p. 29) is the only "knowledge-first" epistemologist to straightforwardly endorse the view that no false belief is rational. Williamson (forth) endorses a guarded version of the view that distinguishes different "grades" of rational and allows that in the primary sense, no false belief is rational. Unger (1975) argues that a belief is rational only if it is adequately based on something known. Unless adequate basing requires entailement, that allows for false rational belief. Unger ultimately argues that no belief (true or false) is rational because nothing is known; the argument does not target false rational belief specifically. Other "knowledge-first" epistemologists avoid rationality-talk altogether, or talk of rationality without saying much on how it connects with knowledge (Williamson, 2000), or explicitly distinguish rationality and justification.

It is open that different bodies of knowledge rationalize the same thing. In excusable error cases, one still knows many things—for instance, things about what seems to be the case. These things may rationalize the same beliefs and intentions that one's knowledge rationalize in less deceptive circumstances.

This paper explores the loophole. A *direct exploitation* of the loophole results in a knowledge-based decision theory (John Hawthorne and Stanley, 2008, Weatherson, 2012). I argue that it fails in the face of what I call asymmetric excusable error cases. However, there is also an *indirect exploitation* of the loophole. On that view, what is known determines *what is supported*, which (in conjunction with standards for "best") in turns determines what is rational. Provided that what is supported is the same across excusable error cases and their non-deceptive counterparts, the view explains how one can be rational in excusable error cases while maintaining that being rational is a matter of what one knows. Note that, unlike dispositional knowledge-first accounts, the view makes rationality a matter of what one actually knows, not a matter of what one is disposed to know.

The upshot is that knowledge-first evidentialism, formulated indirectly, is a serious contender for an account of rationality. Since the view is hardly even discussed, I set the bar relatively low. I will not argue that the view is superior to alternative accounts. I will not present a fully worked-out version of the view—in particular, I will treat the support relation as a primitive. But I will present a simple formal implementation of the view. While I do not want to tie the view to the specifics of that implementation, it will, I hope, give a more concrete feel for how the view delivers some of its goods.

1 The project, the problem and the loophole

Knowledge-first evidentialism about rationality is the idea that being rational is roughly *doing one's best in view of what one knows*. Let me clarify that a bit. First, I'm using "rational" in an ordinary sense, not in a technical, semi-stipulated sense. In that sense being rational is, roughly, doing the reasonable, intelligent, smart, sensible thing; being irrational doing the senseless, stupid, idiotic, crazy thing (Parfit, 2011, p. 33). Second, "doing" stands here for anything that may be called rational or irrational: beliefs, intentions, actions, emotions, and so on. However I will only be concerned with beliefs and intentions. Third, I am primarily concerned with rationality *ex ante.* Suppose the evidence you have points to your arch-enemy's guilt, but you believe they are guilty merely out of spite. *What it is rational for you to believe* is that they are guilty, but you believe it *irrationally*. The former is a matter of *what is the rational thing to do* in your situation; we call that rationality *ex ante.* The latter is a matter of whether *you are being rational in doing* what you do; we call that rationality *ex post*. That is an analogue of the epistemologists' distinction between "propositional" and "doxastic" justification. Fourth, what "best" exactly amounts to does not matter much for my purposes. Rationality may merely require satisficing, that is, doing what's *sufficiently good* in view of what one knows. When one's knowledge leaves the value of one's options uncertain, rationality may more precisely require doing what's *expectably best* in view of what one knows. There may be several notions of good (moral, prudential, and so on), each yielding a different rationality standard. What's good on some of these notions (*e.g.* prudential) may depend on a person's fundamental values. If so, what is rational will not only depend on what one knows, but also on one's fundamental values. The qualification will be left implicit. Finally, I leave open whether the view should be framed in terms of *what one knows* rather than in some other knowledge-based notion such as *what one is in a position to know* (Williamson, 2000, p. 95) or *what one knows by observation alone*.

Why would one hold such a view? Here are five reasons. They are not intended to be conclusive, and each of them would require much more discussion than we can afford here. But I hope they show that the view is worth exploring.

First, the view fits well some patterns of ordinary appraisal (John Hawthorne and Stanley, 2008, p. 571). When we call a person's actions or opinions rational or irrational, we often back up our judgement with claims about what they knew or didn't know, or about some (arguably) knowledge-entailing state such as seeing. Even when we back it up with a simple factual claim, it is arguably assumed that the person knew the fact in question.[2]

Second, many endorse the *evidentialist* idea that being rational is doing what's best in view of one's *evidence*. But it's tempting to say that one's evidence is what one knows (Williamson, 2000), or at least what one knows by observation. From this the view follows.[3]

Third, many assume that being rational is doing what's best in view of *the reasons one has*. But it's tempting to say one's reasons are just what one knows (Unger, 1975; Magidor and J. Hawthorne, forthcoming). From this the view follows as well.

Fourth, there are considerations of systematic unity. During what we may call the

[2]Even when we back it up with a claim about the person's non-factive mental states, one could make the case that the assumption that the state is known is crucial. It's not clear, for instance, that the claim "It was irrational for Amir not to pick the strawberry ice-cream, since that's the flavour he really wanted" would stand if Amir didn't know, or was not even in a position to know, that he wanted strawberry ice-cream.

[3]More generally, any account of evidence as some subset of what one knows that goes beyond the "internal", combined with the evidentialist principle linking evidence and rationality, yields the kind of view I am concerned with.

Chisholmian era, the dominant project in epistemology was to explain knowledge in terms of rationality. One representative view, for instance, holds that one knows p just if, were one to believe all and only truths, what makes one's belief p reasonable would still make it reasonable.[4] If, like me, you do not think that it has been successful, you are left with three options. One is to settle for a two-headed epistemology, where knowledge and rationality belong to distinct set of notions. Another is to account for both in terms of a third notion—the main contender is currently *intellectual virtue*. A third, inspired by Williamson (2000), is to account for rationality in terms of knowledge. The present view aims for the third.

Fifth, the view fits well with how decision theory is applied (Weatherson, 2012, pp. 77-82). Decision theory is standardly presented as a theory of what it is rational to do. To apply it to real-life of imaginary cases we typically frame decisions problems in terms of what an agent *knows*. For instance:

> Suppose [...] that it is known of an urn that it contains either two white balls, two black balls, or a white ball and a black ball. (Savage, 1972, p. 65)
>
> The agent is loath to eat meat on Fridays; is sure that today is Thursday or Friday, but doesn't know which [...]. (Jeffrey, 1983, p. 53)
>
> Let us assume that the physician knows the probability of observing such spots given that the patient has tuberculosis and that she also knows the incidence of TB and the incidence of lung spots. (Resnik, 1987, p. 53)

Similarly, updating is routinely described in terms of what an agent has *learned*, and learning is naturally construed as coming to know. The practice makes straightforward sense on the knowledge-first evidentialist view. We discuss it further in the next section.

