Possibility, Necessity and Probability. A Meditation on Underdetermination and Justification*

Elia Zardini

LOGOS, Logic, Language and Cognition Research Group

Department of Logic, History and Philosophy of Science

University of Barcelona

Northern Institute of Philosophy

Department of Philosophy

University of Aberdeen

elia.zardini@ub.edu

August 29, 2013

1 Introduction and Overview

Metaphysical and *epistemic* modality seem to lie very far apart from one another: for example, it is metaphysically possible for snow to be black (things might have been such that snow would be black), but it is not epistemically possible for snow to be black (we know full well that it is not). Yet, time and again in the history of Western epistemology,

*This paper grew out of my response to Tony Brueckner's paper '$\sim$K$\sim$SK' at the 2009 Arché Basic Knowledge Conference *Contemporary Perspectives on Scepticism* (University of St Andrews). Versions of the material in the paper have then been presented in 2011 at the NIP Basic Knowledge Seminar (University of Aberdeen) and at the COGITO Epistemology Seminar (University of Bologna); in 2012, at the LOGOS Seminar (University of Barcelona); in 2013, at the session *Scepticism and Epistemic Circularity* of the 23rd World Congress of Philosophy *Philosophy as Inquiry and Way of Life* (University of Athens). I'd like to thank all these audiences for very stimulating comments and discussions. Special thanks go to Tony Brueckner, Annalisa Coliva, Dylan Dodd, Miguel Ángel Fernández, Dan López de Sa, Genoveva Martí, Sebastiano Moruzzi, Sven Rosenkranz, Daniele Sgaravatti, Martin Smith, Ernie Sosa, Giorgio Volpe, Richard Woodward, Crispin Wright, José Zalabardo and several anonymous referees. At different stages during the writing of the paper, I've benefitted from an AHRC Postdoctoral Research Fellowship, from a UNAM Postdoctoral Research Fellowship and from a FP7 Marie Curie Intra-European Research Fellowship.

we find epoch-making arguments relying on what is apparently taken to be a compelling transition from the former to the latter. Thus writes René Descartes:

> However, an old opinion sits in my mind, that there is a God who can do everything, and by whom I have been created such as I am. Whence do I then know that he did not determine that there be no earth at all, no sky, no extended bodies, no shape, no magnitude, no space, and that nevertheless all these things seem to me to exist in the same way as they now do? (*First Meditation*)[1]

Arguably, in this and other passages, Descartes is not inferring a lack of *knowledge* or *justification* from a metaphysical possibility; given his declared project in the *Meditations*, he is best seen as inferring a lack of *certainty* from such a possibility. Still, contemporary versions of so-called '*Cartesian scepticism*' have typically taken the apparently compelling transition from metaphysically modal facts variously involving its being the case that P to lack of knowledge or justification for believing that P as the core of a powerful sceptical argument. The crucial transition of the argument is concisely described by Robert Audi in his influential textbook thus:

> Suppose, for instance, that I might be having an auditory hallucination of bird songs. Then my present experience of (apparently) hearing them may not justify my believing that there are birds nearby and is certainly not a sufficient basis for my knowing there are, even if it is true that there are. (Audi [2003], p. 296)

The thought seems to be that, very roughly, if all I have to go on in deciding whether P is something which is metaphysically compossible with or additionally even entailed by something incompatible with its being the case that P, then what I have can't really tell in favour of its being the case that P—its being the case that P is "*underdetermined*" by it. As Audi goes on to explain:

> [...] if our experience *underdetermines* the truth of propositions we commonly believe about the external world, roughly in the sense that it does not decisively indicate their truth as opposed to the truth of skeptical (or other) alternative hypotheses that can explain our experience, how can our experience justify our believing such commonsense propositions? (Audi [2003], p. 299)

I'll be concerned in this paper with a critical scrutiny of this sceptical thought, arguing that, in spite of its historical credentials, there is surprisingly little to it. Epistemic facts are much more independent from metaphysically modal facts—and, in particular, from facts of underdetermination of one kind or another—than the sceptical thought requires. Before embarking on such scrutiny, however, some preliminary remarks about focus and methodology are required.

[1] All translations from the *Meditations on First Philosophy* are mine.

As for *focus*, I'll explicitly examine sceptical underdetermination arguments only in their bearing on *prima facie propositional* justification for *perceptual* beliefs. I think that *prima facie* propositional justification (henceforth simply 'justification') is one of the most fundamental epistemological notions, so that the restriction to it should not detract too much from this paper's interest. Moreover, it should often be clear how my discussion about justification can be naturally extended to cover also *some other epistemological notions* (for example, the notion of *knowledge*),[2] although some of the issues arising here do deserve further investigation that goes beyond the scope of this paper. I regard the further restriction to perceptual beliefs as even less of a limitation: it seems to me that sceptical underdetermination arguments completely analogous to those I'll be considering for perceptual beliefs can be run for beliefs formed by (possibly the combination of) *other faculties, processes and methods* (memory, testimony, induction etc.), and that points completely analogous to those I'll be making apply.

As for *methodology*, I'll argue for claims about justification and other epistemic properties by (among other things) appealing to intuitively compelling and widely accepted ordinary judgements about cases, which entail the fallaciousness of certain kinds of inferences employed in sceptical underdetermination arguments. Such appeal will not be question begging against the sceptic—at least, against the kind of sceptic I primarily have in mind. For the strategy of that sceptic is one of, as it were, *internal undermining*: rather than arguing that perceptual beliefs fail to meet a necessary condition on justification arrived at by relatively unconstrained philosophical reflection, she argues that it is exactly the *ordinary* conception of justification, as articulated in ordinary judgements (in particular, in those in which it is *denied* that someone is justified), that surprisingly enforces sceptical conclusions. (The fact that, in spite of this, we still make anti-sceptical attributions of justification can then be explained in terms of an understandable if inconsequent anti-sceptical bias on our part.) Against such internal sceptic, it is only fair play to show that certain inferences playing a crucial role in her argument have in fact no support at

[2]As for knowledge in particular, since I've often come across the claim that underdetermination arguments are really supposed to be arguments for scepticism about knowledge rather than justification, I'd like to note at least the following. Firstly, some underdetermination arguments for scepticism about knowledge actually proceed *by first establishing scepticism about justification* (see e.g. Brueckner [1994]), and so are immediately affected by my discussion. Secondly, while the points to be made about the *probabilistic* fallacies of the arguments for scepticism about justification apply with somewhat less force to the analogous arguments for scepticism about knowledge (given the wider *gap* between probability and knowledge than between probability and justification), the points to be made about the *intuitive* fallacies of the arguments for scepticism about justification do apply with equal force to the analogous arguments for scepticism about knowledge. Thirdly, at least as far as *perceptual* beliefs are concerned, it is highly plausible to think that none of the usual obstacles typically stand in the way on the path from justified (true) belief to knowledge (perceptual beliefs are typically not based on false lemmas, are typically in the right sort of causal connection with the facts, are typically reliable etc.), and so that *perceptual beliefs typically amount to knowledge if they are justified (and true)*. This highly plausible connection between justification and knowledge for perceptual beliefs gives reasons to be very suspicious of any argument for scepticism about knowledge that would irremediably break down if run instead in terms of justification. It is on the strength of these considerations that it'll sometimes be appropriate in our discussion to refer to works whose primary topic are underdetermination arguments for scepticism about knowledge. Thanks to two anonymous referees for comments that prompted this fn.

all in the ordinary conception of justification, and I'll do so by producing cases—clearly apt to be multiplied *ad libitum*—where, according to intuitively compelling and widely accepted ordinary judgements, those inferences are *unequivocally* fallacious. (Indeed, I'll do more than this: I'll also show that the relevant inferences are fallacious from the point of view of the theory of *probability*, which is widely accepted as a good model of certain epistemological notions involved in the underdetermination sceptic's inferences.) In fact, although I could easily have produced cases involving perceptual beliefs, I'll make the point producing cases involving beliefs formed by (possibly the combination of) other faculties, processes and methods, both to indicate that the underdetermination sceptic's inferences are fallacious *across the epistemic board* and to assuage her worry that the cases involving perceptual beliefs are subject to the *noise* generated by our anti-sceptical bias (see fn 24 for some further background on internal scepticism).[3]

The rest of the paper is organised as follows. Section 2 presents the structure of the sceptical underdetermination argument that will be the focus of the paper. The argument assumes that it is metaphysically possible for a deceived subject to have the same evidence that a non-deceived subject has, and tries to draw consequences about justification from that assumption of metaphysical possibility. Sections 3 and 4 variously object to the transition from the assumption to its supposed consequences. Sections 5 and 6, the central part of this paper, critically consider some influential ways of bridging the gap between the assumption and its supposed consequences, which generally consist in strengthening the assumption from one of metaphysical possibility into one of either counterfactual implication or entailment. Section 7 concludes by drawing some lessons from the foregoing discussion and tracing some of its limits.

2 The Structure of the Sceptical Underdetermination Argument

We can start our discussion following the main lines of Tony Brueckner's reconstruction of the structure of the sceptical underdetermination argument as we'll understand it (see

[3]While decisive against the internal sceptic presented in the text, all this is of course less decisive against *more extreme sceptics* who employ necessary conditions on justification which are arrived at by relatively unconstrained philosophical reflection and which are not satisfied in the cases I'll produce (although I should confess that I myself don't know of any remotely plausible "relatively unconstrained philosophical reflection" establishing the desired necessary conditions). Such sceptics may try to stick to their guns and conclude from the cases I'll produce that the ordinary conception of justification is misguided through and through. Although these sceptics are not the primary target of this paper, I note that the *complete lack of support in the ordinary conception of justification* for their favoured necessary conditions on justification should still give them some pause (as an instance of the general principle that the complete lack of support in the ordinary conception of Fness for her favoured necessary condition on being F should give some pause to someone theorising about Fness), and that they face the problem of the *probabilistic invalidity* of the principles they are relying on no less than internal sceptics do. Thanks to Dan López de Sa, Genoveva Martí and an anonymous referee for pressing me on this issue.

e.g. Brueckner [1994]).[4] Let's take any best candidate for being a perceptually justified proposition, say the proposition that there is a fire in the room (let $FIRE$ be that proposition and let René be a subject for which $FIRE$ is such a best candidate). Let's call one's *evidence* the totality of those entities, whatever they exactly are, that, being involved in some kind or other of *epistemic achievement* on one's part, are apt for helping to constitute one's justification for believing some proposition or other[5,6] (and let's also allow for the natural restricted notion of one's *evidence for a specific proposition*).[7] Thus, for example, René's experience as of a fire in the room, being involved in an (unremarkable) epistemic achievement on René's part and being apt for helping to constitute René's justification for believing $FIRE$, is part of René's evidence. Without further prejudging the issue of the exact general kind and specific nature of René's evidence for $FIRE$, let his evidence in our scenario—whatever it exactly is—be f and let F be a proposition that is a best candidate for representing f.

Now, for f—as for just about any evidence of anyone for any candidate for being a perceptually justified proposition—it would seem possible to contrive a metaphysically possible scenario in which f still is René's evidence[8] but in which $FIRE$ is false. Settle then on your favourite scenario that would seem to fit the bill (let it be adequately described by

[4]As far as I know, the paper that, in contemporary epistemology, has put into centre stage the sceptical underdetermination argument is Yalçin [1992]. A little industry has then followed (in addition to Brueckner's own [1994] paper, see Cohen [1998]; Vogel [2004]; Pritchard [2005]; Brueckner [2005]; [2010]). I recommend all these works for their insightful discussions of issues surrounding the sceptical underdetermination argument. However, such works usually focus on logical and dialectical connections between underdetermination arguments and other arguments for scepticism (typically, closure ones) and do not typically offer an in-depth exploration of the various ways in which the sceptical underdetermination argument's inference from metaphysically modal facts to lack of justification may go, let alone a critical discussion of these (I'll mention in fns 25, 26 and 34 some of the main passages in which such issue emerges in that specific literature or in some other relevant works). The main aim of this paper is to contribute towards such exploration and discussion.

