

## VARIETIES OF INFERENCE?

Anna-Sara Malmgren  
Stanford University

It's orthodoxy in epistemology that the distinction between inferential and non-inferential justification is strictly dichotomous: a justification is inferential or non-inferential but not both (and, barring over-determination, a justified belief is justified inferentially or non-inferentially but not both). In this paper I argue that this can't be right—at least not on standard extant views of that distinction. I'll describe some cases of *prima facie* justified belief that don't fit neatly on either side of it. There's pressure to think of them as inferentially justified. There's pressure to the contrary as well, since it's at best unclear how certain widely held requirements on doxastic inferential justification could be satisfied in these—henceforth, the 'hard'—cases.

One might respond to the tension by concluding that these requirements have to go (or at least be modified or reinterpreted) but it's unclear what to put in their place (or what a plausible modification or reinterpretation would look like). Or by concluding that the inferential/non-inferential distinction isn't, contrary to orthodoxy, exclusive and exhaustive. But whichever of these options is decided on, a substantive challenge then lies ahead: to carve out space for the hard cases—articulate a new framework that allows us to explain the rationality of these beliefs. Towards the end I sketch the direction in which my response to this challenge goes. First I discuss the considerations that motivate it. (This takes up the bulk of the paper.)<sup>1</sup>

In §1, I outline (what I take to be) the received view of the inferential/non-inferential distinction. In §2, I introduce the hard cases. In §3–4, the opposing pressures are brought out. In §5, I outline my positive proposal.

Preliminaries. I take the distinction between inferential and non-inferential *knowledge* to be parasitic on the parallel distinction between justification-types. I use 'inferential/non-inferential', where some use

‘mediate/immediate’, ‘non-basic/basic’, or ‘derived/foundational’. I use ‘rational’ and ‘justified’ interchangeably; likewise, for ‘(good) reason’ and ‘evidence’. I write as though reasons are propositions—rather than, say, *true* propositions or states of affairs. (But much of what follows is recoverable on views that treat reasons as factive.) Propositions are assumed to be relatively fine-grained, but nothing (else) is taken for granted about their metaphysics.

I also assume that justification comes in degrees, and that most if not all justification is defeasible. But I write as if the bearer of justification is full or categorical, rather than graded, belief (or credence). The rationale for that is two-fold: the widespread view of the inferential/non-inferential distinction I want to engage is typically formulated in terms of (justification for/to have) full belief. This may not be essential to it: if credences map on to full beliefs in some straightforward way, we should be able to set up the present problematic in degree-theoretic terms (without substantively affecting the discussion). If they *don't*, it might do more harm than good to try set things up that way. The issue would turn on whether there's (nevertheless) a true description of the actual human mind that trades in credal states—states that are governed by epistemic norms we already follow, when we're rational. That's because the focus here is on certain *realistic* cases, featuring ordinary human agents (in the actual or nearby world/s).<sup>2</sup>

## 1.

### 1.1.

The picture of inferential justification I'm working with—henceforth, the ‘standard conception’—is a skeletal or generic account that, terminological disparities aside, subsumes most (perhaps all) more specific extant accounts in the literature. (An equally skeletal, negative, account of non-inferential justification falls out of it.) The conception tends to serve as the shared starting-point in various contexts where the inferential/non-inferential distinction is discussed or put to work.<sup>3</sup>

It can be crudely summarized thus:

*An inferential justification partly consists in a set of propositions (and one or more inference rules) that together make up a good argument for a given conclusion. An agent has that justification, to believe the conclusion, only if she has antecedent justification to believe each of those propositions. She's justified in believing the conclusion, on the basis of that justification, only if she's antecedently justified in believing each of the propositions that constitute it—and her belief in the conclusion is suitably based (on her beliefs in those propositions).*

More precisely, the standard conception comprises a set of necessary conditions—along the lines of (a)–(g) below. There's debate over what *else*,

if anything, inferential justification requires; over how to cash out some of the theoretical notions that give substance to the conception; and over the scope of inferential and non-inferential justification respectively. (Notably over whether the latter is instantiated at all.) There's also debate over the correct positive characterization of non-inferential justification. But since it's generally assumed that the distinction is dichotomous, non-inferential justification can at least be given an initial negative gloss as justification that's not inferential.

- (a) S has ('propositional') inferential justification to believe  $p$  only if her justification to believe  $p$  at least partly *rests* on her justification to believe at least one other proposition,  $q$ .
- (b) S is inferentially justified in believing  $p$  (has 'doxastic' inferential justification) only if her justification to believe  $p$  at least partly rests on her justification to believe  $q$ —moreover: S *believes*  $p$ , and her belief that  $p$  is *based* on her justified belief that  $q$ .

There are non-circularity constraints on both resting and basing, an explanatory constraint on basing, and a structural or formal constraint on the propositions involved. Perhaps some or all of these are already implicit in (a)–(b), but they're worth spelling out. (Many take the explanation-type involved in basing to be causal, but we don't need to build this in to set up the problem.)