Unfortunately, knowledge-first evidentialist seems to clash with obvious facts about rationality (compare Brown 2008, sec. 5). In the *Good case*, Sarah remembers, and thereby knows, that her car is parked in front of the office. At the end of her workday she decides to walk towards that spot. In the *Bad case*, everything seems to Sarah as it does in the Good case but her car has been stolen. She does not know that her car is parked outside—it is not—, but she takes herself to do so. She decides to walk towards the same spot. In both cases, it's rational for her to believe that her car is parked on that spot and it's rational for her to intend to go there. Yet in the second case, she lacks a

[4] That is the gist of Pollock's (1986) account. The idea is that in Gettier-style cases, acquiring more true beliefs about one's situation would change the basis on which on believes p. Foley (2012) defends a similar view.

crucial piece of knowledge. So, it seems, what is rational doesn't depend on what one knows.

The problem is an instance of the New Evil Demon problem (Lehrer and Cohen, 1983, Cohen, 1984, Wedgwood, 2002). In its extreme version, we compare a Good case like the one above with one in which Sarah is the victim of a Cartesian Evil Demon. In the extreme Bad case, her knowledge is drastically impoverished, yet she seems as rational in her beliefs and intentions as she is in the Good case.

The problem motivates alternative acounts of rationality that reject either the knowledge-first view of evidence or evidentialism about rationality. The alternative accounts face difficulties of their own and the debates are ongoing. I do not aim to rehearse these here.[5] Rather, I intend show that contrary to what most epistemologists assume the New Evil Demon problem does not force us to adopt one of them. If that is right, the alternative accounts cannot be motivated by the New Evil Demon problem alone.

The main idea is this. There is an obvious loophole in the argument against knowledge-first evidentialism (Williamson, 2000, pp. 199–200; Hornsby, 2008; Lord, 201x). What one knows differs across the Good and Bad cases. But two different bodies of knowledge may rationalize the *same* thing. In the ordinary Bad case, Sarah still knows that *she parked the car on that spot*, for instance. In the extreme Bad case, she may still know that *it seems to her that she did* (Williamson, 2000, p. 199) and that *she does not know that the car is not on the spot*, for instance. These bodies of knowledge give her reason to believe that the car is there and to intend to walk towards it. So we could both maintain that what it is rational to do is a function of what one knows *and* that the counterpart states in Good/Bad cases pair are equally rational.

The loophole is worth exploring. It would uphold a simple account of rationality that combines the virtues of evidentialism and externalism. Evidentialism is more readily combined than dispositionalism with our best models of rationality—those of decision theory and formal epistemology. More controversially, putting Bad cases aside, externalist notions of evidence have an easier time explaining what it is rational for us to do. Internalist notions tend to recommend excessive confidence about the inner and have a harder time explaining the rationality of our beliefs about the outer or the past.[6]

However, it turns out that there are two ways of exploiting the loophole. Both take one's evidence and reasons to be what one knows. The *direct strategy* feeds that into a decision-theoretic representation of the choices one faces, and take the resulting recommendations to be rational. As I will argue in next section, that strategy fails. The

[5] See my ***introduction to this volume***.

[6] See the speckled hen problem for the first (BonJour and Sosa, 2003) and the problem of stored belief for the latter (Williamson, 2007).

indirect strategy says that what one knows fixes, but is not identical with, the input of decision theory. As we will see, that strategy shows more promise.[7]

Before we move on, an important note. As Lord (201x) points out, the loophole raises a worry with rationality *ex post*. On a standard view, you act rationally *ex post* (doing something rationally) just if you do what is rational *ex ante* (what it is rational for you to do), and you do so for the right reasons, namely, you do so on the basis of reasons or evidence in virtue of which that thing is rational *ex ante*. Now on the view just sketched, what makes it rational for Sarah to intend as she does is that she parked the car there or that it seems to her that she did. However her intentions are formed on the basis of the false belief that the car is there. So, it seems, the view has to say that Sarah is not intending rationally in the Bad Case. So the problem recurrs with rationality *ex post*. To this Lord replies that, as we naturally imagine the case, Sarah is sensitive enough to the relevant known facts: that she remembered having parked the car there, that it seems to her that the car is there. If, for instance, she did not remember that or it did not seem to her that way, she would stop believing that the car is there. Because she is so sensitive, Lord claims, Sarah's intention is *also* based on those facts. And that, he argues, is all that we need for her to intention to be *ex post* rational. The reply strikes me as being on the right track. I will endorse it here and leave rationality *ex post* aside.

2 The direct strategy

The direct strategy evaluates both belief and intention in view of what one knows. With intention, the result is *knowledge-based decision theory* (Weatherson, 2012, pp. 77–82).[8] Put roughly, the idea is this: if you draw a decision table that reflects what a

[7] Lord (201x) does not distinguish the two strategy. His proposal may be spelt out either way.

[8] The view has also been suggested by (John Hawthorne and Stanley, 2008, pp. 577–80). Showing that they endorse it requires a bit of exegetical work. They explicitly present their view as an alternative to dominant theories of *rationality* (571, 577 and other places). Yet when they state their view they do not use "rational" and cognates but instead "ought", "should", "appropriate" and "excusable". While they call the norm "treat *p* as a reason for acting only if you know *p*" a "norm of practical rationality" (577), it is unclear whether failing to comply with the norm entails being in any way irrational. This depends on whether "excusable" violations are always rational; but they stay silent on this — see 573, 577, 582, 586. Whichever way they go on this, they must endorse the idea that practical rationality is a function of the relevant types of appropriateness and excusability. If not, it is hard to see how their proposal is a theory of rationality. Since they think that appropriateness and excusability in the relevant sense are a function of what one knows, they endorse the knowledge view. More specifically, they take themselves to "gesture at" a type of decision theory in which knowledge delivers epistemic probabilities (580). It is very natural to assume that that knowledge is also meant to deliver decision tables, as in Weatherson's more explicit proposal.
The same page (580) misleads some readers into thinking that Hawthorne and Stanley distance themselves from the proposed decision theory. They write that while "[they] are by no means opposed to a perspective in which claims of practical rationality [. . .] are grounded in a decision theory of the sort [they] gestured at",

person knows, then that table shows you what it is rational for her to intend. To reflect what a person knows, the information on the table must include all the relevant facts that the person knows, and nothing more. In particular, the states of the world it treats as possible should be exactly those states that are compatible with what one knows. There are some details to iron out but the rough picture is enough for now.[9] As we noted above, a major advantage of the view is to make sense of the way in which decision theorists rule states in or out of a decision table. [10]

Unfortunately, the direct strategy does not work. That can be shown by using *Mixed cases*, cases in which one is in a Good case with respect to one proposition but a Bad one with another. Suppose you need to grab a bottle of water to go on a bike trip. In the *Good case*, you know that:

b: there's a bottle of water in the basement.

g: there's bottle of water in the garage.