[5]Thus, throughout, by 'evidence' I mean '*total* evidence'; in the following, for *partial* evidence I'll talk of "pieces" and "parts" of evidence.

[6]On some views, the justification is constituted by certain *objects* consisting in fire-like sense data; on some other views, the justification is constituted by the *event* of undergoing an experience as of a fire in the room; on yet some other views, the justification is constituted by the *fact* that a reliable process has delivered the information that there is a fire in the room. (I don't mean to suggest any strict correlation between the choice of a *general* kind of entity (e.g. events *vs* facts) and the *specific* nature of the evidence (e.g. dogmatism *vs* reliabilism)—various combinations are certainly possible and the ones I've used are merely meant for the sake of example.)

[7]This notion of a subject's evidence is related to but importantly different from the notion usually expressed by the use of the phrase 'a subject's evidence' by those epistemologists who endorse a *propositionalist* conception of evidence. That notion usually applies to those objects (i.e. propositions) possession of which by the subject is the subject's evidence in the sense defined in the text (the possession in question consisting in the subject's bearing an appropriate attitude or some other kind of suitable relation towards such propositions). Having thus explained the relationships between these two notions of evidence, and how to go from the latter to the former, by 'evidence' I'll henceforth mean the notion defined in the text. Also, I emphasise that such notion does not involve any *accessibility* requirement (contrary to the etymology and at least one standard use in contemporary epistemology of the word 'evidence'): the objects, events or facts in question may be such that their existence lies beyond what a subject may even in principle come to ascertain.

[8]In the following, for ease of expression, instead of saying that, in a certain situation, evidence e is one's

the proposition $DEMON$,[9] which we'll thus assume to be incompatible with $FIRE$). It would then follow that:

(MP) It is metaphysically possible that $F \wedge DEMON$ is true

holds.

(MP) has suggested to some epistemologists that:

($\mathrm{NF^F}$) f does not favour $FIRE$ over $DEMON$

also holds. 'Favour' can hardly lay any claim to being taken as an epistemological primitive, and we'll in fact see in section 6 that an unreflective usage of this verb can sometimes contribute to significantly obscuring matters. I'll offer in that section an appealing framework in which to make more precise sense of 'favour'-talk; however, for the time being I'll acquiesce in the habit prevailing in at least a certain strand of discussion of sceptical underdetermination arguments, and thus provisionally treat 'favour' as an epistemological primitive, appealing to unreflective intuitions concerning contents expressed with the help of that verb.

One such intuition concerns the principle that *justification requires the lack of underdetermination by the evidence among alternatives*:

(JNU) If evidence e justifies one for believing a hypothesis H,[10] and H is incompatible with H^*, e favours H over H^*.

Since $FIRE$ is incompatible with $DEMON$, it then follows by ($\mathrm{NF^F}$) and (JNU) that f does not justify René for believing $FIRE$.[11]

---

evidence, I'll simply say that e exists.

[9] I use '$DEMON$' for obvious historical reasons, but our discussion will be neutral as to what $DEMON$ exactly is (see fn 34 for an additional, frequently mentioned constraint on the choice of $DEMON$).

[10] Throughout, I use 'justify' and its likes in the sense of 'justify *by itself*, without the need of anything else' and its likes. An analogous remark holds for 'favour' and 'support'.

[11] On many understandings of favouring (but not on the second one to be introduced in section 6), evidence e not favouring a hypothesis H over an incompatible hypothesis is arguably *much stronger* than e not justifying one for believing H, *and obviously so*. It might then be wondered why the sceptical underdetermination argument interpolates ($\mathrm{NF^F}$) in the route from (MP) to (JNU) rather than directly inferring (JNU) from (MP). The reason is that one natural way in which the inference from (MP) to (JNU) can be motivated is precisely by claiming that (MP) shows that, given f, *$FIRE$ is in no better standing than* a hypothesis that is incompatible with it (as $DEMON$ is). And what that claim amounts to is precisely the validity of the inference from (MP) to ($\mathrm{NF^F}$). More generally, all the sceptical underdetermination arguments I'll be considering fit into the mould—naturally suggested by the label 'underdetermination'—of involving an inference to the *absolute* claim that f does not justify René for believing a hypothesis H from the *relative* claim that f does not favour H over an incompatible hypothesis. In addition to that being a natural route from metaphysically modal facts to (JNU), notice that ($\mathrm{NF^F}$) and its likes can presumably play their mediating role in such route even if they are weakened by, for example, replacing 'does not favour' with 'does not significantly favour'. Thanks to Richard Woodward for discussion that helped to bring out this point.

Now, f is René's evidence. Hence, if f does not justify René for believing $FIRE$, René's evidence does not justify him for believing $FIRE$. However, given the above explication of what evidence is, the principle that *justification requires evidence*:

(JE) If one is justified for believing a proposition, one is so justified by one's evidence

seems to hold. Notice in particular that, given the above explication of what evidence is (and especially fn 7), (JE) does not amount to a confession of *internalism*, since a subject's evidence is allowed to be *external and inaccessible* to the subject (and it does not amount to a confession of *empiricism* either, since a subject's evidence is allowed to be *non-empirical*). However, I should mention here that (JE) would be rejected by certain *concessive anti-sceptical epistemologies* (such as that of Wright [2004]), according to which the underdetermination sceptic is right in thinking that f *in itself* does not justify René for believing $FIRE$, but only f *together* with certain facts that do not involve any epistemic achievement on René's part. In this paper, however, I'll mostly take (JE) for granted and focus instead critically on the subargument which starts with assuming (MP) or strengthenings thereof and ends with the intermediate conclusion that f does not justify René for believing $FIRE$.[12]

Before doing that, let's however conclude our presentation of the whole argument by observing that, since René's evidence does not justify him for believing $FIRE$, it then follows by (JE) that René is not justified for believing $FIRE$. Finally, as already noted, the analogues of (MP) would seem to hold for just about any evidence of anyone for any candidate for being a perceptually justified proposition, and so, given the analogues of the principles and inferences employed in the argument, a wide-ranging scepticism about perceptual justification would ensue. This is the broad structure of the sceptical underdetermination argument as we'll understand it. In the following, we'll start (in sections 3 and 4) by examining the argument as it presently stands, which will provide a useful background for then, in the central part of this paper (in sections 5 and 6), moving on to scrutinise other apparently more troublesome versions of the argument that modify (MP) and possibly (NF^{F}) too (let's call this version of the sceptical underdetermination argument 'the (MP)-argument' and let's make corresponding stipulations).[13]

[12]While I'll thus not directly touch on the question that is of paramount importance for these epistemologies (that is, the question whether (JE) holds), I should remark that my very critical treatment of that subargument does not sit completely comfortably with them, at least to the extent to which they take the compellingness of that subargument as a reason for exploring the possibility that the *prima facie* appealing (JE) might fail.

[13]Some authors (e.g. Stroud [1984], pp. 1–38; Vogel [2004], pp. 426–429) discuss a similar argument (most of the time in terms of knowledge rather than justification), whose crucial assumption is roughly that, in order to be justified for believing $FIRE$ on the basis of his perceptual engagement with the fire in the room, René needs to have an *independent* justification for rejecting the hypotheses incompatible with $FIRE$ whose truth is metaphysically compossible with the rest of f. Although it is not perfectly clear what kind of independent justification is required, since the only problem made out with the rest of f (which, including René's experience as of a fire in the room, would have seemed to constitute a pretty good justification for rejecting hypotheses incompatible with $FIRE$) is simply its metaphysical compossibility with $DEMON$'s truth, I would suppose that the pieces of evidence constituting the independent justification

3 Kinds of Alethic Modalities

There would be much to say about the first step of the sceptical underdetermination argument consisting in the assumption of a certain metaphysically modal fact. In this paper, let's however grant that assumption, which rules out world-involving interpretations of f (as containing, for example, the fire itself, or René's seeing the fire, or René's knowing that there is a fire in the room), and let's assume that it is compatible with f at least containing René's experience as of a fire in the room (see Zardini [2013b] for a discussion of these issues). Let's start then our critical appraisal of the argument by focussing on its second step, consisting in the inference from that fact to lack of favouring. A preliminary consideration here concerns the kind of alethic modality in question. That is *metaphysical* modality and it is interesting to observe that under no decently natural strengthening of the nature of the possibility in question does (MP) clearly retain its plausibility. Natural strengthenings that come to mind are *nomological* and *technological* possibility, and it is unclear whether and how (MP) could be preserved under such strengthenings. But before substantiating that contention, given that nomological and technological possibility will play a crucial role in this section, a couple of words about how I'll understand these notions are in order. The relevant notions of nomological and technological possibility in this kind of discussion are based on the traditional and familiar scholastic picture of a world governed by (non-probabilistic) *laws of nature* but also containing a variety of subjects of *spontaneous* action—i.e. action constrained but not determined by those laws. (Among such subjects count of course human beings.) Technological possibility further constrains the possible spontaneous actions on a subject's part to those made available by the subject's *skills* and *tools*.[14]

Importantly, both nomological and technological possibility are so strong as to allow, contrary to metaphysical possibility, for a non-vacuous *relativisation to times*. In assessing whether (MP) can so be strengthened to a claim of nomological or technological possibility, we must thus attend to the relevant time of relativisation. Now, given the way we broadly

are at least supposed to be metaphysically incompossible with $DEMON$'s truth (at least once taken in conjunction with the rest of f), so that, given f's being *total*, f's justifying René for believing $FIRE$ is actually thought to require that f's existence be metaphysically incompossible with $DEMON$'s truth. If so, my critical discussion of the (MP)-argument carries over to this argument. However, sometimes (as in some options mentioned by Vogel [2004], p. 429), one has the impression that this is not so and that the kind of independent justification required is not even supposed to be constituted by evidence (as though, while metaphysical compossibility with $DEMON$'s truth were damning for the justificatory prospects of objects, events and facts involved in some kind or other of epistemic achievement, it were not damning for the justificatory prospects of objects, events and facts not so involved), so that, given f's being *evidence*, f's justifying René for believing $FIRE$ is, again, actually thought to require that f's existence be metaphysically incompossible with $DEMON$'s truth. If so, again, my critical discussion of the (MP)-argument carries over to this argument (see also fn 34).