- (c) S's justification to believe  $p$  rests on her justification to believe  $q$  only if her justification to believe  $q$  doesn't in turn rest on—is *antecedent* to—her justification to believe  $p$ .
- (d) S's (justified) belief that  $p$  is based on her (justified) belief that  $q$ , only if her belief that  $q$  isn't in turn based on her belief that  $p$ .<sup>4</sup>
- (e) S's belief that  $p$  is (partly) based on her belief that  $q$  only if S's believing  $q$  (partly) *explains* her believing  $p$ .
- (f) S's justification to believe  $p$  rests on her justification to believe  $q$ , only if  $q$  (inferentially) *evidentially supports*  $p$ —alone, or in conjunction with other propositions,  $r \dots r_n$ , that are *available* to S.

I'll sometimes refer to (f) as the 'structural requirement'. Note that it has two parts: first, the requirement that  $q$  evidentially support  $p$  (on its own or together with auxiliary propositions  $r \dots r_n$ ); second, an embedded constraint on the relationship between S and  $r \dots r_n$  (in cases where  $r \dots r_n$  are indeed implicated)—here provisionally labeled 'availability'.

The way that the embedded constraint is in turn cashed out is important:

- (g) for propositional (inferential) justification,  $r$  is available to S only if she has *antecedent justification to believe  $r$* ; and, for doxastic justification,  $r$  is available to S only if she's *antecedently justified in believing  $r$* .

I'll refer to (g) as the 'standard availability constraint', or the 'availability constraint' for short. And (a) through (g) make up the standard conception, as I understand it.

## 1.2.

Some clarifications. The structural requirement, (f), is effectively an instance of the more general constraint that, for a proposition  $q$  to be (good) reason or evidence to believe another proposition,  $p$ ,  $q$  must stand in an appropriate logical—or, more broadly, implication or confirmation—relation to  $p$  (see e.g. Davidson 1986; Pryor 2005). On my usage, any relation that qualifies, for a given type of justification, instantiates the 'evidential support-relation', for that type. One that qualifies for inferential justification in particular instantiates the 'inferential evidential support-relation.' A 'good (inference) rule', in turn, is just a rule that encodes some such relation, and a 'good (argument) form' is the corresponding form (or scheme or pattern)—roughly: a relatively abstract template for particular arguments, at finer levels of grain, which exploit or incorporate that rule.<sup>5</sup>

Note also that I think of an argument as partly individuated by the rule(s) it exploits; specifically: as a set or sequence of propositions—at least one of which is a conclusion and the rest of which are premises—together with one or more inference rules or (candidate) support-relations. On another construal, arguments merely consist in sets of propositions, divided into conclusions and premises, that can be evaluated in different ways—by the standards of (some or other) deductive logic, the standards of (some or other) formal system of inductive implication, etc. The richer notion is more convenient for present purposes, but there's nothing of substance at stake here.<sup>6</sup>

More importantly: a 'good' argument, in the intended sense, needn't *guarantee* inferential justification for an agent with the requisite relation to its premises (in situations where the other requirements above are met).<sup>7</sup> A good argument is just one that's capable of partly constituting justifications of this type. That requires it to have a form that reflects an inferential evidential support-relation between premise- and conclusion-propositions. But what does that come to? Which argument forms qualify?

There's no agreed-on, principled answer to this question in the literature. It's also unclear how to approach it—whether it can be settled independently

of certain other elusive questions about how to unpack the standard conception (see Malmgren ms(a) and below). But there's at least a modest amount of convergence on some *prima facie* paradigms (see e.g. Audi 1986, 2011; Boghossian 2008; Tucker 2012; Schechter forthcoming; Wedgwood 2012). The typical paradigm list includes some relatively simple, deductively valid forms—e.g. arguments by modus ponens and modus tollens, by existential introduction and elimination, conditional proof and reductio—as well as some forms that are broadly inductive—e.g. arguments by statistical-enumerative induction, to both particular and general conclusions, and by abduction. Perhaps the most prominent rival stance is that *only* abductive arguments can constitute inferential justifications (e.g. Harman 1965, 1973; Lycan 1988). But the difference between this and the more permissive view won't matter here.

There's also an intriguing but largely neglected corollary: how to *represent* (or specify or explicate) a given argument form—in how much detail, and with what internal structure—relative to the epistemic and psychological demands imposed by the availability constraint, (g). (E.g. a representation can be more or less fine-grained, and pack more or less informational content into premises and rules respectively.) Unless (the proffered) argument forms are individuated so finely that no representational variation is possible, the corollary question comes up however we answer the first—i.e. whatever argument goodness requires—and the coarser they're carved, the more potential variation there is, some of which makes a difference to the precise demands that (g) issues. We'll touch on this again in §4.