With details suitably filled in, I claim:

(1) In the Good case, the table below reflects your knowledge:

	$b\&g$
Go to the basement	1
Go to the garage	1

That is, you know that $b\&g$, so all the columns in your decision table should be ones in which $b\&g$ obtains. You have also some background knowledge which is not made explicit here but will be constant across our cases. Several $b\&g$ states are compatible with what you know (for instance, some in which it is snowing next winter, some in which it is not) but the differences between them do not matter to your decision. Finally, you know that your options are to go to the basement or the garage (no other or more fine-grained options are relevant) and you know that if you do either you will get a bottle, with a certain value that we represent by the number 1. The table encodes all these items of knowledge, and no more.

"the need to integrate such a theory with reasons for action is still vital". The point of this, however, is not to distance themselves from the proposal; rather, it is to stress that the proposal only covers rationality *ex ante*. That is made clear in the next sentence: "For one thing, there are cases where one does what one ought to do but for the wrong reasons, and this phenomenon needs explanation". The idea is not that knowledge-based decision theory is false; it is that it fails to cover rationality *ex post*.

[9] One thorny detail is that the information under the table is closed under entailment, while a person's knowledge is not. So distant, non-obvious logical consequences of what appears on the table are not supposed to be representative features of the model.

[10] The suggestion is that a table models rationality *if* it reflects an agent's knowledge. It is not that it models rationality *only* if it reflects an agent's knowledge. In game contexts, for instance, it may be useful to include in a table states that are excluded by a subject's knowledge but not by her opponent's.

In the *Mixed case*, exceptional circumstances have led your partner to take away the bottle in the garage. Everything is otherwise as in the Good case. So, I claim:

(2) In the Mixed case, the following table reflects your knowledge:

	$b\&g$	$b\&\neg g$
Go to the basement	1	1
Go to the garage	1	0

That is, you still know that b, but fail to know that g (because g is false). Everything is otherwise as before.

But, by fairly minimal decision-theoretic principles:

(3) In table (1), going to the garage is choice-worthy.

(4) In table (2), going to the garage is not choice-worthy.

These are two limited applications of dominance reasoning. Dominance reasoning is problematic when states causally or evidentially depend on actions and with iterated elimination of choices in games. But nothing of the sort arises here. Now given knowledge-based decision theory, (1)–(4) entail:

(5) it is rational to (intend to) go the garage in the Good case but not in the Mixed case.

Which contradicts the New Evil Demon intuition.

Can defenders of the direct strategy block the argument?

Attempt 1: scepticism. By rejecting knowledge of the external world, one can reject (1) and restore symmetry between the cases. *Reply*. Scepticism blocks the argument but at heavy cost: its sheer implausibility as well as a steep challenge in explaining how anything can be rational (Unger, 1975, chap. 4). The kind of knowledge-first account of rationality I'm interested in here is not sceptical.

Attempt 2: no knowledge of b *in Mixed*. This attempt denies that you know b in the Mixed case. It says that you are in a kind of fake barn situation where your mistake about g also impacts your ability to know b. We should thus reject (2) since the proper table for the Mixed case would be:

	$b\&g$	$b\&\neg g$	$\neg b\&g$	$\neg b\&\neg g$
Go to the basement	1	1	0	0
Go to the garage	1	0	1	0

Since going to the basement does not weakly dominate on that table we need not deem it irrational. *Reply*. That is untenable for some variants of the case. Suppose you are in the basement, looking at the bottle. (The basement is much messier however, so getting that bottle would not cost less time and energy than getting one in the garage.) Whatever happens in the garage does not affect what you know about the basement.

Attempt 3: divide and conquer. This attempt builds on the previous ones but divides variants of the pair into two groups. The *symmetric* ones are those in which, in the Good case, you are equally well positioned with respect to b and g. The *asymmetric* ones are those in which, in the Good case, you are better positioned towards b. A case where you *see* the bottle in the basement but merely *remember* the one in the garage would be an asymmetric one. For symmetric pairs, the attempt uses the "fake barn" verdict of the previous attempt. For asymmetric ones, it claims that, since in the Good case, your epistemic position or evidence for b is better than that for g, it is (if ever so slightly) irrational to go to the garage even in the Good case.

Reply. On the symmetry side it is unclear why the fact that one's position with respect to b and g is *equally good* means that one cannot "lose" knowledge of one without losing knowledge of the other. That would have more plausibility if one's knowledge of both came from the same source; but they may be equally good without coming from the same source. On the asymmetry side, a first issue is to implement the view. Table (1) has to be rejected; but replaced with what? One option is to say that you *fail* to know g in an asymmetric Good case. That invites in rampant scepticism—every small difference in "epistemic position" will be said to forbid indifference and that will require that one fails to know. One may try to contain it by adopting a pragmatic encroachment view on which knowledge of the lesser-positioned proposition is only lost when the asymmetry is relevant to choices one actually faces. Still, that is much more skepticism than pragmatic encroachers typically acknowledge, since cases of choice among quasi-identical options are much more frequent than "high-stakes" choices. Alternatively, one may try to hold onto the idea that both b and g are known in asymmetric Good cases but implement the idea that going to the basement should be preferred in some other way. It's unclear how to do that without rejecting knowledge-based theory. A worse problem is that the asymmetry in the Mixed case need not match that of the Good case. It may go in the opposite direction. To illustrate: in the Good case, you are looking at the bottle in the garage, and in the Mixed case you are looking at the bottle in the garage but its water has been replaced by petrol. In the Good case your position is better with respect to g (seeing) than to b (remembering) but in the Bad case it is worse with respect to g (deceptive appearances) than to b (remembering). The attempt would still say that is rational to go to the garage in the Good case but not in the Mixed one.

Attempt 4. no knowledge of equal goodness in Mixed. This attempt says that in the Mixed case, you not only lack knowledge that the garage is *at least as good* as the basement, but you also lack knowledge that it is *at most as good*. That is, for all you know in the Mixed case, there could be something *better* in the garage—a fancy energy drink, say. The attempts rejects (2) and replaces it with a table such as:

	$b\&g$	$b\&\neg g\&\neg f$	$b\&\neg g\&f$
Go to the basement	1	1	1
Go to the garage	1	0	2

So going to the basement does not dominate in the Mixed case. *Reply*. That is untenable in some variants. Even in the Mixed case you retain a lot of knowledge of the garage: you still know that there are no elephants in it, for instance. You may just as well retain knowledge that it contains nothing better than a bottle of water.