[14]The picture is in strong tension with both *classical* (*determinist*) and *contemporary* (*probabilistic*) physical theories. However, I think that it is arguably the notion of nature as articulated in the traditional and familiar scholastic picture that is at play in the informal notion of "what is possible in nature" that is in turn most relevant for modally sensitive ordinary notions such as epistemic ones—witness the issues raised for the latter notions by the "quantum miracles" allowed by mainstream theoretical physics (Hawthorne [2004], pp. 4–5 is an early reference for this problem).

take the world to be, at any time sufficiently *close* to the time of René's experience as of a fire in the room, it is no longer nomologically or technologically possible that $F \wedge DEMON$ is true (indeed, it is no longer nomologically or technologically possible that $DEMON$ itself is true). One might want to consider instead times that are arbitrarily *remote* in the past from the time of René's experience as of a fire in the room. In fact, it seems that interesting notions of nomological and technological possibility are also defined by taking *different initial conditions* from those in fact obtaining at the world at which the possibility claim is evaluated (with the requirement that such conditions be in some sense compatible with the laws governing that world). However, even considering times that are arbitrarily remote in the past from the time of René's experience as of a fire in the room, and even considering different initial conditions from those in fact obtaining at our world, it is a real question concerning the bounds of nomological and technological possibility how many instances of the envisaged strengthening of (MP) really hold. For one thing, most kinds of illusions we know of (and hence most kinds of illusions which, possibly considering different initial conditions, clearly are technologically or at least nomologically possible) do not seem to display anything like the *vividness*, *richness* and *consistency* of experiences that successfully represent the world.[15] Even when they seem to do so (as during a dream or a drug-induced hallucination), this is typically not because of their *intrinsic character*, but rather because they are accompanied by a *debilitation* of one's ability to evaluate the character of one's own experience.

Now, the inference from (MP) to (NF^{F}) seems to be a transition from *compossibility* (of f's existence with $DEMON$'s truth) to lack of disfavouring (of $DEMON$ over any alternative—in particular $FIRE$—on the part of f). Postponing for the time being a critical discussion of whether that is in general a valid transition, the question has to be raised as to why a privileged role is given in the (MP)-argument to *metaphysical* compossibility, given that, as we've seen in the last paragraph, (MP) is false if it is strengthened, say, to a claim of nomological possibility relative to the present time. Why should metaphysical possibility—rather than, say, nomological possibility relative to the present time—play a privileged role in determining what the evidence favours and fails to favour?

In fact, if a privileged role were given instead to nomological possibility relative to the present time, the situation would be even worse for (NF^{F}). For, if a certain kind of possibility plays a privileged role in determining what the evidence favours and fails to favour, one would expect not only that that kind of compossibility speaks in favour of lack of disfavouring, but also, conversely, that that kind of incompossibility speaks in favour of

[15]The point was forcefully and extensively made by Austin [1962], pp. 47–50. With respect to that particular kind of illusion generated by dreams, it is actually anticipated by Descartes himself in the *First Meditation* when he briefly remarks about his waking experiences that "[...] no such distinct experiences would occur to one who is asleep". I believe that the point is not retracted in the immediately following passage (which may be interpreted in that sense): "As though I did not remember that already some other times I have been deceived in dream by similar thoughts; and now that I think more carefully about these things, I see so clearly that waking can never be distinguished from dream by signs that are certain that I am astonished". Rather, I think that, in that passage, Descartes is pointing out that broadly similar thoughts occurring in dream are sometimes false. That observation may be enough for Descartes' purposes (see section 1), but not for those of the (MP)-argument.

disfavouring (over at least certain alternatives, for example those that are on the contrary compossible with the existence of the evidence). Now, as we've seen in the second last paragraph, at any time sufficiently close to the time of René's experience as of a fire in the room, it is in fact both nomologically and technologically *impossible* that $F \wedge DEMON$ is true. Indeed, given our understanding of what $DEMON$ could be, this means that, at any time sufficiently close to the time of René's experience as of a fire in the room, it is both nomologically and technologically *impossible* that $F \wedge \neg FIRE$ is true—i.e. it is both nomologically and technologically *necessary* that $F \rightarrow FIRE$ is true.[16] Given that, apart from considerations of alethic modality, f, including as it does René's experience as of a fire in the room, seems well positioned to favour $FIRE$ over $DEMON$, if nomological modality relative to the present time played a privileged role in determining what the evidence favours and fails to favour these nomologically modal (relative to the present time) facts would then plausibly settle that f does favour $FIRE$ over $DEMON$. And that seems anyways a plausible view: to repeat, given how things are at any time sufficiently close to the time of René's experience as of a fire in the room, *the laws of nature themselves* rule out the possibility that $F \wedge DEMON$ is true, and that, especially when coupled with the necessity, determined by these laws, that $F \rightarrow FIRE$ is true, seems part of a pretty good reason for thinking that f favours $FIRE$ over $DEMON$.[17]

Similar points apply to the strengthenings of (MP)—counterfactual implication and entailment from $DEMON$ to F—that are exploited by the versions of the sceptical underdetermination argument discussed in sections 5 and 6: if $DEMON$ is not nomologically possible (relative to the present time), then the nomological counterfactual implication and the nomological entailment (relative to the present time) from $DEMON$ to F are at best only *vacuously* true—they do not describe any *non-trivial correlation* in nomologically possible (relative to the present time) worlds—and such truth seems to be vastly outweighed by the *non-vacuous* truth of the nomologically necessary (relative to the present time) implication from F to $FIRE$. While certainly deserving a much more in-depth discussion than is possible in this paper, all these asymmetries between metaphysical modality and other natural kinds of alethic modalities do at least cast some suspicion on the very idea of a transition from metaphysically modal facts to lack of favouring. The rest of this paper will try to substantiate the suspicion.

[16]For those of us who see a difference between any *real implication* and so-called '*material implication*', I observe that '$\rightarrow$' over and above '$\supset$' is fully warranted: the impossibility of $F \wedge \neg FIRE$ is due to the impossibility of the *combination* of F with $\neg FIRE$ rather than to the *independent* impossibility of either (which does not subsist: in particular, even at times sufficiently close to the time of René's experience as of a fire in the room, René might have decided eventually not to light any fire).

[17]I should emphasise that I don't mean in the least to suggest that a necessary implication is *by itself* sufficient for favouring: as I've mentioned in the text, it's the fact that, apart from considerations of alethic modality, f seems well positioned to favour $FIRE$ over $DEMON$ that makes the necessary implication part of a pretty good reason for thinking that f does favour $FIRE$ over $DEMON$. Thanks to two anonymous referees for comments that led to improvements in the last two paragraphs in the text.

4 Metaphysical Compossibility, Positive Probability, Favouring and Justification

The considerations developed in section 3 are meant as preliminary remarks that do cast some suspicion on the very idea of a transition from metaphysically modal facts to lack of favouring, but that by no means tell decisively against it. We should now look a bit more in detail at how that transition in fact fares in the particular case of the inference from (MP) to (NF^{F}).

It doesn't fare very well. Quite generally, evidence e can favour a hypothesis H over an incompatible hypothesis H^* even if it is metaphysically possible that [e exists and H^* is true].[18] For example, my evidence containing my seeming memory of having thought about this paper for a while clearly favours ⟦I have thought about this paper for a while⟧[19] over ⟦I have not thought about this paper for a while⟧ even if it is metaphysically possible that [my evidence exists and I have not thought about this paper for a while].[20] The uncontroversial fact that my evidence containing my seeming memory of having thought about this paper for a while and my not having thought about this paper for a while are metaphysically compossible clearly does not overturn (nor is in any way in contrast with) the equally uncontroversial fact that my evidence containing my seeming memory of having thought about this paper for a while supports[21] ⟦I have thought about this paper for a while⟧ and does not support ⟦I have not thought about this paper for a while⟧,[22] and so that it favours

[18] Throughout, I use square brackets to disambiguate constituent structure in English.

[19] Throughout, I use '⟦φ⟧' to denote the proposition expressed by φ.

[20] The same point could have been made with an example which, instead of relying on the *use of a specific faculty for forming beliefs*, simply relies on *brutely probabilistic considerations*. Thus, suppose that I own one ticket of a fair lottery with 1,000,000 tickets (where the lottery has the peculiarity that its existence and functioning do not require the existence of the external world). Then, my evidence containing the information about the lottery clearly favours ⟦My ticket will lose⟧ over ⟦My ticket will not lose⟧ even if it is metaphysically possible that [my evidence exists and my ticket will not lose]. Similar examples can be given also against the other inferences to be discussed in sections 5 and 6. The same point could also have been made with an example relying on *induction*. By induction, my evidence so far favours ⟦It is not the case that in the next second I'll start having thoroughly realistic experiences as though living in 15^{th}-century Tenochtitlan⟧ over ⟦In the next second I'll start having thoroughly realistic experiences as though living in 15^{th}-century Tenochtitlan⟧ even if it is metaphysically possible that [my evidence exists and in the next second I'll start having thoroughly realistic experiences as though living in 15^{th}-century Tenochtitlan]. Again, similar examples can be given also against the other inferences to be discussed in sections 5 and 6.

[21] Throughout, I take 'support' and its likes to express a gradable notion, and I understand the non-comparative uses of such expressions to mean something along the lines of 'speak to some substantial extent in favour of'. Since evidence (e.g. Smith's having left his fingerprints on the crime scene) can speak to some substantial extent in favour of a hypothesis (e.g. Smith's being the murderer) without justifying outright belief in the hypothesis, I thus take support to be weaker than justification. I'll introduce below in the text a well-known measure over support.

[22] On some views of the structure of memorial support, my seeming memory of having thought about this paper for a while does not really *in itself* support ⟦I have thought about this paper for a while⟧: it is only that state *in conjunction with* my evidence for the relevant general correlations between my seeming memory of having Fed and my having Fed that supports ⟦I have thought about this paper for a while⟧. The point in the text still applies to this conception of the structure of memorial support, as long as my

the former over the latter.[23,24]

Something like this point is relatively uncontroversial in the contemporary epistemological literature under the broad heading of '*fallibilism*'.[25] Sometimes (as e.g. in Vogel [2004], p. 427), the point is however made exclusively by taking as examples relatively *high-level* epistemic practices such as induction (see fn 20), while, in the dialectic against the internal sceptic, it is important to realise that it applies across the epistemic board, including such a *bottom-level* epistemic practice as relying on one's memory (so that the internal sceptic is barred from replying that relatively high-level epistemic practices such as induction somehow deviate from the ordinary conception of justification as articulated in more basic epistemic practices). However, as already variously mentioned, there are versions of the sceptical underdetermination argument that start with metaphysically modal facts stronger than compossibility, and so might succeed even if the (MP)-argument fails. I'll turn to these apparently more troublesome versions in sections 5 and 6, basically arguing that, perhaps surprisingly, *the situation does not improve much in spite of the strengthened metaphysically modal premise and that points similar to that just made apply with equal*

evidence for certain general correlations between my seeming memory of having Fed and my having Fed is metaphysically compossible with [me having a seeming memory of having thought about this paper for a while but not having thought about this paper for a while] (which it certainly is).

[23]Letting H and H^* be incompatible, in this section and section 5 I assume that the inference from 'Evidence e supports hypothesis H and e does not support hypothesis H^*' to 'e favours H over H^*' is valid. The assumption is extremely plausible, especially when treating 'favour' as an epistemological primitive (as we're doing in this section and section 5). I'll define a notion of favouring in section 6 under which the inference arguably fails. As I'll discuss there, that notion will however still be a notion under which the inference from (MP) to (NF^{F}) is fallacious (and under which (JNU) becomes very problematic).

[24]Now that our discussion has become more concrete, let me add some further background about internal scepticism. The internal sceptic proposing the (MP)-argument is struck by the fact that we do not take my evidence containing my seeming memory of having thought about this paper for a while to favour ⟦I have thought about this paper for 100 days⟧ over ⟦I have thought about this paper for 101 days⟧ as well as by the fact that we do not take my evidence containing that seeming memory to justify me for believing that I have thought about this paper for 100 days. She claims that the *explanation* of these ordinary judgements is given by the fact that the existence of my evidence containing the seeming memory is metaphysically compossible with me having thought about this paper for 101 days and by the epistemic principle that metaphysical compossibility of the existence of the evidence with the truth of a hypothesis implies lack of disfavouring of that hypothesis over any alternative on the part of the evidence. She then applies such independently motivated epistemic principle to vindicate the inference from (MP) to (NF^{F}). The example in the text shows that the internal sceptic's extrapolation of the epistemic principle from the relevant ordinary judgements has been overhasty, and that that principle has in fact no support at all in the ordinary conception of justification (of course, this leaves epistemologists with the task of coming up with a better explanation of the relevant ordinary judgements). Analogous comments apply to the other inferences to be discussed in sections 5 and 6. Thanks to Dan López de Sa, Genoveva Martí and an anonymous referee for questions that prompted this fn.

