

Isolating Correct Reasoning

Alex Worsnip

Forthcoming in Magdalena Balcerak Jackson & Brendan Balcerak Jackson (eds.),
Reasoning: New Essays on Theoretical and Practical Thinking, Oxford University Press

Penultimate draft; please cite the published version if possible

Let me start with a threefold distinction between some different normative notions as they apply to attitudinal mental states like beliefs and intentions.

The first is that of the **reasons** or **justification** that one has for an individual attitude. Having a reason for an attitude, I assume, is a matter of there being considerations of some kind that speak in favor of that attitude. Perhaps these reasons have to pass some kind of epistemic filter or in some sense be “available” to the agent to be the kinds of reasons that justify in the operative sense. But at least an important subset of reasons justify attitudes “prospectively”, and then when those attitudes are held in a way that is appropriately responsive to those reasons, they are justified, for want of a better term, “retrospectively”.¹

The second notion is that of the **structural rationality** or **coherence** of certain *combinations of attitudes*. Here the idea is that there are certain combinations of attitudinal mental states that it is rationally impermissible to hold jointly. For example, it’s structurally irrational to simultaneously believe *p* and believe not-*p*, or to simultaneously intend to *F*, believe that one cannot *F* unless one *G*’s, and fail to intend to *G*. Unlike a claim about justification – say, that believing *p* is justified – the claim that one may not rationally (believe *p* and believe not-*p*) is completely silent on the substantive merits of any one individual attitude. It merely puts a rational constraint on which attitudes can be combined.

The third notion is that of **correct reasoning**, and rules thereof. Here reasoning is being understood – perhaps somewhat narrowly or idiosyncratically – in terms of transitions between attitudes. So correct reasoning involves making a correct transition from one attitude to another. For instance, it’s common to say that to move from believing *p* and believing (if *p* then *q*) to believing *q* is correct reasoning. And it’s common to say that moving from intending to *F* and believing that one cannot *F* unless one *G*’s to intending to *G* is correct reasoning.

This paper is part of a broader project investigating the relationship between these three normative notions, in both the practical and doxastic (“theoretical”) cases. My view is that the relationships between these notions are less tight than many have implicitly assumed (when they have even

For helpful discussion and feedback, I’m grateful to David James Barnett, Paul Boghossian, Cian Dorr, Ram Neta, Jim Pryor, Miriam Schoenfield, Daniel Whiting, members of the epistemology reading group at NYU, and the editors of this volume.

¹ The term “prospective” justification comes from Pryor (ms.), and corresponds to what many philosophers call “propositional justification”. The ugly talk of “retrospective” justification is my own. Epistemologists (including Pryor) often label the latter notion “doxastic” justification, but this label inherently resists generalization to non-doxastic attitudes, and my account is supposed to apply to both the doxastic and the practical domains.

distinguished them at all). This paper can be seen as continuing my case for this broad view. In previous work (Worsnip forthcoming), I argued for pulling apart the first and second notions and against some putative connections between them. Here I'm going to argue that the third notion is likewise to be pulled apart from both the first notion and the second. So, after a few more brief remarks on the notion of rules of correct reasoning (section 1), I will argue that attempts to understand rules of reasoning in terms of the justification of individual attitudes fail (section 2). Then, I will go on to further deny that rules of reasoning correspond to either requirements or permissions of structural rationality (section 3).

Out of the ashes of these attempts to reduce or at least identify correct reasoning with one of the other normative notions, I will make some gestures towards an account of correct reasoning as a more *sui generis* notion (section 4). I suggest that this *sui generis* account has some independently interesting results.

1. A bit more on correct reasoning

I don't want to give too specific an account of (rules of) correct reasoning at the outset, lest that prejudice the issues to be discussed. Rather, I am going to assume that there is some intuitive if inchoate notion of (rules of) correct reasoning that we can grasp, and that we are in the course of investigating which (if any) more precise notions in the neighborhood of this intuitive notion should feature in our ultimate theory of normativity. However, we can explicate the notion we're interested in somewhat by paying attention to paradigmatic examples of the kind of rules that it is intuitively correct to reason with. Here are two examples, one theoretical and one practical:

Modus Ponens Rule. From the belief p and the belief *if p then q* , derive the belief q .

Instrumental Rule. From the intention to Φ and the belief that to Φ you must Ψ , derive the intention to Ψ .

I offer these rules only as rough examples. They may need some modification. I even leave it open that these should ultimately be rejected (indeed, I'll consider such a view later on). But even as initially plausible *candidates* to be rules of correct reasoning, they offer some (rough) fix on the notion we are after.

One thing that is for sure is that we will not want to understand these rules so that, whenever a rule is correct, one is *required* to employ such a rule on all the occasions that one has the premise-attitudes that it mentions. For example, it cannot be right that whenever one has the belief p and the belief (if p then q), one is required to come to believe q . One reason for this it is also (sometimes) permissible to give up one of one's original beliefs instead.² This is particularly clear if we imagine the case so that, in one's initial state, one believes that p and believes that if p then q , but also has a positive belief that not- q . In this kind of case, plausibly one should (at least usually) give up one of these three

² This point is forcefully made by Harman (1986: 11-12). See also Broome (1999: 405; 2013: 82), Scanlon (2007: 84) and (for the analogous point in the practical case) Greenspan (1975: 272-3), Darwall (1983: 46-7) and Bratman (1987: 24-7).

inconsistent beliefs, but there is no general reason why one should, in all such cases, give up the belief that not-q (and, indeed, form belief in q) rather than give up the belief that p or the belief that if p then q. There is nothing privileged about reasoning via *modus ponens* rather than *modus tollens*. So we should understand the claim that (e.g.) *modus ponens* is a rule of correct reasoning as coming to something like the claim that *when* one reasons via *modus ponens*, one reasons correctly. There may in any particular case be other ways of reasoning correctly, some of which may be incompatible with reasoning via *modus ponens*.

Second, I do not want to presuppose at this stage that all *valid* reasoning is *correct* reasoning; I will treat that as a substantive issue to be settled by our best theory of correct reasoning. Validity is a logical notion, concerned with what follows from what; correctness is a normative notion. I will try not to use ‘valid’ and ‘correct’ interchangeably.³

A final general point about the general, intuitive notion of correct reasoning is that it focuses specifically on *transitions* between states. So, in the sense we’re interested in, one can be *reasoning* correctly even if one starts with unjustified premises, and so ends up with an unjustified conclusion. What is at fault in such a case is not one’s reasoning itself but one’s premises.⁴ There is a distinctive kind of rational achievement in being good at reasoning, and it is different from that of having generally justified starting-points.

2. Correct reasoning and justification

For the reason just given, someone who wants to link correct reasoning and justification cannot simply propose that correct reasoning always issues in a justified attitude. Nevertheless, many philosophers take a slightly subtler relationship between correct reasoning and justification for granted. They have assumed that rules of correct reasoning *preserve* justification. Let us call this the **preservation thesis**. As Jonathan Way and Daniel Whiting put it, the idea is this:

“If you reason correctly from justified premise-attitudes, you will reach a justified conclusion-attitude.” (Way & Whiting 2016: 1876)^{5,6}

I will begin by taking it that we want the account of rules of correct reasoning to vindicate rules of reasoning that are paradigmatically taken to be correct, such as the *modus ponens* rule and the instrumental rule. (Later, I’ll consider the revisionary strategy of rejecting these rules.) I’ll begin with

³ Harman (1986: 3-6) thinks that to talk of a valid rule of reasoning, or even to talk of deductive reasoning, is to commit a category error: there are only valid/deductive *arguments*. I don’t go this far: I think it is perfectly sensible to talk of deductive reasoning, and the validity of such reasoning.