Attempt 5. Zero probability. Since actuality is compatible with what one knows, a knowledge-reflecting table for the Mixed case must have a $b\& \sim g$ state. But, this attempt suggests, we can assign it a probability zero. If so, it is irrelevant to fix what is rational, and we can restore symmetry. *Reply*. First, one has to justify the zero probability assignment. It's hard to see how it could derive from reasonable priors conditionalized on what one knows in the Mixed case. Second, the attempt faces a dilemma. Either zero-probability states are distinguished from impossibilities or they are not. If they are, as when we say that a win in an infinite lottery is possible even though its probability is zero, dominance reasoning seems sound: you should prefer a bet that pays a lot if 8 comes up and nothing otherwise to one that pays nothing whatever happens.[11] So the view does not avoid the problem. If they are not, the view is in effect treating zero-probability states are excluded from the table even though they are not known not to obtain. Then it is not an instance of knowledge-first decision theory.

Attempt 6. Reject the decision-theoretic claim. The last attempt is to object to the decision-theoretic claims (3) and (4). *Reply.* While dominance is not wholly unproblematic it is hard to motivate an exception here.

So the direct strategy is incompatible with the New Evil Demon intuition. Since

[11]There is of course a tension between saying that you should prefer the dominant option when its dominance only arises from zero-probability states and the idea that you should not sacrifice anything to switch between two options of equal expected value. Still one can at least remove the apparent contradiction by saying that preferring a over b does not require being willing to give up anything for getting one rather than the other.

knowledge-first evidentialism was widely thought to be incompatible with the intuition, the result will not come as a surprise. But the result is worth establishing carefully. For the result does not affect all versions of knowledge-first evidentialism.

3 The indirect strategy

The *indirect* strategy exploits the loophole too but proceeds in two steps. First, what one knows supports a certain picture of how things are. Second, the supported picture fixes what it is rational to believe, to intend and so on. Thus what it is rational to do is still a function of what one knows. But unlike the direct strategy, what one knows does not directly serve as the input of decision theory.

The strategy deals with the New Evil Demon problem as follows. In a Bad case, what is supported is *more* than what one knows: what is supported is (a counterpart of) exactly what one knows in the corresponding Good case.[12] In the Mixed case of the previous section, for instance, even though you do not know that there is a bottle in the garage, what you know supports the hypothesis that there is one. So $b\&g$ is supported, just as it is in the Good case where it is known. So what is supported is the same across both cases. More specifically, the decision table (1) reflects *what is supported* in both cases, even though it only reflects what one knows in the first. Since what it is rational to do is in turn determined by what is supported, what it is rational to do is the same across both cases. So the New Evil Demon intuition is preserved.

Optionally, one may supplement the strategy with the claim that what it is rational to believe is simply what is supported. With that addition, the indirect strategy can be rephrased thus: first, what one knows fixes what it is rational to believe. Second, what it is rational to believe fixes what it is rational to intend, to feel and so on.

To make good on the view we should spell out, or at least argue that there is, a support relation that delivers the desired results. That project is beyond the scope of this paper. Instead I will do three things. First, I will lay out plausible constraints on such a relation. These will give a flavour of what the support relation may be like.

[12] Content externalism makes it impossible to say that in all Good / Bad cases pairs the *very same* beliefs or intentions are rational in both cases. Suppose, for instance, that I have never encountered a certain cat (Felix) before. In the Good case I see Felix on the mat and it is rational for me to believe that Felix is on the mat (though I would not myself put it in those terms; I would say "*that cat* is on the mat"). In the Bad case I have an hallucination that happens to match my experience in the Good case. In the Bad case it is rational for me to have a belief that I would express by "*that cat* is on the mat". But that belief is not a belief that Felix on the mat. Whatever its exact content is, it is sufficient for our purposes that we count that belief as a *counterpart* of my belief that Felix is on the mat in a Good case. Whether we can make systematic sense of the counterpart relation in question is open. If we cannot then it is hard to spell out the New Evil Demon intuition in rigorous terms and that may undermine the intuition itself. That would be a problem for any view that tries to accomodate it, not simply the present one. Thanks to Timothy Williamson here.

Second, I will provide a toy model of a support relation that satisfies these constraints. The model is not supposed to be the last word on the relation but a good first sketch. Third, I address four objections to the view. These should be enough to establish that the indirect strategy is worth exploring.

First, some tentative constraints. Write Kp for "p is known", Sp for "p is supported". I will use "might" as the dual of "know" and write Mp for $\sim K \sim p$, that is, "might p" means that it is not the case that p is known to be false. Note that when I say, "it is not the case that p is known to be false", I do not imply that p *is* false: if p is true, then *a fortiori* it is not known to be false ($p \to Mp$ is the dual of the factivity axiom $Kp \to p$). Here are some desirable principles for a support relation:

$Kp \to Sp$ What is known is supported.

$Sp \to MKp$ What is supported might be known.

Put together these two principles put an upper and lower bound on the supported: it includes *at least* what is known, and *at most* what might be known. The second principle says that if you know that you do not know p, then p is not supported. It is consistent with the idea that what is rationally believed in a Bad case is supported, since a characteristic feature of what is rationally believed in a Bad case is that one cannot know that one fails to know it. By contrast, assuming that in a typical lottery scenario you know that you do not know that your ticket will lose, then in such a scenario the hypothesis that it will lose is not supported. Hence if we equate justified belief with the supported we reject views on which lottery beliefs are justified.[13]

The view rejects both converse principles $Sp \to Kp$ and $MKp \to Sp$. Against the first, the view means to allow that in Bad cases, some things that we do not know are supported. Against the second, there are cases where p is a proposition you cannot even conceive; arguably, it is not the case that you know that you do not know p, and for that matter it is not the case that you know that you do not know $\neg p$ either. But we do not want to say that p is supported, much less that both p and $\neg p$ are.

Consistency The supported is consistent.

Closure The supported is closed under entailment.

Cut If one's total knowledge A supports all and only the propositions in B and if one's total knowledge was B it would support all and only the propositions in C, then $B = C$.

[13] See Smith (2016) for a defense of the idea that lottery beliefs are not justified.

These principles are optional. *Consistency* and *Closure* face trouble in Preface-like cases, in particular. However, they make the support relation formally simpler, so they are good to try at a first pass. The *Cut* principle implies that it does not matter whether the supported is merely supported or known. While appealing, note that we may reject it while maintaining knowledge-first evidentialism. Nor do I see a straightforward argument from the New Evil Demon intuition to that principle.

Goodness In a Perfectly Good case, $Kp \leftrightarrow Sp$ for all p.