[25]Less so if one restricts attention to works on scepticism. For example, some passages in the literature mentioned in fn 4 do express sympathy for the the inference from metaphysical compossibility to lack of justification (see e.g. Cohen [1998], p. 148, fn 11; Pritchard [2005], p. 52). Also, the discussion in fn 13 suggests that another familiar way of running the sceptical underdetermination argument in fact implicitly assumes that the inference from metaphysical compossibility to lack of justification is valid, and the discussion in fn 26 intimates that a recent, very influential treatment of scepticism actually takes as its model of sceptical argument one that involves the inference from metaphysical compossibility to lack of knowledge. Thanks to two anonymous referees for discussion of this literature.

force (which is why it'll be useful to have started our discussion with the (MP)-argument). Before turning to that task, it'll however prove fruitful to consolidate and extend the present point by introducing a new kind of modality which will be very relevant for our subsequent discussion: *probability on the evidence*.

The inference from (MP) to (NF^{F}) is *intuitively* fallacious.[26] By introducing the notion of probability on the evidence, we can show that the inference is also *probabilistically* fallacious (if one so likes, intuitive fallacy about favouring may be thought to stand to probabilistic fallacy about favouring just as intuitive fallacy about conjunction stands to sentential-logic fallacy about conjunction). The notion of probability on the evidence is, very roughly, a notion of the *measure* of the *support* that the evidence gives to a hypothesis.[27] For modelling purposes, we'll take a *hypothesis* (i.e. a proposition) to be a set of worlds (i.e. the set of worlds at which the hypothesis is true) and, given a non-empty set W of (contextually relevant) worlds, we'll take the *probability on the evidence* (in that context) to be a function with certain constraints assigning to each subset of W a real number in $[0, 1]$—the likelihood given the evidence (possessed in that context) that the actual world is a member of that subset. More precisely:

Definition 1. A *probability space* $\mathcal{S}$ is a pair $\langle W^{\mathcal{S}}, \mathsf{Pr}^{\mathcal{S}} \rangle$ where:

- $W^{\mathcal{S}}$ is a non-empty set of worlds;
- $\mathsf{Pr}^{\mathcal{S}} : \wp(W) \mapsto [0, 1]$ is a *probability function* on $W^{\mathcal{S}}$.

A natural kind of possibility definable within the modality of probability on the evidence is constituted by the property of a hypothesis of *having positive probability on the evidence*. Since that modal property is arguably two-way independent from the other modal property we've focussed so far of a hypothesis of *being such that its truth is metaphysically compossible with the existence of the evidence*, a new, independent route for getting to (NF^{F}) opens up, for, instead of taking as a premise the claim of metaphysical compossibility with the

[26] Williamson [2000], pp. 164–183 considers an argument (in terms of *knowledge* rather than *justification*) whose crucial lemma is that in *DEMON* René has the same evidence as he has in his original, epistemically best situation. Williamson's treatment of the argument consists in rejecting the *same-evidence lemma* (in particular, a natural case that could be made in its favour), and does not offer many details on how the argument is in any event supposed to proceed from the lemma. Williamson's scarce remarks to this effect do suggest however that he is envisaging the argument to be continued along the lines of the (MP)-argument (see e.g. Williamson [2000], pp. 169, 181). Be that as it may, Williamson's rejection of the same-evidence lemma relies on a view of evidence which I argue in Zardini [2013b] not to offer the materials for a general answer to the sceptical underdetermination argument. This paper in effect grants the same-evidence lemma and argues that several prominent versions of the ensuing sceptical underdetermination argument break down anyways at some of the following steps. Thanks to an anonymous referee for pressing me on these issues.

[27] 'Very roughly' because the proposed interpretation of probability on the evidence is arguably in tension with some distinctive principles of classical probability (see for example Zardini [2013a]). As such tension will however not be immediately relevant for our issues, for simplicity I'll conduct our discussion keeping fixed both the proposed interpretation of probability on the evidence and the classical theory of probability (see also fn 28). Thanks to Dylan Dodd for urging this clarification.

existence of the evidence (MP), one might try to take as a premise the claim of positive probability on the evidence:

(PP) The probability of $DEMON$ on f is positive

(let's call this version of the sceptical underdetermination argument 'the (PP)-argument' and let's make corresponding stipulations). However, since the memory example used above is an example where the relevant hypothesis is naturally interpreted as having positive probability on the evidence, the inference from (PP) to ($\mathrm{NF^F}$) is no less intuitively fallacious than the inference from (MP) to ($\mathrm{NF^F}$). Moreover, we can now argue that these two routes for getting to ($\mathrm{NF^F}$) are probabilistically blocked, in the sense that there are probabilistic countermodels both to the inference from (MP) to ($\mathrm{NF^F}$) and to the inference from (PP) to ($\mathrm{NF^F}$), and this in turn in the sense that there are probabilistic models of (MP) and (PP) which are also models where $FIRE$ is supported by f but $DEMON$ is not, and so models where $FIRE$ is favoured by f over $DEMON$ (see fn 23).[28]

We can start by very easily showing that there are probabilistic countermodels to the inference from (MP) to ($\mathrm{NF^F}$):

Theorem 1. *There is a sequence of probability spaces $\mathcal{S}^i$ where $H \wedge H^* = \varnothing$ and $H^* \wedge E \neq \varnothing$, but $\mathsf{Pr}^{\mathcal{S}^i}(H|E)$ is arbitrarily high (and can in fact be 1).*

Proof. Consider the sequence of probability spaces $\mathcal{S}^i$ $[i : i \geq 3]$ such that $W^{\mathcal{S}^i} = \{w_j : 1 \leq j \leq i\} = E$ and, for every j $[j : 2 \leq j \leq i]$, $\mathsf{Pr}^{\mathcal{S}^i}(w_j) = 1/i-1$, with $H = \{w_j : 2 \leq j \leq i-1\}$ and $H^* = \{w_1\}$ (set $H = \{w_j : 2 \leq j \leq i\}$ instead to get $\mathsf{Pr}^{\mathcal{S}^i}(H|E) = 1$).

□

We can also very easily show that there are probabilistic countermodels to the inference from (PP) to ($\mathrm{NF^F}$):

Theorem 2. *There is a sequence of probability spaces $\mathcal{S}^i$ where $H \wedge H^* = \varnothing$ and $\mathsf{Pr}^{\mathcal{S}^i}(H^*|E) = \mathfrak{r} > 0$, but $\mathsf{Pr}^{\mathcal{S}^i}(H|E)$ is arbitrarily high below $1 - \mathfrak{r}$ (and can in fact be $1 - \mathfrak{r}$).*

Proof. Consider the sequence of probability spaces $\mathcal{S}^i$ $[i : i \geq 3]$ such that $W^{\mathcal{S}^i} = \{w_j : 1 \leq j \leq i\} = E$, $\mathsf{Pr}^{\mathcal{S}^i}(w_1) = \mathfrak{r}$ and, for every j $[j : 2 \leq j \leq i]$, $\mathsf{Pr}^{\mathcal{S}^i}(w_j) = 1 - \mathfrak{r}/i - 1$, with $H = \{w_j : 2 \leq j \leq i - 1\}$ and $H^* = \{w_1\}$ (set $H = \{w_j : 2 \leq j \leq i\}$ instead to get $\mathsf{Pr}^{\mathcal{S}^i}(H|E) = 1 - \mathfrak{r}$).

□

[28] Although, for simplicity, I'm assuming the standard framework of *classical* probability, I should mention that all the facts to follow also hold, possibly *mutatis mutandis*, in all the main *non-classical* probabilistic frameworks.

Both inferences are thus not only intuitively fallacious, but also probabilistically fallacious.[29] As for the probabilistic fallacy, the (MP)-sceptic or (PP)-sceptic could deny the relevance for her argument of probabilistic considerations. That would strike me as a bad move on the part of the internal sceptic. I'm only mentioning it to set it aside. More plausibly, while accepting the relevance of probabilistic considerations, the (MP)-sceptic or (PP)-sceptic could argue that the underdetermination scenario exhibits additional features that are lacking from our models, so that theorems 1 and 2 fail to apply to her argument (in the sense that the kinds of models on which these theorems rely are ruled out because of their failure to exhibit such additional features).[30] This idea is certainly interesting and worth pursuing further, but the burden is now clearly on the (MP)-sceptic or (PP)-sceptic to discharge her argumentative obligations: in order to rescue an *instance* of a *form* of argument which, in its generality, has been acknowledged by all parties to be fallacious, the (MP)-sceptic or (PP)-sceptic faces the daunting task of establishing that the admissible probabilistic models do not include any of the models on which theorems 1 and 2 rely—that is, that the admissible probabilistic models must have very specific and by no means obvious features. That is no longer internal scepticism as we know and love: instead of a *seemingly compelling* argument with seemingly compelling premises leading to a seemingly repugnant sceptical conclusion, what we have is an *embarrassingly fallacious* argument (the one from (MP) or (PP) to (NF^{F})), which could only acquire some compellingness after being patched up with the heroic (and hitherto unattempted) establishment of utterly unobvious probabilistic assumptions (analogous points about the role of probabilistic models will apply to the other versions of the sceptical underdetermination argument considered below).[31]

[29] Of course, theorems 1 and 2, along with the other theorems in sections 5 and 6, do not imply that $FIRE$ is favoured by f over $DEMON$ or similar claims. But that's alright, as recall that my aim is not to establish that or similar claims, but only to argue that the several versions of the sceptical underdetermination argument against them are probabilistically fallacious, and for this it is sufficient to show that there are probabilistic models where the premises of the relevant argument are true but the conclusion is false.

[30] A natural specific version of this idea would have the (MP)-sceptic or (PP)-sceptic arguing that the only mathematical probability functions that represent genuine possible measures of the support that the evidence gives to a hypothesis must satisfy additional constraints that are not satisfied by the probability functions of our models, so that theorems 1 and 2 fail to apply to her argument (in the sense that the kinds of models on which these theorems rely are ruled out because their probability functions fail to satisfy such additional constraints). Thanks to an anonymous referee for pushing this version of the idea.

[31] For example, the (MP)-sceptic could rule out the kinds of probabilistic countermodels on which theorems 1 and 2 rely if she could establish that the only admissible probabilistic models are such that, if H^* is incompatible with $FIRE$ but its truth is metaphysically compossible with f's existence, then the conditional probability of H^* on F is not lower than the conditional probability of $FIRE$ on F. But that is an utterly unobvious claim (of course, if at least certain kinds of scepticism about perceptual justification are right, the claim does follow, but that is hardly relevant, since the claim is supposed to serve in an argument aimed at establishing scepticism in the first place). In fact, the more general claim got by replacing F and $FIRE$ with any evidence E and hypothesis H is certainly false (as shown by the various examples introduced above in the text) and even easily refutable (letting $D = \llbracket$A 6-sided die has been cast$\rrbracket$ and $L_i = \llbracket$The die has landed on $i\rrbracket$, the general claim entails that, for every relevant admissible probability space $\mathcal{A}$, $\mathsf{Pr}^{\mathcal{A}}(L_1|D) = \mathsf{Pr}^{\mathcal{A}}(L_2|D) = \mathsf{Pr}^{\mathcal{A}}(L_3|D) = \mathsf{Pr}^{\mathcal{A}}(L_1 \vee L_2|D)$, which is inconsistent with the principles of classical probability), and it is very unclear why the specific instance in which the evidence is F and the hypothesis is $FIRE$ should be any more plausible. Thanks to an anonymous referee for prompting me to think about this option.