⁴ Again, see Harman (1986: 7).

⁵ Wedgwood (2012: 273) endorses a very similar thesis.

⁶ Are we talking about prospective or retrospective justification here? As stated, I assume Way & Whiting’s principle is about retrospective justification. But one could also state an analogue of the principle for prospective justification: that if one *could* correctly reason from prospectively justified premise-attitudes to some conclusion-attitude, then the conclusion-attitude is prospectively justified. What I say will be adaptable as an argument against either claim.

the modus ponens rule, before turning more briefly to the instrumental rule, arguing that in both cases, the preservation thesis fails.

In the current operative sense of ‘preserve’, to say that the modus ponens rule preserves justification is to say that, for all cases, if one is justified in believing p and justified in believing (if p then q), then one is justified in believing q . The case that I will use to undermine this claim is familiar (though it is typically used for other purposes).⁷ Considering a long chain of deductions via modus ponens:

P_1
 If P_1 , then P_2
 So, P_2
 If P_2 , then P_3
 So, P_3
 [...]

 If P_{n-1} , then P_n
 So, P_n

Let’s call the premises in such a chain that are not themselves derived from other premises in the chain the “non-derived” premises. In the above chain, P_1 is a non-derived premise, and so are all the conditionals ((If P_1 , then P_2), (If P_2 , then P_3), etc). By contrast, P_2 , P_3 , etc, are derived in the course of the chain of deductions.

If this chain of deductions is long enough, then even though one is justified in believing every non-derived premise in the deduction, one may not be justified in believing P_n . This is because each premise, compatibly with its being justified, carries some small risk of being false.⁸ As the chain of deductions continues, these small risks aggregate, until one arrives at conclusions that may (unless they enjoy some other, independent support) have a significant chance of being false. To put the point more explicitly in terms of evidential support: supposing one has excellent but less than infallible evidence for P_1 , and excellent but less than infallible evidence for (If P_1 , then P_2),⁹ then barring further independent evidence for P_2 , one’s evidence for P_2 will be slightly worse than one’s evidence for either of the *individual* premises from which it is derived. And then, supposing one’s evidence for (If P_2 , then P_3) is likewise excellent but less than infallible, one’s evidence for P_3 will again be slightly worse. And so on, until we reach some number n that is high enough such that one’s evidence for P_n is too weak to justify one in believing P_n .

But this is inconsistent with the thesis that the *modus ponens* rule preserves justification, in the way that we are currently understanding it. We are supposing that every non-derived premise in the

⁷ The case derives from the preface paradox, originally introduced by Mackinson (1965). See Foley (1993), Christensen (2004) and Sturgeon (2008) for particularly powerful reiterations.

⁸ Some would challenge this modestly fallibilist idea, that being justified in believing p is compatible with some small risk of being mistaken. See Sutton (2007), Littlejohn (2012), and perhaps the latter-day Williamson, though not the Williamson of *Knowledge and its Limits* (cf., e.g., Williamson 2000: 9). I cannot defend it in the space available here, though I have defended a more robust form of fallibilism elsewhere (Worsnip 2015).

⁹ And that one’s evidence for P_1 and one’s evidence for (If P_1 , then P_2) are at least somewhat independent of each other.

chain of deductions is justified. So P_1 and (If P_1 , then P_2) are both justified. So, if the *modus ponens* rule preserves justification, P_2 is also justified. Since (If P_2 , then P_3) is another non-derived premise, it too is justified. So if the *modus ponens* rule preserves justification, P_3 is also justified. And so on, until we reach the result that P_n is justified. That contradicts the result from the previous paragraph, that even when all the non-derived premises in the chain are justified, P_n can fail to be justified. So the *modus ponens* rule does not preserve justification in the current sense.

I am, of course, crucially working here with an “on-off” notion of justification here. But that is necessitated by the way we are currently understanding the claim that correct reasoning preserves justification. Our way of understanding this, taking our lead from Way & Whiting, amounts to the claim that if one reasons correctly from justified premise-attitudes, one will arrive at justified conclusion-attitudes. It is hard to see how to make sense of this claim as employing anything other than an “on-off” notion of justification.¹⁰

Perhaps, however, this shows that Way & Whiting’s principle is not the best way to interpret the spirit behind the preservation thesis. Suppose we begin instead with a scalar notion of justification. Having done so, one might say, it becomes clear that the problem with the long chain of deductions via *modus ponens* is still ultimately a problem with the justification that one *starts* with for the premises employed in this long chain of deductions, rather than its failing to be preserved. Even if this scalar justification suffices for each of those premises to be justified in some “on-off” sense, it is still less than perfect, and this is what results in the conclusion of the argument being unjustified (in both a scalar and an “on-off” sense). One might then understand the preservation thesis as claiming that any deficiency in one’s justification for the conclusion of an instance of correct reasoning can be traced to a deficiency in one’s justification for its (non-derived) premises.

However, even this new interpretation of the preservation thesis fails. To see this, consider a variant on the long chain of deductions where each step involves a *single-premise* deduction. In contrast to the deduction via *modus ponens*, which moves from two premises – P_1 and (If P_1 , then P_2) – to the conclusion P_2 , we are now to suppose that P_1 simply *entails* P_2 ; no conditional premise is needed for the inference to be valid. And similarly for every other step. So the long chain of deductions simply looks like this:

P_1
So, P_2
So, P_3
[...]
So, P_n

Unlike the chain of deductions via *modus ponens*, this whole chain uses only one non-derived premise, P_1 . If every step is valid – as we are to suppose – then everything else in the chain follows

¹⁰ This is compatible, of course, with thinking that in some sense *underlying* this “on-off” notion of justification is a more scalar notion of justification, such that one’s being “on-off” justified in believing p is at least sometimes a matter of meeting some (perhaps vague and/or situation-dependent) threshold of more gradable or scalar justification for p .

from P_1 alone. Consequently, if there are deficiencies in one's justification for P_n that are traceable to the justification of its non-derived premises, they must be deficiencies in one's justification for P_1 .

Let us now suppose that one does have strong justification for P_1 . Could it still be the case that one's justification for P_n is poor? Here is one reason to think so:¹¹ we are not logically infallible, and we sometimes make mistakes when doing deductions. Just as we can be mistaken about the premises of arguments, then, we can also be mistaken about transitions from one step to another. For each step one makes in a deductive argument, then, there is in at least some good sense a small chance that one has made an invalid deductive inference. But *these* risks of error do still aggregate even in a chain of single-premise deductions, and when such chains are long, they will become significant.¹²

In fact, I think that there are two quite distinct ways that these risks of error can arise. One goes through a kind of higher-order evidence. If there are enough steps in the chain, and one has reason to think that one is not logically infallible, one should suspect that there's a very good chance one has gone wrong somewhere in the chain. And, at least on most theories of higher-order evidence,¹³ this has at least some effect on one's justification for believing the conclusion, P_n , at least on the justification supplied by the chain of deductions.