That principle characterizes Good cases. Good cases are not misleading: one's knowledge in such cases doesn't support anything beyond what is known. Note that we need not assume that the garden-variety stories we usually call "Good cases" are *Perfectly* Good cases. It may be any ordinary human situation involves some degree of mismatch between what is known and supported. Garden-variety Good cases are only meant to be Good on some salient propositions.

NED constraint The supported in a Bad case is (a counterpart of) the supported in a corresponding Good case.

The constraint is a target and a guide to when the relation must hold. For an illustration, consider applying it to *Defeat cases*. In case A, at stage 1 you mistakenly think you perceive a red ball; at stage 2 you learn that it was white but illuminated by red lights. In case B, at stage 1 you perceive a red ball; at stage 2 you are mislead into thinking that it was white but illuminated by red lights. With details suitably filled in, $A1$ and $B1$ are a NED pair with $B1$ as the Good case and $A1$ the Bad one, and $A2$ and $B2$ are a NED pair with $A2$ as the Good case and $B2$ as the Bad one. Let p be the proposition that there is a red ball. Assuming that at $B1$, you know p, by the *NED constraint* p is supported at $A1$. Assuming that at $A2$, you know $\neg p$, by the *NED constraint* $\neg p$ is supported at $B2$, so by *consistency*, p is not supported, and by $Kp \rightarrow Sp$, p is not known at $B2$. Thus the indirect strategist is committed to the traditional view that in a case such as B knowledge is destroyed by misleading counterevidence. [14]

Here is a toy model. For simplicity, assume that what is known is closed under entailment, so we can model what is known at w with the set of possible worlds $K(w)$ compatible with all that is known at w. Say that $w \geq w'$ iff $K(w) \subseteq K(w')$: that is, w is *at least as good* as w' if at w one knows everything that one knows at w' and possibly more. Let $top(w)$ be the best worlds among those at least as good as

[14] Lasonen-Aarnio (2010) argues that knowledge defeat is in tension with some *accounts* of knowledge, in terms of safety or sensitivity for instance. It is not in tension with knowledge-first epistemology *per se*, however.

w: $top(w) = \{w' : w' \geq w$ and $w' \geq w''$ for every w'' s.th. $w'' \geq w'\}$. Intuitively, $top(w)$ are the best-case knowledge scenarios one might hope for in the light of what one knows at w. Assuming finitely many worlds and since $w \geq w$, $top(w)$ is not empty.[15] Now we say that p is supported iff it is known at *all* the best-case knowledge scenarios compatible with what one knows. That is, we let $S(w) = \bigcup_{w' \in top(w)} K(w')$ and we say that p is supported at w iff $S(w) \subseteq [p]$, where $[p]$ is the set of worlds at which p holds. These models reflect an *optimistic* theory of support: look at the best-case epistemic scenarios compatible with what you know; if something is known in all of them, then it is supported. We can verify the following:

- **Supervenience**. If $K(w) = K(w')$ then $S(w) = S(w')$.
- $Kp \to Sp$. For each $w' \in top(w)$, $K(w') \subseteq K(w)$, so $S(w) \subseteq K(w)$.
- $Sp \to MKp$. If $w' \in top(w)$ then $K(w') \subseteq S(w)$ and $K(w') \subseteq K(w)$, so since by factivity $w' \in K(w')$, $w' \in K(w)$. So if $S(w) \subseteq [p]$ then for some world $w' \in K(w)$, $K(w') \subseteq [p]$. So if p is supported at w then there is a world w' compatible with $K(w)$ where one knows p.
- The converses are not true: the supported is neither identified with the known nor what might be known. For some models, $Sp \to Kp$ and $Sp \to p$ fail. Take $K(w_1) = \{w_1, w_2\}$ and $K(w_2) = \{w_2\}$. We have $w_1 \notin S(w_1) = \{w_2\}$. For some models, $MKp \to Sp$ fails. Take $K(w_1) = \{w_1, w_2, w_3\}$ and $K(w_2) = \{w_2\}$ and $K(w_3) = \{w_3\}$. We have $S(w_1) = \{w_2, w_3\}$. With $[p] = \{w_2\}$ at w_1 MKp is true but Sp false.
- **Consistency**. $top(w) \subseteq S(w)$. Since $top(w) \neq \varnothing$, $S(w) \neq \varnothing$.
- **Closure**. What is supported at w is all and only the propositions that hold at all worlds in $S(w)$.
- **Cut**. Suppose that what is supported at w coincides with what is known at w^*: $S(w) = K(w^*)$. Since $K(w^*) = S(w) \subseteq K(w)$, $w^* \geq w$. Now let $w' \in top(w)$. Then for any w'' s.th. $w'' \geq w'$, $w' \geq w''$. Moroever, since $w' \in top(w)$, $K(w') \subseteq S(w) = K(w^*)$, so $w' \geq w^*$. Hence $w' \in top(w^*)$. Conversely, let $w' \in top(w^*)$. Then $w' \geq w^* \geq w$ and for any w'' s.th. $w'' \geq w'$, $w' \geq w''$, so $w' \in top(w)$. Hence $top(w) = top(w^*)$ and $S(w^*) = S(w)$.

[15] To cover the infinite case, in which we might have an infinite series of better and better worlds, we may define $top(w)$ as the set of worlds at least as good as w such that any better world is itself bettered: $\{w' : w' \geq w$ and for any $w'' > w'$ there is $w''' > w''\}$.

If what is supported at w was known at a world w^*, it would only support itself at w^*.

- **Perfect goodness**. For some models and world w, $K(w) = S(w)$. Take partition models, for instance: $K(w') = K(w)$ for all $w' \in K(w)$.

What about the NED constraint? First, it is important not to be misled into thinking that cases of strictly better epistemic states abound. Suppose you are looking at an unmarked clock (Williamson, 2011) and you know that its hand is between 2 and 4. It is presumably compatible with what you know that you know more about the hand: for instance, it may be compatible with what you know that you stand a bit nearer to the clock and know that it is between $2:05$ and $3:55$. However, you also know things about *what you do not know* about the clock: for instance, you know that you do not know that it is between $2:30$ and $3:30$. Now arguably, if you were a bit closer you would know more *about the hand* but less *about what you do not know*: for instance, if you were closer you might not be able to know that you do not know that it is between $2:30$ and $3:30$. So while in that situation there are many propositions that you might (for all you know) know even though in fact you do not know them, it is far from obvious that in that situation you might (for all you know) have a strictly greater body of knowledge than the one you actually have. In short, it is far from obvious that the situation is not a Perfectly Good case. By contrast, a typical Bad case is one in which some strictly better epistemic state is left open by what you know. There is no plausible candidate for a proposition that one knows in a Bad case but not in the corresponding Good one.[16] In particular, if you know that you do not know p in the Bad case, then you do so as well in the Good case. So we should expect the optimist model of support to satisfy the NED constraint.