5 Counterfactual Implication, Favouring and Justification

The (MP)-argument thus fails dramatically. There is however an important and interesting rejoinder to the failure of the (MP)-argument that the underdetermination sceptic has available. Although such rejoinder will turn out eventually not to improve much the compellingness of the sceptical underdetermination argument, it will pay to investigate it in-depth. The rejoinder generally focuses on the fact that the metaphysically modal relations between $DEMON$'s truth and f's existence are much *stronger* than what is adverted to in (MP), with different versions of the rejoinder differing in exactly which stronger relation is focussed on. The first version of the rejoinder, to be discussed in this section, focuses on the observation that not only is $DEMON$'s truth *metaphysically compossible* with f's existence, it is in fact the case that, *were* $DEMON$ true, f *would* exist (in common parlance, $DEMON$'s truth *counterfactually implies* f's existence). I'll argue however that, although it thus avails itself of a stronger (and correct) assumption about metaphysical modality, perhaps surprisingly the resulting argument doesn't fare much better than the (MP)-argument, and indeed fails on similar grounds.

In order to formulate the first version of the rejoinder to the failure of the (MP)-argument, we first need to make an important modification to the (MP)-argument. The modification appeals to the principle that *the justification constituted by the evidence is closed under known logical entailment*:

(CEJ) If evidence e justifies one for believing that P_0, that P_1, that $P_2\ldots$, and one knows that $[\![P_0]\!]$, $[\![P_1]\!]$, $[\![P_2]\!]\ldots$ entail $[\![Q]\!]$, then e justifies one for believing that Q.

(CEJ) (as well as its cousin (CEAJ) to be discussed in section 6) would require a lot of qualifications to stand a chance to avoid the many problems of principle and of detail raised for it in the vast contemporary epistemological literature on the topic, but I'll have to omit all discussion of these and rest content with the assumption that, in the underdetermination scenario, we can safely ignore the need for any such qualification.

With (CEJ) in place, and assuming that René knows that $FIRE$ entails $\neg DEMON$, f's justifying René for believing $FIRE$ entails that f justifies René for believing $\neg DEMON$. However, just as the (MP)-argument asked whether f favours $FIRE$ over $DEMON$, the version in question of the sceptical underdetermination argument asks whether f favours $\neg DEMON$ over $DEMON$. And with $\neg DEMON$ now in place of $FIRE$, it would seem that the underdetermination sceptic could argue as follows. Let evidence e be *sensitive* to a proposition $[\![P]\!]$ iff, were it not the case that P, e would not exist.[32] Given the observation made above that $DEMON$'s truth counterfactually implies f's existence (and given the metaphysical possibility of $DEMON$'s truth), it follows by standard counterfactual reasoning that:

[32]Dretske [1971] introduces a very similar notion. I concur with Dretske in focussing on sensitivity as a property of *evidence* (or *reasons*) rather than *beliefs*, contrary to what many authors do.

(NS) f is not sensitive to $\neg DEMON$

holds.

(NS) has suggested to some epistemologists that:

($\mathrm{NF^N}$) f does not favour $\neg DEMON$ over $DEMON$

also holds.[33] With ($\mathrm{NF^N}$) in place, the new argument can then proceed in a way similar to the (MP)-argument. By ($\mathrm{NF^N}$) and (JNU), f does not justify René for believing $\neg DEMON$, and so, by (CEJ), f does not justify René for believing $FIRE$. But f is René's evidence, and so René's evidence does not justify him for believing $FIRE$. It then follows by (JE) that René is not justified for believing $FIRE$ (let's call this version of the sceptical underdetermination argument 'the (NS)-argument' and let's make corresponding stipulations).[34]

(NS) is certainly a better basis for arguing for ($\mathrm{NF^N}$) than (MP) or (PP) are for arguing for ($\mathrm{NF^F}$): at the very least, the inference from (NS) to ($\mathrm{NF^N}$)—and, more generally and *modulo* (JNU), the requirement that evidence justifying one for believing a proposition be sensitive to that proposition—avoids some of the most glaring counterexamples to which the inference from (MP) or (PP) to ($\mathrm{NF^F}$) is subject. Having said that, I happen to be in agreement with a large part of the contemporary epistemological literature that the inference from (NS) to ($\mathrm{NF^N}$) is still fallacious, and, more generally, that the requirement that evidence justifying one for believing a proposition be sensitive to that proposition is

[33]Let evidence e be *susceptible* to a proposition $[\![P]\!]$ iff it is not the case that, were it not the case that P, e would exist. Then, $DEMON$'s truth counterfactually implying f's existence is tantamount to f's not being susceptible to $\neg DEMON$. And, given the metaphysical possibility of $DEMON$'s truth, it follows by very plausible counterfactual reasoning that f's non-susceptibility to $\neg DEMON$ is even stronger than f's non-sensitivity to $\neg DEMON$. Although, for reasons of continuity with some prominent strand in the extant literature, I focus on f's non-sensitivity to $\neg DEMON$, I should note that all the points to be made about it apply with equal force to f's non-susceptibility to $\neg DEMON$.

[34]While casually mentioning counterfactual implications when considering grounds for a claim of lack of favouring, Brueckner [1994]; [2005]; [2010] constantly settles for metaphysical compossibility as the metaphysically modal relation constituting the official ground for the claim of lack of favouring, and thus overlooks natural versions of the sceptical underdetermination argument—such as the (NS)-argument—that are not immediately exposed to the fallibilist worries he expresses about the (MP)-argument (see fn 25). Cohen [1998] does consider arguments that, on his favoured interpretation of how their premises are grounded, focus instead on counterfactual implication, where the implication is however required to be from $DEMON$'s truth not merely to f's existence, but to $DEMON$'s truth *explaining* f's existence. I agree with Cohen that such explanatory power gives more strength to the underdetermination sceptic's intuition pump and note that all the arguments in this paper are compatible with this additional constraint on the choice of $DEMON$. Schiffer [2004] considers an argument very similar to the one discussed in fn 13, with the modification that the crucial assumption is rather, roughly, that, in order to be justified for believing $FIRE$ on the basis of his perceptual engagement with the fire in the room, René needs to have an *independent* justification for rejecting the hypotheses incompatible with $FIRE$ whose truth counterfactually implies the rest of f. Schiffer's argument seems reducible to the (NS)-argument (or, more accurately, to something along the lines of the simpler argument based on counterfactual implication mentioned in fn 37) just like the argument discussed in fn 13 seemed reducible to the (MP)-argument.

spurious (although, unfortunately, these points are not frequently made in relation to the sceptical underdetermination argument). However, before briefly turning to my reasons for so thinking, I'd like to mention and dismiss a common worry about the (NS)-argument.

Until the end of this section, let's take (JNU) for granted, and so let's take for granted that the validity of the transition from insensitivity to lack of favouring implies in effect the requirement that evidence justifying one for believing a proposition be sensitive to that proposition. The worry referred to at the end of the last paragraph is then that requiring the evidence to be sensitive is in conflict with requiring the justification it constitutes to be closed under known entailment as (CEJ) would have it (see Brueckner [1994], p. 828; [2005], p. 388; [2010], p. 368).[35] This is so, I take it, because many ordinary bodies of evidence like f are sensitive to "*light-weight*" propositions like $FIRE$, but, as we've seen, are not sensitive to "*heavy-weight*" propositions like $\neg DEMON$, in spite of the fact that the latter are known to be entailed by the former.

But, keeping fixed the sensitivity requirement, is that a good reason to reject (CEJ)? Clearly, it is a good reason only if sensitivity is not simply a *necessary* condition for justification (so that the failure of the evidence to be sensitive to a heavy-weight proposition implies that the evidence does not justify one for believing the heavy-weight proposition, hence falsifying the consequent of (CEJ)), but also—possibly together with other conditions uncontroversially met by typical beliefs in light-weight propositions—a *sufficient* condition for justification (so that the sensitivity of the evidence to a light-weight proposition implies that the evidence justifies one for believing the light-weight proposition, hence verifying the antecedent of (CEJ) and thus—given the falsification of the consequent of (CEJ) on the grounds just mentioned—falsifying (CEJ) itself). That was in effect the view of at least one prominent defender of sensitivity (see Nozick [1981], pp. 167–288), at least as far as knowledge is concerned (see fn 35; notice also that Nozick works with the belief version of sensitivity, see fn 32). But that is a commitment that the underdetermination sceptic could fairly reject: to think that sensitivity is a necessary condition for justification by no means commits one to thinking that it is also a sufficient condition for justification, so that no conflict is involved in accepting the former and rejecting the latter (indeed, it is easier to gather reasons for why sensitivity is not a sufficient condition for justification than it is for why it is not a necessary condition), and so that no conflict is involved in accepting both sensitivity (as a necessary condition on justification) and (CEJ). In sum, both sensitivity (as a necessary condition on justification) and (CEJ) enjoy a certain *prima facie* plausibility. They are jointly consistent, and, as we've seen, jointly entail scepticism (*modulo* (JE)). That's a perfectly decent sceptical argument—if it is unsound, it is not because its two key premises are in conflict with one another.

The problem with the (NS)-argument is thus not the conflict between its two key premises, sensitivity (as a necessary condition on justification) and (CEJ), which in fact does not subsist. In my view, it rather lies entirely with the first such premise, sensitivity (as a necessary condition on justification). On reflection, as has been pointed out by many authors, it is just not true that evidence justifying one for believing a proposition has to be

[35]The worry is usually cast in terms of knowledge rather than justification, but its structure clearly allows for it to be replicated in terms of justification in relation to the (NS)-argument.

sensitive to that proposition.[36] For example, my evidence containing my seeming memory of having thought about this paper for a while clearly justifies me for believing that I can think about something for a while, even if, in the extremely far-fetched circumstance in which my psychological functions were so pathologically affected as to prevent prolonged thinking about any issue, they would also try to hide this from personal-level awareness by producing false memories to the contrary (in particular, the memory of having thought about this paper for a while), and so even if my evidence containing my seeming memory of having thought about this paper for a while is not sensitive to ⟦I can think about something for a while⟧. Moreover, my evidence containing my seeming memory of having thought about this paper for a while clearly favours ⟦I can think about something for a while⟧ over ⟦I cannot think about something for a while⟧, so that the problem with the requirement that evidence justifying one for believing a proposition be sensitive to that proposition is shown really to lie with the transition from insensitivity to lack of favouring rather than with the transition, licensed by (JNU), from lack of favouring to lack of justification. Notice that this last point can also be argued from the stronger observation that my evidence containing my seeming memory of having thought about this paper for a while clearly supports ⟦I can think about something for a while⟧ and does not support ⟦I cannot think about something for a while⟧ (see fn 23).