If this were the only way in which the risk of error could arise, it might encourage a picture on which each step in the chain of deductions itself transmits justification perfectly, as it were, only for this justification to *then* be "undermined" by the higher-order (justified) belief that one may have gone wrong somewhere in the deduction. But I think there is another way in which risk arises that calls into question the transmission of justification in each step. This is supplied by the actual risks that, given one's logical fallibility, one always incurs when engaged in the activity of deductive inferences. Here the risk is not created by *reasons to think* that one is logically fallible, as with the higher-order risk, but by one's actual logical fallibility. The thought is that at each step of the deduction, the risk (given one's logical fallibility) that one is not in fact performing a valid deductive inference means that a small amount of justification is "lost", such that the conclusion is (again, barring independent evidential support) slightly less justified than the premise.

This may seem controversial, so here's a way of motivating it. Suppose that some proposition p entails some other proposition q , but that the entailment is highly non-obvious. Now consider a person who reasons from a belief in p to a belief in q , but without any grasp of this entailment whatsoever. His inference is valid only by fluke: there is in a good sense, then, a strong chance that

¹¹ The line of argument here is directly indebted to that given by Lasonen-Aarnio (2008) and Schechter (2013); see also Pryor (ms.: 7-8). See DeRose (1999: 23, fn. 14) for an earlier voicing of a closely related thought.

¹² These claims, once again, are somewhat controversial. One might claim that, as long as one doesn't *in fact* make a logical mistake in deduction, merely having *evidence* that one has made such a mistake has no impact on one's justification for believing the conclusion. (See Wedgwood 2012: esp. 290-4; see also remarks in this direction made by Weatherston (ms.) and Lasonen-Aarnio (2014: esp. 322-3).) Some of what I say below will put pressure on such claims. But obviously this is a bigger topic than I can decisively settle here. And I should be up front that the kind of way of thinking I'm

¹³ Even theories on which the higher-order evidence is not necessarily *decisive* with respect to what first-order attitude one should take typically allow that such higher-order evidence has *some* impact. See e.g. Kelly (2010), Pryor (2013: 99-100), Lasonen-Aarnio (forthcoming).

his inference is invalid, even though it is in fact valid. I hope you share my sense that such a person's justification for *q* might then be considerably worse than his justification for *p*.¹⁴

Now, I am not imagining the person who performs a chain of single-premise inferences this way: I am assuming that the entailments in her chain are not especially opaque, and that for each step of her deduction, she has the normal reasons to (implicitly) think she is inferring validly that we typically have when we perform simple deductions. We could say that she *grasps* the entailments, as long as we do not let the factivity of 'grasp' push us in an unwarrantedly infallibilist direction: though her inferences *are* valid, she is – like all of us – subject to making occasional mistakes, even with simple inferences, and so there is in a perfectly good sense some small chance at each stage that she is inferring invalidly.

But now the question is this. If a *strong* chance that one has inferred invalidly can result in one's conclusion being (much) less justified than one's premise, why couldn't a *slight* chance that one has inferred invalidly result in one's conclusion being (slightly) less justified than one's premise? In both cases, the diminishing of justification occurs notwithstanding the actual validity of the inference. I think we should conclude that both cases are possible.

We have to tread carefully in drawing the lesson for the preservation thesis. As I have just claimed, deductively *valid* inferences can sometimes fail to fully preserve justification. But it is open to the defender of the preservation thesis to deny that whenever an inference is deductively valid, it is in the relevant sense thereby *correct*. Indeed, it seems that, it seems that a defender of the preservation thesis will *have* to deny that all valid inferences are correct, in order to deal with the simpler cases where the reasoner infers validly completely by fluke.¹⁵ She must then identify some privileged set of deductively valid inferences that constitute the rules of correct (deductive) reasoning, and that always preserve justification. The simple, very general rule to always infer *q* from *p* when *p* entails *q* will not be in this set, since such entailments can be highly non-obvious and opaque.

However, I have been assuming that the *modus ponens* rule *will* be in this set. And so now we simply need the point that, just as there can be some small risk of making inferential mistakes in some simple single-premise deduction, likewise there can be a small risk of making inferential mistakes in a two-premise deduction via *modus ponens*. So, in our original long chain of deductions via *modus ponens*, the reasoner has some reason to worry – if the chain is long enough – that not every single one of these deductions is in fact a legitimate instance of *modus ponens*. Even if it happens that each one of them *is* a legitimate instance of *modus ponens*, there is still in the relevant sense, from the reasoner's perspective, a risk that it might not have been (in the same way that a belief can be true but in a perfectly good sense carry a risk of being false). From the reasoner's perspective, one of her steps might in fact have been an instance of affirming the consequent, or have involved an equivocation such that it superficially resembles an instance of *modus ponens*, but in fact invokes a subtly different proposition in the antecedent of the conditional premise than in the unconditional premise. These are the kinds of mistakes we can make, especially when reasoning implicitly and quickly, as we often do.¹⁶

¹⁴ See also Pryor (ms.: 4). Is it merely *retrospective* justification that he lacks? I don't think so: if the entailment is sufficiently non-obvious, such that he couldn't possibly have grasped it given his situation and cognitive capacities, I think he lacks prospective justification as well.

¹⁵ Cf. Wedgwood (2012: 279).

¹⁶ See, e.g., Kahneman (2011: 45).

As in the single-premise case, such risks slightly weaken our justification for the conclusions of our reasoning, even when such reasoning is correct. Deficiencies in one's justification for the conclusion of such reasoning that are traceable to these risks are not traceable to a lack of justification for the non-derived premises in one's reasoning. They are incurred in the course of the reasoning itself, notwithstanding its correctness. Thus, even the scalar interpretation of the preservation thesis fails.¹⁷

It's now time to consider a more radical response to the objections I have voiced so far. I have been assuming that we have some fix on intuitively plausible rules of correct reasoning –the *modus ponens* rule being one such rule – and that it is the job of a general account to prove its extensional adequacy against this background. But if one is convinced that what it *is* for a rule of reasoning to be correct just is for it to preserve justification, one might instead try rejecting the idea that the *modus ponens* rule is, as it stands, a rule of correct reasoning. Perhaps the correct rule will have to be some heavily caveated version of the *modus ponens* rule.¹⁸ Maybe it tells us to infer via *modus ponens* when there are no relevant risks of the kind that I have been urging prevent it from always preserving justification. Other caveats may be needed too.

I do not think we should be casual about the idea that we can simply caveat the *modus ponens* rule to take account of the problems I have been discussing, and thereby solve the problems for the preservation thesis.¹⁹ First, there is the general worry that the caveats to the *modus ponens* rule are gerrymandered to save the preservation thesis, and that they are either independently implausible or render the resulting preservation thesis trivial. But even setting this broad methodological point aside, it is very hard to get the caveats to the *modus ponens* rule right so that they both save the preservation thesis from the above counterexamples *and* give us a plausible account of the *modus ponens* rule itself. Here is a first attempt at a caveated rule:

Modus Ponens Rule-Caveated. (From the belief *p*, and the belief *if p then q*, derive the belief *q*), unless there are risks sufficient to disrupt the preservation of whatever justification one has for these two beliefs to the belief in *q*.

This caveated rule faces a dilemma. Either we are to understand 'disrupt' as covering any kind of slight disruption to the preservation of justification, or we are to understand it as covering only *significant* disruptions to the preservation of justification. On the former interpretation, the caveat seems to exclude every (or at least nearly every) instance of *modus ponens* reasoning from falling under the *modus ponens* rule. For any inference carries with it *some* risk of error, and so the preservation of one's

¹⁷ One might now suggest that, having gone scalar about justification, we should also go scalar about belief and work with credences instead of all-out beliefs, and rules of reasoning for credences. Perhaps then we can say (for example) that the correct *modus ponens* rule should go from high credences in *p* and (*if p then q*) to a *slightly* lower credence in *q*. (See Wedgwood 2012: esp. 288-9.) But I don't think that a single rule of this kind that would always preserve justification can be specified, even in principle. This is because how much justification is "lost" in an inference will vary from case to case depending on the particular thinker's epistemic position with respect to the inference.