The optimistic theory is one example of a theory of support. Others are available. One avenue is to use *normality* considerations: what is supported is roughly what obtains in the most normal scenarios compatible with what one knows (compare Smith, 2016, Goodman and Salow forthcoming). These may deliver Good case-Bad case parity. It is less clear that they deliver the principle that what is supported might be known—but the principle is not sacrosanct. Another is to identify the supported with what one might know (Rosenkranz, 2017). If you think that one's "internal" state supervenes on what one knows, Bird's (2007) and Ichikawa's (2014) view that what is supported is what some of one's internal duplicates know would also provide a conception of support compatible with knowledge-first evidentialism. The latter two avenues

[16] Modulo worries about content externalism, as pointed above. The toy model is not designed to deal with those.

straightforwardly provide Good case–Bad case parity, but they do not ensure that the supported is Consistent, as we will see shortly.

Let me now consider five objections to the indirect strategy.

Lesser Weight. In the Good case, one knows p. In the Bad case, one knows less, for instance, one merely knows *that someone claimed that* p. The indirect strategy says that p is supported in both cases. It is thus committed to the view that *someone claimed that* p supports p just as well as p itself does. But that seems wrong: on any plausible notion of support, the latter conclusively supports p while the first does not.

Reply. Note first that what supports is a *body* of evidence. What a body of evidence supports depends on what is in it but also on what is *not* in it. Talk of what a *piece* of evidence supports is derivative at best and possibly misleading. Consider for instance the three bodies of evidence: (1) $\{p$, it seems that $p\}$, (2) {it seems that p}, (3) {it seems that p, p is not part of the evidence}. The first fully supports p and the third weakly or not at all. On the present view, the second fully supports p too. Why so? Not because *it seems that* p is itself a strong piece of evidence—it is not, as (3) shows. Rather, because that body of evidence does *not* contain the fact that p itself is not part of the evidence. On the present view, that is why that body supports p more than the third: not because of what it includes, but of what it fails to include. If, by contrast, we merely focus on how strong a *piece* of evidence *it seems that* p is we would overlook the difference between the second and the third.

Now the view acknowledges that compared to (1), the support given by (2) has shortcomings. It is essentially based on ignorance and liable to support false propositions. It is liable to be overturned by a strict increase in one's evidence. However, the view holds that those differences do not affect rationality.

Bad Conclusions. Suppose your background knowledge (p) is this: your colleague Bertrand is in good health and has a high incentive to skip work. In the Good case, he calls in sick with a true and well-evidenced story and you thereby come to know that he is sick. In the corresponding Bad case, Bertrand made the story up. In the Good case, what you know entails and hence supports the belief that Bertrand is sick. But the Bad case, what you know—p and the fact that he reports being sick—intuitively supports the conclusion that he is faking it. So the view must either adopt *ad-hoc* support relations or violate the NED constraint.

Reply. The case does not raise more trouble than *Lesser Weight*. Clearly, if your evidence included the claim that Bertrand reports being sick *and it is not part of your evidence that he is*, then that would support the hypothesis that he is faking it. But a crucial aspect of the situation is that your current evidence does not include the claim that your evidence fails to include that Bertrand is sick. Here as in *Lesser Weight*, I do not see any incoherence in thinking that as a result of this piece of 'ignorance', your evidence supports more, though in a somehow defective manner.

Preface-style worries. The indirect strategy allows the supported to be inconsistent. Suppose you are inquiring into who ate the cookies and interview three equally trustworthy witnesses whose claims are jointly inconsistent: (a) Amir ate a cookie, (b) Bertrand ate a cookie, (c) they did not both ate a cookie. Now suppose we can construct three variants A, B, C in which everything seems the same but in variant A, one knows a, in B, one knows b and in C, one knows c. That is, A and one of B or C are a NED pair for a, and analogously for b and c. By the principle that what is known is supported, a is supported in A, b in B, c in C. By the *NED constraint* and the assumption that we can construct such a trio, each of a, b and c is supported in each case. So the supported is inconsistent.[17]

Reply. Note that the problem arises for any view that grants that such a trio of cases is possible, that it is rational to believe what one knows, and that it is rational to believe in a Bad case (the counterpart of) what it is rational to believe in a corresponding Good case. Such views must allow that it is rational to believe each of a, b, c. Presumably, they deny closure: it is not rational to believe the conjunction of a, b and c. That is the natural outcome of Bird's (2007) and Ichikawa's (2014) view that it is justified—for our purposes, rational—to believe what an internal duplicate of you knows. For if each of A, B, C is an internal duplicate of the others, then each is the internal duplicate of someone who knows a, b and c respectively. But since the conjunction of a, b and c is contradictory, none of them is the duplicate of somebody who knows the conjunction. The same result arises on Rosenkranz's (2017) view that one has justification—for our purposes, it is rational for one—to believe something just if one is not in a position to know that one is not in a position to know it. If in our trio, for each of a, b, c one is not in a position to know that one is not in a position to know it, then it follows that it is rational to believe each of them. However, one is in a position to know that one is not in a position to know all three. So it is not rational to believe their conjunction.

Alternatively, one may deny that such a trio is possible. Either one fails to know any of a, b or c in the cases or the cases are not "Good case-Bad case" correspondents. For the first line, we may argue thus: in A, upon hearing the witnesses claims that b and c, one is not in a position to know that one of b or c is false. However, if one knew a, one would be in a position to know that one of them is false.[18] So one does not know

[17] Thanks to Clayton Littlejohn here.

[18] This step uses single-premise closure for being in a position to know. While the Preface-style cases under consideration put pressure on multi-premise closure, they do not give reason to reject single-premise closure.

a. One's ability to know a is thus defeated by the contrary testimony in favour of b and c. As we have seen, indirect strategists are plausibly committed to defeat anyway. Explaining how knowledge defeat arises would be nice, but not required. It is enough for its defenders to point out that a defeat verdict is independently plausible. For the second line, we may argue thus: if in A, one really does know a, then one is in a position to know that one of b or c is false.[19] By contrast, if in B one really knows b, it is presumably not rational for one to believe that one of b or c is false.[20] But, assuming that is rational to believe what one is in a position to know, it is rational to believe that in A. So if one really knows a in A and one knows b in B, the cases are not Good case–Bad case pairs. Perhaps the difference in knowledge requires some relevant difference in phenomenology; perhaps the cases show that sameness of phenomenology is not sufficient to have a proper Good case–Bad case pairs. The two lines are mutually compatible of course: it is enough to argue that any putative trio will fall on one or the other type.