Again, the inference from (NS) to ($\mathrm{NF^N}$) is not only intuitively fallacious, it is also probabilistically fallacious. To see this, we need to enrich somewhat our probability spaces in order to be able to model counterfactual implications in them:

Definition 2. A *counterfactual probability space* $\mathcal{S}$ is a triple $\langle W^{\mathcal{S}}, C^{\mathcal{S}}, \mathsf{Pr}^{\mathcal{S}}\rangle$ where $W^{\mathcal{S}}$ and $\mathsf{Pr}^{\mathcal{S}}$ are as in definition 1 and $C^{\mathcal{S}} : W^{\mathcal{S}} \times \wp(W^{\mathcal{S}}) \mapsto \wp(W^{\mathcal{S}})$ is such that $C^{\mathcal{S}}(w, X) \subseteq X$.

Definition 3. Given a counterfactual probability space $\mathcal{S}$, a *counterfactual implication* $X > Y$ is the operation on pairs of propositions having as value the proposition $\{w : C^{\mathcal{S}}(w, X) \subseteq Y\}$.

(This is a gross oversimplification of the best semantics for counterfactuals on the market, but will do well enough for our purposes.) We can now show:

Theorem 3. *There is a sequence of counterfactual probability spaces $\mathcal{S}^i$ where $\neg H > E = W^{\mathcal{S}^i}$, but $\mathsf{Pr}^{\mathcal{S}^i}(H|E)$ is arbitrarily high (and can in fact be 1).*

[36] Vogel [1986] is an early reference in which several kinds of problematic consequences of sensitivity, at least as far as knowledge is concerned, are pointed out, but, contrary to the problematic consequence to be presented in the text, none applies to such a bottom-level epistemic practice as relying on one's memory (notice also that Vogel [1986] works with the belief version of sensitivity, see fn 32). A few authors (such as e.g. Sosa [1999]) accept the thrust of such counterexamples to sensitivity and advocate the different constraint of *safety*, where, for our purposes, evidence e is safe for a proposition ⟦P⟧ iff it could not easily have been the case that [it is not the case that P and e exists]. Setting aside whether unsafety induces lack of favouring, it will suffice to note that, contrary to the other metaphysically modal relations discussed in this paper, it can straightforwardly be argued that, given the way the non-epistemic world is, f is not unsafe for $\neg DEMON$ (nor for $FIRE$). This last point is related to the one made in section 3 (although it should be noted that, generally, e nomologically necessarily (relative to the present time) implying that P is two-way independent from e being safe for ⟦P⟧).

Proof. Consider the sequence of counterfactual probability spaces $\mathcal{S}^i$ $[i : i \geq 3]$ such that $W^{\mathcal{S}^i} = \{w_j : 1 \leq j \leq i\} = E$, $C^{\mathcal{S}^i}(w, \neg H) = \{w_1\}$ and, for every j $[j : 2 \leq j \leq i]$, $\mathsf{Pr}^{\mathcal{S}^i}(w_j) = 1/i - 1$, with $H = \{w_j : 2 \leq j \leq i - 1\}$ (set $H = \{w_j : 2 \leq j \leq i\}$ instead to get $\mathsf{Pr}^{\mathcal{S}^i}(H|E) = 1$).

□

6 Entailment, Favouring and Justification

The (NS)-argument thus fails dramatically. That argument was the first version of the underdetermination sceptic's rejoinder to the failure of the (MP)-argument and focussed on the observation that, were $DEMON$ true, f would exist. In fact, a second and even stronger observation is available, on which a second and even stronger version of the rejoinder focuses. To wit, not only does $DEMON$'s truth *counterfactually imply* f's existence, it in fact *entails* it. I'll argue however that, although it thus avails itself of a stronger (and correct) assumption about metaphysical modality, perhaps surprisingly the resulting argument doesn't fare much better than the (MP)-argument or the (NS)-argument, and indeed fails on similar grounds.

Can the evidence really disfavour a hypothesis whose truth is not merely metaphysically compossible with, or does not merely counterfactually imply, but does so much as entail the existence of the evidence? It is not clear what the answer to this question should be, but by adding a couple of plausible extra assumptions it *would seem* that one could argue that the answer should be negative.[37] These assumptions are that the unconditional probability of the existence of f is lower than maximum and that the unconditional probability of $DEMON$ is higher than minimum (notice that this latter assumption would be entailed by (PP)). With these assumptions in place, we can turn to a well-known theorem of classical probability:

[37]There is admittedly some pull towards a negative answer, so a simpler version of the sceptical underdetermination argument could *directly* infer (NF^{F}) from the simple fact that $DEMON$ entails F. Notice that an analogous simpler version *directly* inferring (NF^{F}) from the simple fact that $DEMON$ counterfactually implies F was available in the case of the (NS)-argument (among other arguments, Huemer [2001] also offers and discusses what is in essence such version). Both these simpler versions could be thought of as appealing to the general intuition that E cannot favour H over H^* if H^* in some relevant sense genuinely *implies* E. That is an intuition that does have some pull, but it'd be nice to set it on at least *prima facie* firmer ground (given, among other things, that the analogous intuition in favour of directly inferring (NF^{F}) from (MP) has proven so unreliable). Shifting the focus from $FIRE$ to $\neg DEMON$ (courtesy of (CEJ)), I have tried to do so in section 5 by appealing to facts concerning sensitivity (which are still of a metaphysically modal nature and which follow from the fact concerning counterfactual implication), and will try to do so in this section by appealing to facts concerning probability raising and lowering (which are of a probabilistic nature and which follow from the fact concerning entailment). Be that as it may, the examples and models that I have offered and will offer against those more sophisticated versions of the sceptical underdetermination argument also tell against the simpler versions just mentioned (since they are examples and models that not only verify the facts about sensitivity and probability raising and lowering but also the facts about counterfactual implication and entailment). Thanks to an anonymous referee for urging me to consider explicitly these simpler versions of the sceptical underdetermination argument.

Theorem 4. *Given a probability space $\mathcal{S}$, if:*

(i) H^* *entails* E*;*

(ii) $\mathsf{Pr}^{\mathcal{S}}(E) < 1$*;*

(iii) $\mathsf{Pr}^{\mathcal{S}}(H^*) > 0$*,*

then $\mathsf{Pr}^{\mathcal{S}}(H^*|E) > \mathsf{Pr}^{\mathcal{S}}(H^*)$*.*

Corollary 1. *Given a probability space $\mathcal{S}$ satisfying conditions (i)–(iii),* $\mathsf{Pr}^{\mathcal{S}}(\neg H^*|E) < \mathsf{Pr}^{\mathcal{S}}(\neg H^*)$*.*

By stipulation, $DEMON$ and F satisfy condition (i). Moreover, we've plausibly just assumed that F satisfies condition (ii) and that $DEMON$ satisfies condition (iii), so that theorem 4 and corollary 1 apply to $DEMON$, F and $\neg DEMON$. It follows that:

(HL) The conditional probability of $DEMON$ on F is *higher* than its unconditional probability; the conditional probability of $\neg DEMON$ on F is *lower* than its unconditional probability

holds.

(HL) may suggest to some epistemologists that (NF^{F}) holds (from which a new sceptical underdetermination argument along the lines of the (MP)-argument follows; let's call this version of the sceptical underdetermination argument 'the (HL^{F})-argument' and let's make corresponding stipulations), and, with even more apparent compellingness, that (NF^{N}) holds (from which a new sceptical underdetermination argument along the lines of the (NS)-argument follows; let's call this version of the sceptical underdetermination argument 'the (HL^{N})-argument' and let's make corresponding stipulations).[38]

[38]Versions of the sceptical underdetermination argument based on (HL) have not been discussed in the literature mentioned in fn 4. In fact, with the notable exception of Huemer [2001] (who, among other arguments, also offers and discusses what is in essence the (HL^{F})-argument), originally (HL) was mostly employed in debates about basic justification and "Moorean" arguments (Hawthorne [2004], pp. 73–77 is an early reference). This has then given rise to the employment of (HL) in formulating sceptical arguments (in terms of justification or knowledge) along the broad lines of arguments discussed in this section (Silins [2005], pp. 395–398; Weatherson [2007]; Schiffer [2009] are cases in point, although they all employ (HL) only in the construction of a subargument aimed at showing that *experience* does not justify René for believing $\neg DEMON$, while my notion of evidence also comprises whatever *non-empirical* considerations may be available to him). I won't try to determine in this paper how the arguments those authors discuss are best construed in detail given the distinctions drawn in this section. I wish to note, however, that, whereas Weatherson [2007] explores a way of resisting his target argument which consists in the rejection of the Bayesian assumption that the evidence's effect on probability is given by the prior conditional probability on the evidence, my own critical discussion is consistent with that assumption and problematises instead the other assumption implicitly at work in these arguments to the effect that the *probabilistic* fact stated in (HL) has the intended consequences at the *probabilistic* level of favouring and at the *epistemic* level of justification.

At this juncture of the dialectic, and as warned in section 2, it really becomes crucial how we understand the epistemologically non-primitive notion of favouring. Up to now, we've been assuming as a *sufficient* condition for evidence e favouring hypothesis H over an incompatible hypothesis H^* that e supports H and does not support H^* (see fn 23). That would be however too strong a *necessary* condition: favouring can occur even when e (non-comparatively) supports neither H nor H^* and even when e (non-comparatively) supports both H and H^*, for, in both kinds of cases, e may still (comparatively) support H more than it supports H^* (see fn 21). While favouring is thus not reducible to any obvious logical construction out of *support simpliciter*, I think that it can in fact be understood in terms of *degrees of support*. And since we're understanding probability exactly as a measure of the degree of the support that the evidence gives to a hypothesis, I think that, for our purposes, favouring can be understood in terms of probability.

Indeed, in a probabilistic framework, more than one interesting relation can be defined that could reasonably be labelled 'favouring'. Up to now, I've in fact been implicitly assuming the natural, *static* and *absolute* understanding of favouring:

(ABS) e favours H over H^* iff $\mathsf{Pr}(H|E) > \mathsf{Pr}(H^*|E)$.

However, on the (ABS)-understanding, even given (HL), ($\mathrm{NF^F}$) and ($\mathrm{NF^N}$) still do not follow. For example, my evidence containing my seeming memory of having thought about this paper for a while clearly (ABS)-favours ⟦I have thought about this paper for a while⟧ over ⟦I have never thought about this paper before and, moreover, I have had a severe indigestion causing a seeming memory that I have thought about this paper for a while⟧, and clearly (ABS)-favours ⟦It is not the case that [I have never thought about this paper before and, moreover, I have had a severe indigestion causing a seeming memory that I have thought about this paper for a while]⟧ over ⟦I have never thought about this paper before and, moreover, I have had a severe indigestion causing a seeming memory that I have thought about this paper for a while⟧.

Again, the inference from (HL) to (ABS)-($\mathrm{NF^F}$)[39] or (ABS)-($\mathrm{NF^N}$) is not only intuitively fallacious, it is also formally fallacious:

Theorem 5. *There is a sequence of probability spaces $\mathcal{S}^i$ where conditions (i)–(iii) of theorem 4 hold, but E (ABS)-favours H over H^* and $\neg H^*$ over H^* to an arbitrarily high degree (in the sense that $\mathsf{Pr}^{\mathcal{S}^i}(H|E) - \mathsf{Pr}^{\mathcal{S}^i}(H^*|E)$ and $\mathsf{Pr}^{\mathcal{S}^i}(\neg H^*|E) - \mathsf{Pr}^{\mathcal{S}^i}(H^*|E)$ are arbitrarily high).*

Proof. Consider the sequence of probability spaces $\mathcal{S}^i$ $[i : i \geq 2]$ such that $W^{\mathcal{S}^i} = \{w_j : 1 \leq j \leq i\}$ and, for every j $[j : 1 \leq j \leq i]$, $\mathsf{Pr}^{\mathcal{S}^i}(w_j) = 1/i$, with $E = \{w_j : 1 \leq j \leq i-1\}$, $H = \{w_j : 2 \leq j \leq i\}$ and $H^* = \{w_1\}$ (let H^* be defined as $E \wedge \neg H$ to have a formal entailment from H^* to E).