¹⁸ Remarks by Schechter (this volume) suggest that this is his response to the problems he himself noted with long chains of deductions.

¹⁹ Lasonen-Aarnio (2014: 322-3) also argues against such a caveating of our rules of inference. However, she assumes that this bolsters the conclusion that we are always *justified in believing* the conclusions of such inferences even when we have evidence that we have inferred wrongly. This implicitly assumes the preservation thesis, in taking it for granted that if we can avoid caveating our rules of reasoning, parallel conclusions about justification will follow.

justification is always *slightly* disrupted in inference. On the latter interpretation, the caveat fails to deal with the examples involving long chains of inference. For the whole point of those examples is that at no one stage at which the rule applies is the disruption large. Rather, through many repeated steps, the inferences eventually lead to an attitude which is much less justified than the attitude or attitudes we began with. So on this interpretation, the preservation thesis will still fail even for the caveated *modus ponens* rule.

It seems that the only way to get around this dilemma is to return to an “on-off” notion of justification. The cases where it would be incorrect to infer via *modus ponens*, on this view, will be those where there are risks such that, although the premises are each on-off justified, the conclusion is not on-off justified. This may be because the premises only *just* meet the threshold for justification.

This sounds more promising. But there is another problem. Remember that a piece of reasoning that starts with unjustified premises can be perfectly correct *qua reasoning*. This is why the justification-based account of rules of reasoning talks about the *preservation* of justification, not about the outright justification of the conclusion of the reasoning. The caveated *modus ponens* rule does not say anything inconsistent with this in itself. But now consider a more complex case. Suppose that the premises of a putative *modus ponens* inference are (on-off) justified, but only just – such that the conclusion of one’s reasoning is not (on-off) justified. Suppose further that one overestimates the (scalar) justification of the premises, thinking that they *easily* meet the threshold for on-off justification. As such, one takes it that it is perfectly safe to infer the conclusion – and does so.

This reasoning is not vindicated as correct by the caveated *modus ponens* rule, since there are risks sufficient to disrupt the preservation of one’s on-off justification from premises to conclusion. But that is an odd result, especially when juxtaposed with the claim that in the simpler case where one reasons via *modus ponens* from straightforwardly *unjustified* premises, one *is* reasoning correctly. If there is nothing wrong with one’s reasoning in the simpler case, what could be wrong with one’s reasoning in the more complex case? In both cases, one makes a mistake only in beginning with premises that are not well-supported enough to justify the conclusion – not, intuitively, a mistake in one’s reasoning itself. The caveated version of the *modus ponens* principle yields the bizarre result that when one makes a *big* mistake about the justification of one’s premises (considering them justified then they are not), one reasons correctly – but when one makes a *smaller* mistake about the justification of one’s premises (getting it right that they are on-off justified, but considering them to be somewhat *more* justified than they are), one reasons incorrectly!

It may seem like there’s a clear fix to the caveated *modus ponens* rule to take care of this problem. One could say the caveat should cover cases in which one *believes* there to be risks that disrupt the preservation of justification, rather than cases in which there *are* such risks. This avoids the problem raised above, but at the price of once again opening the way for failures of the preservation thesis even given the caveat on the *modus ponens* rule. For consider a case where there are risks that disrupt the preservation of justification, but one lacks the belief that there are such risks. If these were instances of correct reasoning – as they are according to the current proposal – they would be instances of correct reasoning that do not preserve justification. So this kind of caveat would not save the preservation thesis.

Thus, I think the better thing to say is that the original, uncaveated version of *modus ponens* itself is the rule of correct reasoning, but that risks associated with the preservation of justification can be reasons, from time to time, not to employ it, notwithstanding its correctness. As such, the preservation thesis cannot be sustained. The *modus ponens* rule provides a counterexample. Very similar strategies could be mimicked for other rules of theoretical reasoning. I want to round off this section by briefly sketching how the problems developed here also extend – less familiarly – to the instrumental rule and thus to rules of practical reasoning.

The extension is simple. Consider a long chain of practical reasoning via the instrumental rule, where ‘A’ stands for an action:

Intend(A_1)
 Believe(To A_1 , Must A_2)
 So, Intend(A_2)
 Believe(To A_1 , Must A_3)
 So, Intend(A_3)
 [...]
 Intend(To A_{n-1})
 Believe(To A_{n-1} , Must A_n)
 So, Intend(A_n)

Let us grant that both one’s intention to A_1 , and each of the instrumental beliefs in the chain, is justified (that amounts to granting that each non-derived premise in the chain of reasoning is justified). Even when this is so, each instrumental belief comes with a small risk of error. One’s justification for an intention to take a means tracks both one’s justification for intending the relevant end, and one’s justification for believing that the means is necessary for the end. So, since one’s belief that to A_1 , one must A_2 comes with a small risk of error, one’s justification for intending to A_2 is slightly worse than one’s justification for intending to A_1 . Likewise, one’s justification for intending to A_3 is slightly worse than one’s justification for intending to A_2 . And so on, such that if there are enough steps in the chain, one’s justification for intending to A_n may be very weak (at least bracketing independent reasons to A_n that are not derivative on one’s reason to A_1). Put another way, each instrumental belief in the chain introduces an extra risk that it may not be the case that one must A_n in order to A_1 , and these aggregate so that overall it may be very doubtful whether one must A_n in order to A_1 . Then, one’s justification for intending to A_n – again, insofar as it is derivative on one’s justification for intending to A_1 – may be very weak.

Moreover, one may be (justifiably) doubtful about whether every step in the chain is a correct instance of instrumental reasoning. It could be – from one’s own perspective – that some step involved the fallacious, affirming-the-consequent-esque pattern [Intend(Φ), Believe(To Ψ Must Φ), so Intend(Ψ)]. Or one step may have involved the pattern [Intend(Φ), Believe(To Φ Can Ψ), so Intend(Ψ)]. And so on. Provided these mistakes are a risk, even a scalar interpretation of the preservation thesis will fail for instrumental reasoning just as it failed for deductive reasoning: there

will be deficiencies in one's justification for intending to A_n that are not traceable to deficiencies in one's justification for any of the premises of the chain of reasoning.

3. Correct reasoning and structural rationality

What I've said so far has put pressure on attempts to forge a tight link between correct reasoning and justification. But according to many philosophers – especially philosophers of practical rationality – there is an important distinction between justification (as we have been understanding it so far) and what we may call *structural rationality*.²⁰ And perhaps even if there is no tight link between correct reasoning and justification, there is a tight link between correct reasoning and structural rationality.