Our toy model guarantees Consistency and supports the denial approach. Consider first a scenario in which for each of a, b, c, one fails to know it but also fails to know that one fails to know it. One knows, however, that one does not know all three. What are the best-case knowledge scenarios compatible with what one knows? There are two main possibilities. Either one knows that one fails to know any conjunction of *two*, or one does not. If the former, the best worlds compatible with what one knows are ones where just one of a, b, c is known. If the latter, they are ones where two of a, b, c are known. Either way, it is not the case that a is known at all the best worlds, and similarly for b or c: so a is not supported, nor is b or c. However, in the former option, at each best world one knows that at least two of a, b, c are true; in the latter, at each best world one knows that at least one of a, b, c is true. So these claims are supported in the corresponding variants. More generally, the toy model diagnoses Preface-style cases with n perfectly symmetrical claims as follows: a basic claim is supported if and only if it is known, and the disjunctive propositiont that *at least* $m - 1$ *claims are true* is supported, where m is the smallest number such that one is in a position to know that does not know that m claims are true. The diagnosis has a flexibility that delivers some intuitive verdicts. If in a given Preface case it does not seem rational for one to believe that *at most one claim in the book is mistaken*, then we would argue that one must in

[19] Here as well, using single-premise closure. See the previous footnote.

[20] One may argue thus: assuming that it is rational to believe what one knows, given that one knows b it is rational for one to believe b. So if it was also rational for one to believe that one of b or c is false then it would be rational for one to believe that c is false, but it is not so. However, the second step replies on closure for rational belief, which is contentious in the present context. I rather rely on the straight implausibility of the claim that in the intended case B one knows b and it is nevertheless rational for them to believe that one of b or c is false.

fact be in a position to know that they do not know that all but one of the claims are true.

The indirect strategy may give up consistency and closure or deny that Preface-style cases are subject to the NED constraint. Our toy model provides a path for the second reply that preserves Consistency and is at least not in obvious conflict with the NED intuition.

Conditionalization. The New Evil Demon problem still arises for credences. Grant that one's credences are rational only if they match the a reasonable prior conditionalized upon one's evidence. Consider the case used in *Bad Consequences* above. A reasonable prior assigns a non-zero probability to Bertrand lying about being sick. In the Good case, one learns that he is sick; in the Bad case, one merely learns that he claims to be sick. Thus conditionalizing on what one knows would yield credence 1 that Bertand is sick in the Good case but a lower credence in the Bad case. But that conflicts with the New Evil Demon intuition.

Reply. The indirect strategist rejects the idea that one should conditionalize on one's *evidence*. Rather, one should conditionalize on *the supported*. In Good cases, that amounts to the same. In Bad cases, that amounts to conditionalizing upon what it is rational to flat-out believe. The upshot is the same, so the NED intuition is respected.

It is of course debatable whether it is rational to assign credence 1 in everything that one knows, let alone everything that one rationally believes. But that is a distinct difficulty for knowledge-first evidentialism from the New Evil Demon problem. So we leave that debate aside.[21]

The conditionalization idea tells in favour of Consistency. If we want to think of rational credences as matching a reasonable prior conditionalized upon some input, we need the input to be consistent. If we directly use one's knowledge as the input, we face a New Evil Demon problem. If we adopt an indiret strategy and use some notion of the supported instead, then what is supported had better be consistent. If we want the supported to be the same across Good and Bad cases, we cannot require the supported to be true. It is not trivial to think that there is a notion of the supported that does not require truth but nevertheless guarantees consistency. Fortunately, however, our toy model shows that there are such notions.

Crypto-internalism. While knowledge-first evidentialism about rationality is advertised as a paradigmatic externalist view, the indirect strategy makes it barely distinguishable for an internalist one on which one's evidence is what appears to one to be so (Huemer, 2007) or what one non-inferentially rationally believes (Comesaña, forth).

[21] See the exchange between Kaplan and Williamson in Greenough and Pritchard (2009) for some discussion.

For the indirect strategy distinguishes two notions, the known and the supported, where the supported plays the traditional roles of evidence as the input to decision theory and conditionalization, and in light of the NED constraint, the supported matches what appears to one to be the case or what is basically rational for one to believe. Save for the claim that the supported supervenes on what one is in a position to know, which internalists need not quarrel with anyway, and the terminological difference of using "evidence" for the latter rather than the former, the view agrees with internalist evidentialism.[22]

Reply. Not squarely sitting on one side of the internalist–externalist divide is not necessarily a bad thing, since the labels are not used here for a specific claim and its negation but for more general outlooks. Admittedly, our aim has been to reconcile knowledge-first evidentialism with the kind of intuition about rationality that motivates internalism. However, the indirect strategy's outlook has some characteristically externalist features.

First, in contrast with views like Huemer's, the indirect strategy does not assign a fundamental role to phenomenology. For a start, the supported need not concide with what appears to be so. It is open, for instance, to say that the supported includes memories even in the absence of mnemic experiences, that it includes propositions about particular objects even though one could have phenomenally identical experiences about other objects, and that it does not include all the content of one's phenomenal experiences.[23] We have also seen how Preface-style cases may lead one to reject the idea that rational belief supervenes on what appears to be so. These claims do not obviously conflict with NED-like intuitions. Moreover, even if the supported coincides with appearances, the indirect strategy is compatible with a kind of externalism in philosophy of mind that reduces appearances to the supported and hence explains appearances in terms of knowledge.

Second, in contrast with views like Comesaña's, the indirect strategy does not use rationality as a primitive. What is rational is explained in terms of what is known—and

[22] *Contra* the objector, internalists may deny that appearances supervene on what one is in position to know—and if they do, "the supported" is not merely another label for appearances. For instance, they may think that some qualitative differences in experience cannot be captured in one's knowledge, so that appearances may differ across two subjects in position to know exactly the same thing. However, it is open for internalists to think that any difference in appearances or any difference in what it is basically rational for one to believe results in some difference in what one is in a position to know. That follows in particular from accessibilist-internalist claims that one is always in position to know what appearances are like or what it is basically rational for one to believe. (On the notion of "accessibilist" internalism see Conee and Feldman, 2004 and Bergmann, 2006.)

[23] Thus the indirect strategy does not face the standard problems for phenomenalist internalist views: the problem of stored beliefs, singular contents, and the speckled hen. See the ***introduction to this volume*** for further discussion and references.

if needed other notions involved in the account of support, such as a "normality" ranking. That is in line with a broadly naturalist outlook available to internalists and externalists alike in which the normative is explained in terms of the non-normative, with the characteristically externalist addition that the *explanans* is knowledge. If successful the strategy has more explanatory benefits than one that treats rationality as a primitive. For instance, in its toy implementation we were able to derive claims about the consistency of what it is rational to believe, about equal rationality across some Good–Bad cases pairs, about what is rational to believe in Preface-style cases, and about defeat.