□

[39] Given that two distinct notions of favouring are introduced in this section, principles and arguments employing 'favour'-talk need to be correspondingly disambiguated. I'll do so by prefixing their names with the label of the relevant notion of favouring.

It would thus seem that in order to be able to put theorem 4 to do sceptical underdetermination work, we should assume a more *dynamic* and *relative* understanding of favouring:

(REL) e favours H over H^* iff $\mathsf{Pr}(H|E) - \mathsf{Pr}(H) > \mathsf{Pr}(H^*|E) - \mathsf{Pr}(H^*)$

(or something along these lines).[40] However, even on the (REL)-understanding, even given (HL), ($\mathrm{NF^F}$) still does not follow. For example, my evidence containing my seeming memory of having thought about this paper for a while clearly (REL)-favours ⟦I have thought about this paper for a while⟧ over ⟦I have never thought about this paper before and, moreover, I have had a severe indigestion causing a seeming memory that I have thought about this paper for a while⟧.

Again, the inference from (HL) to (REL)-($\mathrm{NF^F}$) is not only intuitively fallacious, it is also formally fallacious:

Theorem 6. *There is a sequence of probability spaces $\mathcal{S}^i$ where conditions (i)–(iii) of theorem 4 hold, but E (REL)-favours H over H^* to an arbitrarily high degree (in the sense that $(\mathsf{Pr}^{\mathcal{S}^i}(H|E) - \mathsf{Pr}^{\mathcal{S}^i}(H)) - (\mathsf{Pr}^{\mathcal{S}^i}(H^*|E) - \mathsf{Pr}^{\mathcal{S}^i}(H^*))$ is arbitrarily high).*

Proof. Consider the sequence of probability spaces $\mathcal{S}^i$ $[i : i \geq 4]$ such that $W^{\mathcal{S}^i} = \{w_j : 1 \leq j \leq i\}$ and, for every j $[j : 1 \leq j \leq i]$, $\mathsf{Pr}^{\mathcal{S}^i}(w_j) = 1/i$, with $E = \{w_j : j \text{ is a Fibonacci number}\}$, $H = E \setminus \{w_1\}$ and $H^* = \{w_1\}$ (let H^* be defined as $E \wedge \neg H$ to have a formal entailment from H^* to E).

□

What does follow is of course (REL)-($\mathrm{NF^N}$), which is immediate given theorem 4 and corollary 1. And with (REL)-($\mathrm{NF^N}$) in place, it would look like we could run the (REL)-($\mathrm{HL^N}$)-argument. Couldn't we?

Not so quick. To appreciate the problem with that line of thought, we should move on to inspect in greater detail the sceptical underdetermination argument's third step consisting in the inference from lack of favouring to lack of justification. Now, no matter how things may stand with other understandings of (JNU) (for example, as (ABS)-(JNU)), (REL)-(JNU) is arguably false. To see this, consider the two properties related by (JNU): favouring and justification. On the one hand, for our purposes, favouring can be understood as a probabilistic notion. On the other hand, while I don't think that justification is a probabilistic notion, I think that it is safe to assume that, insofar as it is connected with probabilistic support (for example, in the sense of the latter being a necessary condition for the former), justification for believing a proposition is connected with the proposition's probability on the evidence meeting a certain (possibly only contextually determined, possibly

[40] For example, substituting (straightforward) *division* for (straightforward) *subtraction*. I'll henceforth assume the particular way of spelling out the dynamic and relative understanding of favouring which uses straightforward subtraction, but the substance of my discussion will apply just as well to other natural candidates for construing the dynamic and relative understanding of favouring.

vague) *threshold*. And that last property is in turn much more related to (ABS)-favouring than to (REL)-favouring.

The last statement was deliberately vague and the precise sense in which it is true will depend on the precise relationship postulated in one's theory between justification and probabilistic support. For example, if the postulated relationship is the (very plausible) one consisting in the fact that probabilistic support is necessary for justification, and if the operative threshold for probabilistic support $\mathfrak{t}$ is (again, very plausibly) $> .5$, (ABS)-favouring is also going to be necessary for justification, while (REL)-favouring is not. For suppose that evidence e justifies one for believing a hypothesis H. Then, since probabilistic support is necessary for justification, E[41] probabilistically supports H. And, since probability $> .5$ is necessary for probabilistic support, $\mathsf{Pr}(H|E) > .5$. If H^* is incompatible with H, $\mathsf{Pr}(H^*|E) < .5 < \mathsf{Pr}(H|E)$, and so e (ABS)-favours H over incompatible hypotheses, while it may still not (REL)-favour it. More generally, upon getting the final part of e, it seems that one's new body of evidence can *still on the whole* constitute a justification for believing $\neg H^*$ (if $\mathsf{Pr}(\neg H^*|E)$ is sufficiently high), even though it may be a slightly weaker justification than the one one had before getting the final part of e (because $\mathsf{Pr}(\neg H^*|E) < \mathsf{Pr}(\neg H^*)$). So it seems that (JNU) itself is only plausible when favouring is understood as (ABS)-favouring rather than (REL)-favouring. But, as we've seen, on that understanding (NF^{N}) hasn't yet been vindicated. In sum, we have (REL)-(NF^{N}), but we do not have (REL)-(JNU) (and thus we cannot run the (REL)-(HL^{N})-argument); we have—we may grant—(ABS)-(JNU), but we do not have (ABS)-(NF^{N}) (and thus we cannot run the (ABS)-(HL^{N})-argument): either way, the inference to f's not justifying René for believing $\neg DEMON$ would be a fallacy of equivocation on 'favour'.

I'll come back to the (REL)-(HL^{N})-argument at the end of this section, but let's first consider another line of thought that has been suggested to some epistemologists by (HL) and that might be prompted by the last paragraph. Notice that it's true that, because $\mathsf{Pr}(\neg DEMON|F) < \mathsf{Pr}(\neg DEMON)$, it still seems odd to think that, upon getting the final part of f, René can, on that basis, *learn* $\neg DEMON$ (or *acquire* a justification for believing it). In order to put that intuition (together with corollary 1) to do sceptical underdetermination work, we would however need principles similar to but substantially different from (CEJ) and (JNU), and consequently run an argument similar to but substantially different from the arguments considered so far.

Firstly, (CEJ) should be replaced by the principle that *the acquisition of justification constituted by the evidence is closed under known logical entailment*:

(CEAJ) If one has acquired a justification for believing that P_1, that P_2, that P_3... constituted by evidence e, and one knows that $[\![P_1]\!]$, $[\![P_2]\!]$, $[\![P_3]\!]$... entail $[\![Q]\!]$, then one has acquired a justification for believing that Q constituted by e.

The contemporary epistemological literature is used to calling a principle like (CEAJ) a *transmission* principle: while a *closure* principle states a sufficient condition for *having* a

[41] Henceforth, E is a proposition that is a best candidate for representing e.

justification for believing the conclusion in terms of having a justification for believing the premises, a *transmission* principle states a sufficient condition for *acquiring* a justification for believing the conclusion in terms of having a justification for believing the premises—or, as in the case of (CEAJ), in terms of acquiring one such.[42] In turn, the difference between one's *having* and one's *acquiring* a justification constituted by the evidence is that one's having a justification constituted by one's current evidence does not imply that *that justification is a justification one did not have before having that specific evidence* (contrary to one's acquiring a justification constituted by one's current evidence). For example, René has a justification constituted by his current evidence f for believing that 2+3=5, but, upon getting the final part of f (resulting from his perceptual engagement with the fire in the room), he has not acquired any justification for so believing: a previous part of f—involving his arithmetical skills—already provided him with a justification, and that is exactly the same justification provided by f in its entirety (different bodies of evidence can constitute the same justification if they do not differ in the relevant respects).[43]

I should note that, although (CEAJ) is a transmission principle, it is not objectionable on the usual grounds adduced by the concessive anti-sceptical epistemologies introduced in section 2 (see e.g. Wright [2007]), for (CEAJ) concerns the transmission of justification constituted *by evidence*, and hence falls outside the relevant scope of those epistemologies, which hold that evidence in itself hardly ever suffices to constitute justification. But I should also note that this is not to suggest that (CEAJ) in its full generality is acceptable. For example, René knows that ⟦A fire is burning in René's room⟧ entails ⟦René has a room⟧, and, upon getting the final part of f, he has acquired a justification for believing the former proposition constituted by f, but has not acquired any justification for believing the latter proposition (although he does have a justification for believing it constituted by f): a previous part of f—involving his knowledge of the ownership of the room—already provided him with a justification, and that is exactly the same justification provided by f in its entirety. Having noted all that, I should add that, in its specific application to the argument from $FIRE$ to $\neg DEMON$, (CEAJ) retains high plausibility (I defend this claim to some extent in Zardini [2013a]).

Secondly, (JNU) should be replaced by the principle that *acquisition of justification is incompatible with probability lowering*:

(AJIPL) If one has acquired a justification for believing a hypothesis H constituted by evidence e, then $\mathsf{Pr}(H|E) \geq \mathsf{Pr}(H)$.

(AJIPL) is a restatement of the intuition introduced in the third last paragraph. By

[42]Notice that, both in the case of (CEJ) and in the case of (CEAJ), the justification had or acquired for believing the conclusion might crucially rely on one's knowledge of the entailment, which is part of one's evidence.

[43]I should emphasise that, upon getting the final part of evidence e, one can acquire a justification for believing a hypothesis H constituted by e while already having *another* justification for believing H constituted by a previous part of e. For example, upon getting the final part of f, René can acquire a justification for believing $FIRE$ constituted by f while already having another justification for believing $FIRE$ constituted by a previous part of f—involving, say, the testimony of someone that saw the fire in the room. Thanks to an anonymous referee for urging this clarification.

(AJIPL) and corollary 1, René has not acquired a justification for believing $\neg DEMON$ constituted by f, and so, by (CEAJ), René has not acquired a justification for believing $FIRE$ constituted by f. Now, f is René's current evidence and we may assume that, before his perceptual engagement with the fire in the room, René did not have any justification for believing $FIRE$. Hence, if René has not *acquired a justification* for believing $FIRE$ constituted by f, f does not constitute *any justification at all* for believing $FIRE$. It then follows by (JE) that René is not justified for believing $FIRE$ (let's call this version of the sceptical underdetermination argument 'the AJ-argument' and let's make corresponding stipulations).

For all of its undeniable *prima facie* plausibility, the AJ-argument seems to be blocked by the same kind of example used against the (ABS)-(HL^{N})-argument: my evidence containing my seeming memory of having thought about this paper for a while not only clearly (ABS)-favours ⟦It is not the case that [I have never thought about this paper before and, moreover, I have had a severe indigestion causing a seeming memory that I have thought about this paper for a while]⟧ over ⟦I have never thought about this paper before and, moreover, I have had a severe indigestion causing a seeming memory that I have thought about this paper for a while⟧; that evidence also seems to constitute a justification for believing the former proposition, and I seem to have acquired such justification precisely upon getting the seeming memory, *contra* (AJIPL).