One proposal about this difference is suggested by at least one possible interpretation of some remarks of Nadeem Hussain's (ms.). The idea here is that reasoning is correct when it brings one to satisfy structural requirements.²¹

However, as it stands this cannot be right. First, the purported condition for correct reasoning is not necessary. As I said in section 1, the notion of correct reasoning is not such that whenever reasoning is correct, one is always *required* to undertake such reasoning. Although this claim is not quite the same as the claim that reasoning is correct when it brings one to satisfy structural requirements, the same sorts of cases undermine both claims. For example, consider disjunction introduction. It is surely correct reasoning to move from believing p to believing $(p \vee q)$. But is there a structural requirement to the effect that believing p *without* believing $(p \vee q)$ is always rationally forbidden? It seems not. Suppose you believe that it's raining. If the purported structural requirement held, you'd be irrational if there were even one disjunction of the proposition that it's raining and some other proposition q (quantifying over all possible propositions q), such that you failed to believe the disjunction. For example, you would be irrational if you failed to believe the disjunction that either it's raining or V.I. Lenin was an off-spin bowler for Nottinghamshire County Cricket Club. But that is not plausible: you are surely not irrational for lacking this belief. (You might not even have the concepts, such as the concept of an off-spin bowler, that are plausibly required to have it.) Rationality does not require you to fill your mind with trivial, useless extensions of your existing belief state.²² But then it seems that disjunction introduction is correct reasoning, without bringing one to satisfy any structural *requirement*.²³

²⁰ For this terminology, see Scanlon (2007) and Fogal (2015).

²¹ See, e.g., Hussain ms.: 46.

²² Yet again, this point was heavily stressed by Harman (1986: 5-6, 12), in a slightly different context. See also Broome (2013: 158-9, 246-7.)

²³ Hussain recognizes this problem. His reply is that the requirements of rationality supply norms of how you ought to reason *if* you are going to reason. I do not think this helps, however. First, as it stands the proposal is too vague. Perhaps I am, and ought to be, reasoning about *something*, but that still doesn't mean that I ought to be reasoning from the belief that it's raining to the belief that either it's raining or V.I. Lenin was an off-spin bowler for Nottinghamshire County Cricket Club. Second, and perhaps more importantly, the proposal simply does not make this original problem go away. Suppose I *do* in fact engage in the reasoning via disjunction introduction, moving from the belief that it's raining to the belief that either it's raining or V.I. Lenin was an off-spin bowler for Nottinghamshire County Cricket Club. My reasoning is correct. But then Hussain's claim that (correct) reasoning involves coming to satisfy rational requirements still entails

Second, the purported condition for correct reasoning is not sufficient.²⁴ Suppose you intend to attend your court appearance, and believe that to do so you must get your car fixed, but you fail to intend to get your car fixed. You violate the instrumental requirement, and if you eliminate any of these three states, you will *ipso facto* no longer be violating (this particular instance of) this requirement. But there are routes of reasoning that achieve this, yet do not constitute correct reasoning. For example, suppose your aunt Hilda fixes cars, and you believe that to get your car fixed you must call Hilda. And suppose you do intend to call your aunt Hilda. If you were to reason from these two states (intending to call Hilda and believing that to get your car fixed you must call Hilda) to intending to get your car fixed, your reasoning would not be correct: it would be like a practical version of affirming the consequent.²⁵ Nevertheless, in adopting the intention to get your car fixed, you would come to satisfy a structural requirement that you previously failed to satisfy, since you would no longer be in the combination of states where you intend to attend your court appearance, believe that do so you must get your car fixed, but fail to intend to get your car fixed.²⁶

The first, and perhaps also the second, of these problems arise for Hussain's view because he tries to tie correct reasoning to structural *requirements*. One might think that the solution here is to tie correct reasoning not to structural requirements but to structural *permissions*. Indeed, this tack is taken by John Broome, who claims that "if it is correct to reason to some conclusion, that is because rationality permits you to reach that conclusion" (Broome (2013: 219). Here Broome is still speaking of what we have been calling structural rationality, but crucially it is on Broome's view rational *permissibility* rather than rational *requiredness* that is most tightly linked to correct reasoning. According to Broome, (correct) reasoning will often bring one into satisfaction of rational requirements, and is one of our main ways of doing this.²⁷ But equally, "in many cases, you commit no offense against rationality by failing to do a piece of reasoning that would have been correct had you done it."²⁸

Broome develops his account by appealing to the notion of a "basing permission", which specifies that it is rationally permissible to *base* some particular attitude on some other attitude or attitudes.²⁹ According to Broome, "each permission will determine a rule, and reasoning by correctly following that rule will be correct."³⁰ For instance, reasoning by *modus ponens* is correct because it is always (structurally) rationally permissible to base a belief in q on believing p and believing that if p then q. Here is the way Broome writes out the basing permission that corresponds to the *modus ponens* rule (very slightly adjusted to match my notation):

that there is some rational requirement that I now satisfy and previously did not satisfy. And this does not seem to be so. Stipulating that it may be a case where it's not the case that one ought to reason has not solved the problem.

²⁴ For a structurally similar example to the one I am about to give, see Broome (2013: 246).

²⁵ Compare: I intend to go to the bar; I believe that to drink five bottles of tequila I must go to the bar; so I intend to drink five bottles of tequila.

²⁶ Nor would you have purchased this compliance at the price of creating a violation somewhere else in your belief states. It is not structurally irrational for you to be in a combination of states where you intend to call Hilda, believe that to get your car fixed you must call Hilda, and intend to get your car fixed. What went wrong was how you got to these states, not the combination you ended up with.

²⁷ See Broome (2013: 207).

²⁸ *Ibid.*: 219.

²⁹ *Ibid.*: 189-90.

³⁰ *Ibid.*: 247; see also 255.

Modus Ponens Permission. Rationality permits N that N believes q at some time on the basis of believing p at some time and believing (if p then q) at some time.³¹

Now, it is not immediately obvious what this kind of construction, with a permission to believe some unspecified variables, means. When Broome states (wide-scope) *requirements* of structural rationality, the variables in schema like the above one are supposed to quantify over all propositions.³² So, for example, for all propositions p, it supposed to be the case that you are rationally required not to both believe p and believe not-p. If you do this, whatever the proposition, you are irrational (according to the requirement). This is easy to make sense of.

What about a permission, like the *modus ponens permission*, however? Broome tells us that “a basing permission is nothing other than the negation of a basing prohibition.” A basing prohibition is a requirement that *forbids* basing some attitude on some other attitudes. Given this, the *modus ponens* permission amounts to the following:

$\neg(\text{Rationality requires of N that N does not believe q at some time on the basis of believing p at some time and believing (if p then q) at some time}).$

Unfortunately, the lack of quantifiers makes this claim ambiguous, depending on whether the negation go in front of the quantifiers or after them:

- (a) $\neg(\forall N \forall p \forall q (\text{Rationality requires of N that N does not believe q at some time on the basis of believing p at some time and believing (if p then q) at some time})).$
- (b) $\forall N \forall p \forall q \neg(\text{Rationality requires of N that N does not believe q at some time on the basis of believing p at some time and believing (if p then q) at some time}).$

These claims are not equivalent. I am genuinely unsure as to which Broome intends. Plausibly, when Broome says that a basing permission is just the negation of a basing prohibition, the more natural reading is (a). After all, a full statement of a basing prohibition would include its quantifiers. It is (a) that it equivalent to saying that there is no basing prohibition forbidding one from believing q on the basis of believing p and believing in p then q. On the other hand, if we read the original statement of the *modus ponens* permission as itself implicitly universally quantified (as Broome’s claims about requirements are), then we get (the logical equivalent of) reading (b). So I am not sure how to read Broome here.

I’ll now argue that neither reading manages to get Broome what he wants: reading (a) is too weak to vindicate the proposed connection between structural rationality and correct reasoning, while reading (b) is too strong to be plausible.