Third, the indirect strategy is consistent with the idea that Good cases have a primacy over Bad ones. In the optimistic theory of support, for instance, we can acknowledge that beliefs are equally supported in Good and Bad cases while claiming that the support they receive in Bad cases is different in kind and deficient insofar as it essentially involves some form of ignorance of the limits of one's knowledge (compare Pritchard, 2012, this volume***).

The indirect strategy is better pictured as follows. Evidence traditionally plays three roles: it is *what is given*, the *input* to formal principles of rationality such as conditionalization and decision-theoretic principles, and what *rationalizes belief and action*. If the indirect strategy is right no single thing plays all roles. *Knowledge* is what is given to you: your access to the world. The *supported* is what serves as input to formal principles of rationality. Both rationalize belief and action but they do so differently: indirectly so for the former, directly so for the latter. In Good cases the two coincide, so one single body of propositions plays all roles. In Bad cases, however, the given is impoverished in ways that are not themselves given, and as a result it is rational to believe and act as if more was given than what is. The resulting picture is externalist in the sense that rationality is ultimately a matter of what you know—what facts you are aware of—rather than, say, the inner coherence of your mental life.

4 Conclusion

Knowledge-first evidentialism about rationality is the idea that it is rational to do what is best in view of what one knows. Even though the project is well-motivated, it seems to fly in the face of obvious facts about cases of excusable error: namely, the differences in knowledge, including drastic ones, need not make any difference in what it is rational to do. For that reason, few philosophes deem it worth of serious consideration. The verdict is far from obvious, however, for it is in principle possible for distinct bodies of knowledge to rationalize the same thing. That is the loophole in the New Evil Demon problem.

We have explored two strategies to rescue knowledge-first evidentialism through the loophole. The direct one uses what one knows directly as the the input to formal principles of rationality (decision-theoretic principles, conditionalization) to determine what it is rational for one to do. I have argued that it fails with slight more complex excusable error cases. The indirect strategy takes one's knowledge to determine a body of propositions, *the supported*, which is itself to serve as input to formal principles. I have laid out some desirable properties for such a notion, including that of lining up with plausible intuitions on excusable error cases. I have given a toy theory—the *optimistic* theory of support—that instantiates them and pointed out some alternatives. Finally, I have considered a range of objections and suggested ways to answer them. I have fallen short of arguing that the indirect version of knowledge-first evidentialism compares favourably with alternative accounts of rationality. But given how little consideration the view has received so far I hope that the foregoing is enough to make it a contender worth further exploration.

References

Bergmann, Michael (2006). *Justification without Awareness: A Defense of Epistemic Externalism*. Oxford University Press.

Bird, Alexander (2007). "Justified Judging". In: *Philosophy and Phenomenological Research* 74.1, pp. 81–110.

BonJour, Laurence and Ernest Sosa (2003). *Epistemic Justification: Internalism Vs. Externalism, Foundations Vs. Virtues*. Wiley-Blackwell.

Brown, Jessica (2008). "Subject-Sensitive Invariantism and the Knowledge Norm for Practical Reasoning". In: *Noûs* 42.2, pp. 167–189.

Cohen, Stewart (1984). "Justification and truth". In: *Philosophical Studies* 46.3, pp. 279–295.

Comesaña, Juan (forth). "A Plea for Falsehoods". draft of may 2017.

Conee, Earl and Richard Feldman (2004). *Evidentialism*. Oxford University Press.

Foley, Richard (2012). *When is True Belief Knowledge?* Princeton University Press.

Goodman, Jeremy and Bernhard Salow (forthcoming). "Taking a Chance on KK". In: *Philosophical Studies*.

Greenough, Patrick and Duncan Pritchard, eds. (2009). *Williamson on Knowledge*. Oxford University Press.

Hawthorne, John and Jason Stanley (2008). “Knowledge and Action”. In: *Journal of Philosophy* 105.10, pp. 571–590.

Hornsby, Jennifer (2008). “Acting for Reasons”. In: *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press, pp. 244–61.

Huemer, Michael (2007). “Compassionate Phenomenal Conservatism”. In: *Philosophy and Phenomenological Research* 74.1, pp. 30–55.

Ichikawa Jenkins, J. (2014). “Justification is Potential Knowledge”. In: *Canadian Journal of Philosophy*, pp. 184–206.

Jeffrey, Richard (1983). *The Logic of Decision*. University of Chicago Press.

Lasonen-Aarnio, Maria (2010). “Unreasonable Knowledge”. In: *Philosophical Perspectives* 24.1, pp. 1–21.

Lehrer, Keith and Stewart Cohen (1983). “Justification, Truth, and Coherence”. In: *Synthese* 55.2, pp. 191–207.

Lord, Errol (201x). “Defeating the Externalis’s Demons”. In: *this volume*. draft of june 2013.

Magidor, Ofra and J. Hawthorne (forthcoming). “Reflections on the Ideology of Reasons”. In: *Oxford Hanbook of Reasons and Normativity*. Oxford University Press.

Parfit, Derek (2011). *On What Matters*. Vol. 1. Oxford University Press.

Pollock, John L. (1986). *Contemporary Theories of Knowledge*. Rowman & Littlefield.

Pritchard, Duncan (2012). *Epistemological Disjunctivism*. Oxford University Press.

Resnik, Michael D. (1987). *Choices: An Introduction to Decision Theory*. University of Minnesota Press.

Rosenkranz, Sven (2017). “The Structure of Justification”. In: *Mind*.

Savage, Leonard J. (1972). *The Foundations of Statistics*. New York: Dover.

Smith, Martin (2016). *Between Probability and Certainty: What Justifies Belief*. Oxford University Press.

Sutton, Jonathan (2007). *Without Justification*. MIT Press.

Unger, Peter (1975). *Ignorance : a Case for Scepticism*. Oxford University Press.

Weatherson, Brian (2012). “Knowledge, Bets and Interests”. In: *Knowledge Ascriptions*. Ed. by Jessica Brown and Mikael Gerkken. Oxford University Press.

Wedgwood, Ralph (2002). “Internalism Explained”. In: *Philosophy and Phenomenological Research* 65.2, pp. 349–369.

Williamson, Timothy (forth). “Justification, Excuses and Sceptical scenarios”. In: *this volume.*

— (2000). *Knowledge and its Limits*. Oxford University Press.

— (2007). “On Being Justified in One’s Head”. In: *Rationality and the Good*. Ed. by Mark Timmons, John Greco and Alfred Mele. Oxford University Press.

Williamson, Timothy (2010). “Reply to Schiffer”. In: *Williamson on Knowledge*. Ed. by Patrick Greenough and Duncan Pritchard. Oxford University Press, pp. 357–64.
— (2011). “Improbable Knowing”. In: *Evidentialism and its Discontents*. Ed. by T. Dougherty. Oxford University Press.