Even clearer counterexamples to (AJIPL) emerge once one starts to reflect on cases that demonstrate the *gap* between *probability* and *justification*. I have developed the point at length elsewhere (see Zardini [2013a]), and will simply rest content in this paper with a rough sketch of the structure of the problematic cases (in effect, *cross-temporal* versions of the *cross-modal* cases discussed by Smith [2010]). For example, suppose that there are i papers, in fact so many that I can only have thought about a minuscule number of them and that, before I start reflecting on the matter, no paper is more likely than any other to have been thought about by me. Suppose also that the conditional probability of me having thought about a paper given the seeming memory that I have thought about it for a while is maximum while the conditional probability r of me not having thought about a paper only for one minute given the seeming memory that I have thought about it for a while is very high but not maximum. (Crucially for what follows, the negation in sentences like 'I have not thought about this paper only for one minute' is supposed to take *wide scope*, so that the sentence is understood to be true *not only* if I have thought about this paper for more than one minute, *but also* if I have not thought about this paper at all.) Then, clearly, upon getting the seeming memory that I have thought about this paper for a while, I acquire a justification for believing that I have not thought about it only for one minute. However, given any non-maximum value for r, a corresponding value for i can be chosen so that, upon getting the seeming memory, the probability of ⟦I have not thought about this paper only for one minute⟧ goes down (since it will then be more likely, upon getting the seeming memory, that my fallible memory is misrepresenting precisely the fact that I have thought about this paper only for one minute than it will be likely, before getting the seeming memory, that I have thought about this paper in the first place given the i-many papers there are). (AJIPL) is false.

To conclude, although I've expressed reasons for serious concern about the underdetermination sceptic's employing (REL)-favouring and corresponding principles, let's pursue that track a bit further to see where it actually leads. That track, which represents in effect the last-ditch effort on the part of the underdetermination sceptic in the dialectic of this paper, consists, to repeat, in the (REL)-(HL^{N})-argument. Notice first that we have not only (REL)-(NF^{N}), but also the stronger result that f in fact (REL)-favours $DEMON$ over $\neg DEMON$. But notice also that analogues of both these facts are no longer guaranteed once we switch to anti-sceptical hypotheses *more specific* than $\neg DEMON$ (such as e.g. the hypotheses $F \wedge \neg DEMON$ and $FIRE$): as shown by theorem 6, it is probabilistically consistent that such hypotheses are (REL)-favoured by f over $DEMON$, and indeed that this is so to an arbitrarily high degree.

Given that (CEJ) is not under question (so much so that it is in fact an essential component of the (REL)-(HL^{N})-argument), a new reason emerges for finding (REL)-(JNU) objectionable, reason which is even stronger than the one indicated above. To recall, above I've argued that, given plausible assumptions about the connection between justification and probabilistic support, (ABS)-(JNU) can in fact be vindicated, while (REL)-(JNU) cannot be vindicated in the same way, and that in effect, contrary to (REL)-(JNU), it seems that, upon getting the final part of e, one's new body of evidence can still on the whole constitute a justification for believing $\neg H^*$ in spite of the dynamic and relative fact that $\mathsf{Pr}(\neg H^*|E) < \mathsf{Pr}(\neg H^*)$. What is emerging now is that, whenever E and H^* satisfy conditions (i)–(iii) of theorem 4 and H is incompatible with H^*, given (CEJ) not only is (REL)-(JNU) incompatible with the apparent possibility, upon getting e, of one's body of evidence constituting a justification for believing $\neg H^*$ *in spite of the negative fact* that $\mathsf{Pr}(\neg H^*|E) < \mathsf{Pr}(\neg H^*)$, it is also incompatible with many apparent possibilities, upon getting e, of one's body of evidence constituting a justification for believing H *(partly) thanks to the positive fact* that $\mathsf{Pr}(H|E) > \mathsf{Pr}(H)$! For that would imply, by (CEJ), that one's body of evidence constitutes a justification for believing $\neg H^*$, which contradicts the conjunction of (REL)-(JNU) with the indisputable fact that $\mathsf{Pr}(\neg H^*|E) < \mathsf{Pr}(\neg H^*)$. Thus, given (CEJ), (REL)-(JNU) entails that evidence e may not constitute a justification for believing a hypothesis H which is even (REL)-favoured by e over any incompatible hypotheses just because one knows that H entails something ($\neg H^*$) which has the only flaw [of not being so (REL)-favoured] simply because logically much weaker. Come again?

That seems very poor reasoning by anyone's lights, anti-sceptics and sceptics alike. Moreover, it leads to something well worth calling '*infallibilism*' *across the board*. For, whenever E and H^* satisfy conditions (i)–(iii) of theorem 4 and H is incompatible with H^*, given (CEJ) (REL)-(JNU) entails that, upon getting e, one's evidence does not constitute a justification for believing H, no matter how arbitrarily close $\mathsf{Pr}(H|E)$ gets to 1 (for one knows [that H entails $\neg H^*$] and e does not (REL)-favour $\neg H^*$ over H^*). Setting henceforth aside the point about explanation raised in fn 34,[44] since just about any interesting body of evidence satisfies condition (ii), $E \wedge \neg H$ will be the required H^* unless $\mathsf{Pr}(E \wedge \neg H) = 0$—that is, unless either $\mathsf{Pr}(\neg H|E) = 0$ or $\mathsf{Pr}(E) = 0$. And as the second disjunct of the

[44]This is arguably no real limitation anyways, as very explanatory, if very silly, candidates for being H^* will arguably be available for just about any E and H.

last condition can safely be ignored, that condition can be taken to be equivalent with the condition that $\mathsf{Pr}(H|E) = 1$. Therefore, for just about any interesting body of evidence e, given (CEJ) (REL)-(JNU) entails that, upon getting e, one's evidence can constitute a justification for believing any hypothesis H whatsoever only if $\mathsf{Pr}(H|E) = 1$. That should strike everyone, anti-sceptics and sceptics alike, as an extremely problematic infallibilist consequence which, keeping fixed the highly plausible (CEJ), reflects very badly on (REL)-(JNU) and thus on the (REL)-(HL^{N})-argument.

7 Conclusion

Thus, although they avail themselves of a stronger (and correct) assumption about metaphysical modality, perhaps surprisingly the entirety of the (NS)-argument, (HL^{F})-arguments and (HL^{N})-arguments don't fare much better than the (MP)-argument, and indeed fail on similar grounds. A rather straightforward explanation of this phenomenon is given by the fact that *all these arguments rely on principles and inferences that after all still have infallibilist consequences*. Since, if it is metaphysically possible that e exists and H is false, $E \wedge \neg H$ will be able to play the role of $DEMON$ in the (NS)-argument, the principles and inferences involved in that argument have the infallibilist consequence that e can only constitute a justification for believing H if it is not metaphysically possible that e exists and H is false (just as the (MP)-argument would have it). And since, assuming that $0 < \mathsf{Pr}(E) < 1$, if $\mathsf{Pr}(H|E) < 1$ it follows, essentially by the reason given at the end of section 6, that $E \wedge \neg H$ will be able to play the role of $DEMON$ in the (HL^{F})-arguments and (HL^{N})-arguments, the principles and inferences involved in those arguments have the infallibilist consequence that e can only constitute a justification for believing H if $\mathsf{Pr}(H|E) = 1$ (just as the (PP)-argument would have it).[45]

Having noticed that, I'd like to close by tracing some of the limits of the foregoing discussion. It does not show that the sceptical underdetermination manoeuvres I've been considering are bound to fail—it only shows that the underdetermination sceptic has (much) more work to do in order to support the crucial principles and inferences involved. As we've had occasion to appreciate, there is no straightforward connection between the metaphysically modal facts appealed to by the underdetermination sceptic and the epistemic conclusions she intends to follow from those facts. The epistemic cannot so easily be read off the metaphysically modal. And also the probabilistic facts that are available to the underdetermination sceptic are far from lending any straightforward support to her target conclusions. For all I've said, however, there might be more complicated and sophisticated considerations vindicating the crucial principles and inferences involved in the versions of the sceptical underdetermination argument I've examined in this paper. And, of course, there might be other interesting sceptical arguments relying on the idea that what we believe is in some sense underdetermined by our evidence. There *might* be...

[45] A similar consideration holds for the (AJ)-argument, although it may carry some qualification owing to the fact that (CEAJ) in its full generality is not acceptable.

References

Robert Audi. *Epistemology. A Contemporary Introduction to the Theory of Knowledge*. Routledge, London, 2nd edition, 2003.

John Austin. *Sense and Sensibilia*. Oxford University Press, Oxford, 1962.

Anthony Brueckner. The structure of the skeptical argument. *Philosophy and Phenomenological Research*, 54:827–835, 1994.

Anthony Brueckner. Fallibilism, underdetermination, and skepticism. *Philosophy and Phenomenological Research*, 71:384–391, 2005.

Anthony Brueckner. ∼K∼SK. In *Essays on Skepticism*, pages 367–381. Oxford University Press, Oxford, 2010.

Stewart Cohen. Two kinds of skeptical arguments. *Philosophy and Phenomenological Research*, 58:143–159, 1998.

Fred Dretske. Conclusive reasons. *Australasian Journal of Philosophy*, 49:1–22, 1971.

John Hawthorne. *Knowledge and Lotteries*. Oxford University Press, Oxford, 2004.

Michael Huemer. The problem of defeasible justification. *Erkenntnis*, 54:375–397, 2001.

Robert Nozick. *Philosophical Explanations*. Harvard University Press, Cambridge MA, 1981.

Duncan Pritchard. The structure of skeptical arguments. *The Philosophical Quarterly*, 55: 37–52, 2005.

Stephen Schiffer. Skepticism and the vagaries of justified belief. *Philosophical Studies*, 119: 161–184, 2004.

Stephen Schiffer. Evidence = knowledge: Williamson's solution to skepticism. In Patrick Greenough and Duncan Pritchard, editors, *Williamson on Knowledge*, pages 183–202. Oxford University Press, Oxford, 2009.

Nicholas Silins. Deception and evidence. *Philosophical Perspectives*, 19:375–404, 2005.

Martin Smith. What else justification could be. *Nous*, 44:10–31, 2010.

Ernest Sosa. How to defeat opposition to Moore. *Philosophical Perspectives*, 13:141–153, 1999.

Barry Stroud. *The Significance of Philosophical Scepticism*. Clarendon Press, Oxford, 1984.

Jonathan Vogel. Tracking, closure, and inductive knowledge. In Steven Luper-Foy, editor, *The Possibility of Knowledge*, pages 197–215. Rowman and Littlefield, Totowa, 1986.

Jonathan Vogel. Skeptical arguments. *Philosophical Issues*, 14:426–455, 2004.

Brian Weatherson. The bayesian and the dogmatist. *Proceedings of the Aristotelian Society*, 107:169–185, 2007.

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

Crispin Wright. Warrant for nothing (and foundations for free)? *Proceedings of the Aristotelian Society Supplementary Volume*, 78:167–212, 2004.

Crispin Wright. The perils of dogmatism. In Susana Nuccetelli and Gary Seay, editors, *Themes from G.E. Moore: New Themes in Epistemology and Ethics*, pages 25–48. Oxford University Press, Oxford, 2007.

Ümit Yalçin. Skeptical arguments from underdetermination. *Philosophical Studies*, 68:1–34, 1992.

Elia Zardini. Confirming the less likely, discovering the unknown. Dogmatisms: surd and doubly surd, natural, flat and sharp. In Dylan Dodd and Elia Zardini, editors, *Scepticism and Perceptual Justification*. Oxford University Press, Oxford, 2013a. Forthcoming.

Elia Zardini. The modal metaphysics of evidence. ms, 2013b.