Begin with (a). (a) just says that, quantifying over all propositions p and q, it is not *always* forbidden to believe q on the basis of believing p and believing (if p then q). This reading of the *modus*

³¹ *Ibid.*: 191.

³² Or attitudes or actions, in other cases.

ponens permission is extremely weak. It is equivalent to an *existentially* quantified reading of the *modus ponens* permission as originally stated: there are *some* cases where it is permissible to believe *q* on the basis of believing *p* and believing (if *p* then *q*). But once we fix on this understanding of basing permissions more generally, Broome's core claim that every basing permission generates a rule of correct reasoning is implausible. For example, consider the following claim:

- (1) $\neg(\forall N \forall p \forall q (\text{Rationality requires of } N \text{ that } N \text{ does not believe } q \text{ at some time on the basis of believing } p \text{ at some time}))$.

What this says is that there are some propositions *p* and *q* such that it is permissible to believe *q* on the basis of believing *p*. This is surely true. But surely there is no *general* rule of correct reasoning that says:

From the belief *p*, derive the belief *q*.

Thus, there is no rule of correct reasoning corresponding to this very weak kind of basing permission.

One might now say that the commitment should only be that *when* it is permissible to believe *q* on the basis of believing *p*, it is correct to reason from believing *p* from believing *q*. But if that is the relationship between structural rationality and correct reasoning then, likewise, we would say only that *when* it is permissible to believe *q* on the basis of believing *p* and believing (if *p* then *q*), it is correct to reason from believing *p* and believing (if *p* then *q*) to believing *q*. But that is less than what the *modus ponens* rule says: the *modus ponens* rule says that quite generally it is correct to reason from believing *p* and believing (if *p* then *q*) to believing *q*. So this way of developing the view would not vindicate the *modus ponens* rule (as Broome explicitly promised to do).

Consider now claim (b). (b) says, in effect, that it is *always* rationally permitted to believe *q* on the basis of believing *p* and believing (if *p* then *q*). This claim, I think, is implausibly strong. Even confining ourselves to purely structural rationality (and so bracketing whether your beliefs in *p* and (if *p* then *q*) are themselves justified), there are cases where, in believing *q* on the basis of believing *p* and believing (if *p* then *q*), you are not structurally rational.

For a first example, suppose that you believe that it is raining, and believe that if it is raining then it is not raining. And suppose that on the basis of these two beliefs, you believe that it is not raining. This combination of states is structurally irrational, since it involves your believing that it is raining and believing that it is not raining; that is, believing contradictory propositions.³³

³³ It might be objected here that the irrationality is not *specifically* in the basing of the belief that it is not raining on the beliefs that it is raining and that if it is raining, then it is not raining. Rather, it is just in the initial state of believing that it is not raining and believing that if it is raining, that it is not raining. I am not sure about this: there does seem to *also* be something irrational about forming the belief that it is not raining partly on the basis of the belief that it *is* raining. But even if we concede the point, it does not save the present way of reading the *modus ponens* permission as quantifying over all propositions. For Broome makes it clear that he intends the *modus ponens* permission to be read as shorthand for 'Rationality permits *N* that *N* believes *p* at some time, *N* believes (if *p* then *q*) at some time, and *N* believes *q* at some time on the basis of believing *p* at some time and believing that (if *p* then *q*) at some time' (see the general form of basing permissions set out on p. 190). And, read as quantifying over all propositions, this full statement of the

For a second example, return to the long chain of *modus ponens* deductions that we discussed in section 2. In this case, we said, there is a slight risk of error associated with each premise in the deduction, and these risks aggregate so that eventually the probability of the conclusion is low. But now let us suppose that the agent herself *recognizes* these risks of error, and even recognizes how they aggregate. Would she be even *structurally* rational to base belief on the conclusion of the long chain on its premises? I think not. But if it were always structurally rational to base a belief *q* on a belief *p* and a belief (if *p* then *q*), then each new conclusion drawn in the chain would be structurally rational, and so ultimately would the final conclusion.

I conclude that neither Hussain nor Broome has given us a convincing account of a tight relation between correct reasoning and structural rationality. Of course, there might be other ways of trying to do so. I can think of only one that seems initially promising. One might say that, if one begins with a belief set that does not violate any structural requirements, and reasons correctly, one will never then be left in a position where one does then violate a structural requirement. That is: reasoning correctly can never introduce any *new* structural irrationality into one's belief set.

However, I think once again that the long chain of deductions case is a counterexample to this claim.³⁴ Suppose that one believes every (non-derived) premise of the long deduction, and assigns a very high but non-one credence to each such premise. Surely there need be no structural irrationality involved in this combination of states so far. Let us now also add that, recognizing the way that the risk aggregates, one has a credence < 0.5 in the conclusion of the long deduction. This does not seem to introduce any structural irrationality into one's state either.³⁵ But, one can correctly reason, via a series of *modus ponens* inferences, from the beliefs in the premises to belief in the conclusion of the long deduction. However, it is structurally irrational (I claim) to believe a proposition while also assigning it credence < 0.5 (thus regarding it as more likely to be false than true). So we have a case where correct reasoning has introduced structural irrationality into a previously structural rational belief state.³⁶

4. Correct reasoning as a *sui generis* notion

permission is clearly falsified by a case where *p* and *q* are contradictories: rationality does not permit this combination of states in such a case, since rationality does not permit believing *p* and (if *p* then *q*) when *p* and *q* are contradictories.

³⁴ Here I build on a similar argument from another context in Worsnip (2016).

³⁵ If you think it does, there is a slightly more complex way to run the argument. We can make the case one where the long chain of deductions issues in a conclusion that, of logical necessity, is true *only* if all the premises are true. The simplest way to do this is to make the conclusion the conjunction of all the premises, as in the preface case. Now, one can actually *correctly reason* from the original non-one credences in the premises to a low credence in the conclusion (given suitable probabilistic independence between the premises). So someone defending the claim that correct reasoning never introduces any structural irrationality will now either have to accept that the low credence in the conjunction does not introduce any structural irrationality, or concede that their own view is false.

³⁶ This is interesting, since some (e.g. Fogal 2015) think of preface-type cases as involving conflicts between structural rationality and justification, whereby structural rationality requires one to believe the logical conclusions of one's other beliefs, yet one would not be justified in doing so. My argument suggests that in preface-type cases, believing the logical conclusions of one's other beliefs can be *structurally* irrational.

One reaction to what I have argued so far would be to give up on the notion of correct reasoning, or at least that of any systematic *rules* of correct reasoning that will include *modus ponens*, the instrumental rule, and so on. However, I do think there is an intelligible sense in which when one reasons with *modus ponens*, one reasons correctly, and so I am sympathetic to the idea that correct reasoning can be understood as a *sui generis* notion, one that is not reducible either to justification or to structural rationality. In the space remaining I can only make some brief remarks about such a *sui generis* notion and its virtues.

So far, I have been maintaining a fairly loose fix on the notion of correctness, trying to leave open what exactly it amounts to. But now I want to propose that the notion of correctness as it features in “correct reasoning” should be understood in a way analogous to how it features in “correct belief”. Several philosophers have defended the idea that there is a notion of correctness for belief such that a belief is correct iff it is true.³⁷ Although, when it comes to belief, correctness is coextensive with truth, correctness is a normative concept in a way that truth itself is not.

A further way to bring out the contrast between correctness and truth is to see that correctness can apply to intentions (and actions), whereas truth cannot. What unifies the notion of a correct belief and a correct intention is that in both cases the notion of correctness corresponds to the maximally “objective” (or, as I like to call it, the “super-objective”) reading of ‘ought’, the one that is relative to all the facts and not at all constrained by the agent’s epistemic situation. For belief, relative to an informational body containing all the facts, rather than just those that constitute the agent’s evidence, the agent just ought to believe whatever is true. For intention (or action), the agent ought to (intend to) do whatever will in fact satisfy the ideal practical norms. So, for example, if the ideal practical norms tell one to maximize utility, the agent “super-objectively” ought to (intend to) do whatever will *in fact* maximize utility, even if she is in no epistemic position to know what this is.

Clearly, it may not always be in any ordinary sense of the terms either justified or rational to believe or intend what it is correct to believe or intend. There are beliefs which would be true but are not supported by one’s evidence, and there are actions that will, as a matter of fact, satisfy the ideal practical norms, though one has no way of knowing this. Indeed, in both cases, one’s evidence can strongly suggest that the correct belief or intention is in fact incorrect. In these cases, the correct beliefs and intentions would still (of course) be correct, but would not be justified.

I suggest that we should understand the correctness of following rules like *modus ponens* and the instrumental rule in the same way. It is always, in the super-objective sense, correct to reason by *modus ponens* or by the instrumental rule, but, as I have been urging in the last two sections, it is not always justified or rational to do so.

Why take these rules to be correct as rules of reasoning in the same way that true beliefs are correct as beliefs? Such reasoning obviously does not always yield correct *beliefs*: reasoning by *modus ponens* can lead one to false belief. Nevertheless, these rules do *preserve* correctness (though, as I argued in section 2, they do not always preserve *justification*), and this is one reason to think that reasoning by them is correct *qua reasoning*, where this (as I said in section 1) focuses on the status of the transition between states rather than the status of the states themselves.

³⁷ See, especially, Shah (2003); Wedgwood (2002, 2013); Gibbard (2005).

To spell this out: since correctness for belief is truth, deductive rules of inference like *modus ponens* – which are truth-preserving – are also correctness-preserving. That is: if one begins with *correct* premises and reasons by *modus ponens* (or any other valid rule of inference), one will always arrive at a correct conclusion. Similarly, the instrumental rule is correctness-preserving: if you super-objectively ought to Φ , and it's true that to Φ you must Ψ , then you super-objectively ought to Ψ . Again, the analogy to the practical case here shows that the notion of correctness-preservation is not just a trivial way of restating the (non-normative) notion of truth-preservation. Interestingly, this suggests an important parallel between deductive reasoning in the theoretical case and instrumental reasoning in the practical case: both are correctness-preserving.

This is not to say that all correct reasoning is correctness-preserving;³⁸ nevertheless, that some rule of reasoning is correctness-preserving is at least a good reason to take it to be correct.

This picture allows us to issue a qualified defense of the much-debated claim that, as it is sometimes put, “logic is normative for thought.”³⁹ On my account, this is true in the sense that deductive logic supplies us with rules of reasoning that have a normative status and role as rules of *correct* reasoning – they are not merely inert descriptive principles. But we can also preserve the intuitive thought that there are sometimes situations when we (in a very good sense) *shouldn't* reason with deductive rules – in just the same way that often there is a belief that is (as a matter of fact) correct, but that a responsible epistemic agent *shouldn't* hold.

Moreover, on this picture, *all* (deductively) valid reasoning is correct reasoning, since all valid reasoning preserves correctness. In section 1 I bracketed this question as one to be settled by our substantive theory, and in section 2 I considered some views on which not all valid reasoning is correct reasoning. Given an attempt to maintain a tight link between correct reasoning and justification, it seems that one is under pressure to claim that not all valid reasoning is correct, since reasoning that is only flukily valid, or where the agent is in no position to grasp the validity of the reasoning in question, does not appear to be justified or to preserve justification. But then this leads one into the rather difficult project of non-arbitrarily demarcating the “simple” or “basic” rules which are rules of *correct* reasoning from other deductively valid inference-patterns that are not to count as correct.⁴⁰

On the present picture, by contrast, we do not need to engage in this project, because calling reasoning correct does not commit us to thinking of its instances as justified or justification-preserving. As I have argued, even when one follows very simple rules of reasoning – such as *modus ponens* – there are small risks of error that can disrupt the preservation of justification. The difference between such rules and much more complex but valid rules of reasoning, then, is not one of kind but of degree. Reasoning by the more complex rules will be unjustified more often, and involve greater disruptions of the preservation of justification, because these complex rules tend not to be ones that the agent has even an implicit grasp of when she reasons. All of this is, however, at least somewhat

³⁸ Most obviously, inductive and abductive reasoning are not (always) correctness-preserving. One might take this to show that there are two interestingly different kinds of rules for reasoning: those that are strictly correctness-preserving and those that are not. One might also claim that the latter category seems to involve a “looser” kind of reasoning that is less strictly rule-governed.

³⁹ For discussion, see (e.g.) Harman (1986), Sainsbury (2002), Field (2009) and MacFarlane (ms.).

⁴⁰ See, e.g., Wedgwood (2012: 279). For discussion of the difficulties of various different ways of identifying such a privileged set of basic rules, see Schechter (this volume).

dependent on the agent in question: some thinkers have greater logical capacities than others, a point that again makes it unattractive to try to demarcate some privileged subset of the valid rules that are also to count as correct, where validity does not entail correctness. What we can say, instead, is that the correlation between correctness and justification will be weaker with complex rules than with simple ones.

This leads us to what I suspect many may think is the elephant in the room here. I have tried to pull correct reasoning apart from the justification and rationality of the states that one reasons from and to. And I am also saying that saying that reasoning is correct does not involve any commitment to saying that the reasoning itself is justified either. But if this is right, and if correctness is best understood as a super-objective normative notion, then shouldn't we also be able to give some account of the *justification* of reasoning itself? And perhaps that is what we should really care about. So in giving an account of correct reasoning where correctness is understood in this super-objective way, we have changed the subject.

Now, one possibility here – which I have been ignoring until now – is to deny that, properly speaking, justification ever attaches to reasoning itself, as opposed to the states that it begins and ends with. In the theoretical case, the kind of justification involved is supposed to be *epistemic* justification. But plausibly, reasoning is an *act*.⁴¹ Can an *act* ever be *epistemically* justified? There is at least a challenge to say how this can be so.

That said, I will concede for the sake of argument that we can talk of the justification (as distinguished from correctness) of reasoning. Even given this, however, what I want to deny is that there are *rules* of justified reasoning in the same way that there are *rules* of correct reasoning. On the picture I have been arguing for, it can be unjustified to follow pretty well any rule of reasoning, given the right setup and background conditions. So if we are looking for rules of reasoning the following of which will always be justified, we are in for a disappointment. So as long as we are looking for *rules* of reasoning that have some positive normative status, I maintain, the super-objective notion of correctness is the only notion we have to work with. If that's right, it's not changing the subject to fix on this notion of correctness in looking for an intelligible notion of “rules of correct reasoning”.

Nevertheless, we can still ask: on what will the justification of the employment of a rule of correct reasoning in some particular case depend? It will at least partly depend, I think, on whether one implicitly grasps the correctness of the rule that one is following (if one is following a rule at all).⁴² That is the obvious difference between the ordinary, good case of justified reasoning on one hand and the case in which someone performs a valid inference by “fluke”, or when one makes an inference that is too complex to justifiably proceed in a single step, on the other.⁴³

I've heard it objected at this point that this claim raises some kind of problem analogous to that which the tortoise foists on Achilles in Lewis Carroll's (1895) famous parable. I do not think that it does. What the tortoise insists on – and creates so much trouble by insisting on – is treating facts about entailments – about some set of premises entailing a conclusion – *themselves* as premises of the

⁴¹ See, e.g., Broome (2013: 235-42); Pettit (2007).

⁴² Somewhat similar suggestions are made by Fogal (2015: 60), Pryor (ms.: 9-10), and – for the non-basic rules that he does not regard as *always* justified to employ – Wedgwood (2012: 279).

⁴³ I also think that when one has misleading evidence that one is following a rule of correct reasoning, but isn't actually doing so, there's a perfectly good sense in which one's reasoning is justified. Here I agree with Fogal (2015: 59).

argument, before the argument is to be accepted as valid. Transposed into the language of reasoning, and rules of reasoning, the idea would be that the *modus ponens* rule itself should have the relevant entailment-fact included as one of the beliefs that one must reason from. In other words, the *modus ponens* rule should in fact read:

Tortoisey Modus Ponens Rule. From the belief *p*, the belief *if p then q*, and the belief (if (*p* and *if p then q*), then *q*),⁴⁴ derive the belief *q*.

As Carroll shows, once we start down this road, we are led into a horrible regress, since one could just as well demand that one also include the belief (if (*p* and *if p then q* and (if (*p* and *if p then q*), then *q*)), then *q*), and so on *ad infinitum*.

However, I have not said that we should modify the *modus ponens* rule in this way. Rather, I have said that it is a condition of being justified in employing the *modus ponens* rule – as it originally stood – that one have some (implicit) grasp of its correctness. This is more than a terminological difference, because it does not generate the same regress that we get when we make the tortoise-y modification of the *modus ponens* rule itself.⁴⁵

If the implicit grasp of a rule's correctness plays a crucial role in understanding its justification, then the notion of a rule of correct reasoning turns out to do important theoretical and explanatory work even when giving an account of the *justification* of reasoning.⁴⁶ This reinforces, again, that the account of correct reasoning I have been suggesting is not just a way of changing the subject from what we were interested in. We began with an inchoate notion of correct reasoning – presupposing very little about how correctness was to be understood – and we have been looking for sensible, well-defined, theoretically explanatory notions in its neighborhood. I think we have found one in the way that I have proposed that we understand correct reasoning.

References

- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Broome, J. (1999). 'Normative requirements,' *Ratio*, 12: 398-419.
- (2013). *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.
- Carroll, L. (1895). 'What The Tortoise Said To Achilles,' *Mind*, 4: 278-80.
- Christensen, D. (2004). *Putting Logic in its Place*. Oxford: Oxford University Press.
- Darwall, S. (1983). *Impartial Reason*. Ithaca: Cornell University Press.
- DeRose, K. (1999). 'Introduction: Responding to Skepticism,' in DeRose & Warfield (eds.), *Skepticism: A Contemporary Reader*. Oxford: Oxford University Press.

⁴⁴ This already errs in representing the entailment in question merely as a *conditional* like the original (if *p* then *q*) premise.

⁴⁵ Contrast Harman (1986: 7-8), who arguably leads us into a tortoisey mistake when he claims that when one affirms the consequent, one is only making a mistake in the beliefs that one begins with (namely, a belief about what implies what), rather than in proceeding "in accordance with an incorrect rule of revision".

⁴⁶ Similarly, Wedgwood (2013) argues that the notion of correct belief plays an important role in helping us to understand justified and rational belief derivatively.

- Field, H. (2009). 'What is the Normative Role of Logic?,' *Proceedings of the Aristotelian Society Supplementary Volume*, 83: 251-68.
- Fogal, D. (2015). *Bad Attitudes: Rationality and its Discontents*. Doctoral dissertation, New York University.
- Foley, R. (1993). *Working without a Net*. Oxford: Oxford University Press.
- Gibbard, A. (2005). 'Truth and Correct Belief,' *Philosophical Issues*, 15: 338-350.
- Greenspan, P.S. (1975). 'Conditional Oughts and Hypothetical Imperatives,' *Journal of Philosophy*, 72/10: 259-276.
- Harman, G. (1986). *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.
- Hussain, N. (ms.). 'The Requirements of Rationality,' draft manuscript, Stanford University.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kelly, T. (2010). 'Peer Disagreement and Higher Order Evidence,' in Feldman & Warfield (eds.), *Disagreement*. Oxford: Oxford University Press.
- Lasonen-Aarnio, M. (2008). 'Single premise deduction and risk,' *Philosophical Studies*, 141/2: 157-173.
- (2014). 'Higher-Order Evidence and the Limits of Defeat,' *Philosophy and Phenomenological Research*, 88/2: 314-45.
- (forthcoming). 'Enkrasia or evidentialism? Learning to love mismatch,' *Philosophical Studies*.
- Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge, UK: Cambridge University Press.
- MacFarlane, J. (ms.). 'In What Sense (If Any) Is Logic Normative For Thought?,' draft manuscript, University of California, Berkeley.
- McDowell, J. (1998). *Mind, Value & Reality*. Cambridge, MA: Harvard University Press.
- Pettit, P. (2007). 'Rationality, Reasoning and Group Agency,' *dialectica*, 61/4: 495-519.
- Pryor, J. (2013). 'Problems for Credulism,' in Tucker (ed.), *Seemings and Justification*. Oxford: Oxford University Press.
- (ms.). 'The Merits of Incoherence,' draft manuscript, New York University.
- Sainsbury, R.M. (2002). 'What logic should we think with?,' *Royal Institute of Philosophy Supplement*, 51: 1-17.
- Scanlon, T.M. (2007). 'Structural Irrationality,' in Brennan, Goodin, Jackson & Smith (eds.), *Common Minds: Themes from the Philosophy of Philip Pettit*. Oxford: Oxford University Press.
- Schechter, J. (2013). 'Rational Self-Doubt and the Failure of Closure,' *Philosophical Studies*, 163/2: 428-452.
- (this volume). 'Small Steps and Great Leaps in Thought: The Epistemology of Basic Deductive Rules.'
- Shah, N. (2003). 'How Truth Governs Belief,' *Philosophical Review*, 112/4: 447-482.
- Sturgeon, S. (2008). 'Reason and the Grain of Belief,' *Noûs*, 42/1: 139-165.
- Sutton, J. (2007). *Without Justification*. Cambridge, MA: MIT Press.
- Weatherson, B. (ms.). 'Do Judgments Screen Evidence?,' draft manuscript, University of Michigan.
- Wedgwood, R. (2002). 'The Aim of Belief,' *Philosophical Perspectives*, 16: 267-297.
- (2012). 'Justified Inference,' *Synthese*, 189: 273-295.

- (2013). 'Doxastic Correctness,' *Proceedings of the Aristotelian Society Supplementary Volume*, 87, 217-234.
- Way, J. & Whiting, D. (2016). 'If You Justifiably Believe That You Ought to Φ , You Ought to Φ ,' *Philosophical Studies*, 173/7: 1873-1895.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
- Worsnip, A. (2015). 'Possibly false knowledge,' *Journal of Philosophy*, 112/5: 225-46. (2016).
- (2016). 'Belief, credence, and the preface paradox,' *Australasian Journal of Philosophy*, 94/3: 549-562.
- (forthcoming). 'The conflict of evidence and coherence,' *Philosophy and Phenomenological Research*. DOI: 10.1111/phpr.12246