Note that (g) concerns the availability of a proposition to play a certain contributory role in an agent's inferential justification: the role specified in (f). It doesn't amount to or alone entail, say, the requirement that—to play that role—a proposition must be reflectively accessible to her. (Or that she must believe, or in some other way 'take', the proposition to be her reason.) It just says that she must have antecedent justification to believe the proposition, or, for doxastic justification, an antecedently justified belief in it.

Last, as the paradigms above make clear, (g) needs at least one upfront proviso or fix since some seemingly good arguments have premises that aren't appropriately believed at all—only supposed, or conditionally believed, until discharged (cf. Dogramaci 2013; Wright 2014). But nothing here turns on this, so we can safely restrict the ensuing discussion to broadly axiomatic arguments: arguments without such premises.<sup>8</sup>

### 1.3.

How do we—how should we—*apply* the inferential/non-inferential distinction to undecided cases? How settle whether a *prima facie* justified belief—token or type, actual or hypothetical—falls on one or the other side

(going along, for now, with the prevailing assumption that these are the only non-skeptical options)?

What I call the ‘dual asymmetric dependence test’ is a defeasible first-round test for (doxastic) inferential justification: one I think we commonly already use, in our initial classification of cases, and that integrates nicely with the standard conception—in particular with (a)–(d). The test breaks down into one for (non-circular) resting and another for basing, but we can simplify away from this.

In brief: if the resting- and basing-relations are both asymmetric (when things go well), we should expect an inferentially justified belief to asymmetrically counterfactually depend on other justified beliefs of the agent, other things equal. More specifically, that (when S is inferentially justified in believing *p*, there’s at least one other proposition *q* such that) *ceteris paribus*, if S didn’t have justification for believing *q*, S wouldn’t have justification for believing *p*, but not *vice versa*, and if S didn’t believe *q*, S wouldn’t believe *p*, but not *vice versa*. (Simplifying: *ceteris paribus*, if S weren’t justified in believing *q*, S wouldn’t be justified in believing *p*, but not *vice versa*.<sup>9</sup>) If that’s right then, conversely, this type of dual dependence fallibly indicates that S is inferentially justified—or, better: that her justification for believing *p* (non-circularly) rests on her justification for believing *q*, and that S’s belief that *p* is (non-circularly) based on her belief that *q*. I’ll sometimes express this by saying that S’s justified belief that *p* ‘inferentially depends’ on her justified belief that *q*. (The dual asymmetric dependence test is, in the first instance, a test for inferential dependence. And note that we can recognize the existence of inferential dependence without committing to its coextension with—or even the existence of—inferential justification, as captured by the standard conception as a whole.)

The *ceteris paribus* clause is important. There are a number of alternate ways for the asymmetric counterfactuals to be sustained—e.g. a mischievous demon, or other spurious agent or mechanism, could interfere. Less fantastically: the counterfactuals could be grounded in certain hierarchical relations between belief-contents, or, arguably, in relations of ‘enabled-on-enabler’ dependence—i.e. the kind of dependence that a fact (or state/event/condition) has on the facts (states/events/conditions) that merely enable it, as opposed to those that explain it. (More on this shortly.) It also seems clear that inferential justification or dependence could occur without the asymmetric counterfactuals: again because of interfering agents or mechanisms, or certain defeater patterns, or simply due to over-determination of (or readily available backup-routes to) justification and/or belief. These are all circumstances where other things aren’t equal.<sup>10</sup>

But the modest role I want to grant the dual asymmetric dependence test is just this: if an agent and a set of otherwise suitable propositions pass it—and no plausible competing diagnosis suggests itself—it’s reasonable to

take the hypothesis of inferential justification (or dependence) seriously, and proceed to investigate it further.

How do we do *that*?

A natural suggestion—still within the current framework—is to consider whether the remaining requirements that make up the standard conception are satisfied in the case under consideration. (If the case is hypothetical: whether it plausibly could, or would, be realized in such a way that they're satisfied.) That is: whether it's plausible that the candidate base belief, the belief that  $q$ , explains the other, the belief that  $p$ —and, crucially, that there's a good argument from  $q$  to  $p$ , each of whose (non-redundant) premises are believed with antecedent justification by the agent.

This strategy is complicated by the fact that the conception is underspecified with respect to the question of argument representation, mentioned above—what I call the problem of 'relevant completeness' (Malmgren ms(a), ms(b); see also §4)—and by the absence of an agreed-on general criterion for argument goodness. It seems that all we have to work with, as regards goodness, is a small selection of *prima facie* paradigms, from which it's contentious how to generalize. And how a given argument form is represented can impact what exactly (g) requires of the agent, because it partly determines what counts as the full set of essential premises of arguments with that form.

In addition: it's sometimes unclear how even to *initially* regiment the argument that's at play in a given case—e.g. which familiar argument form, if any, on the paradigm list, to assimilate it to. As we'll see, the hard cases are cases in point.

## 2.

### 2.1.

I'll introduce the hard cases by example(s). The first three are my primary case studies.<sup>11</sup>

BERTRAND	Bertrand is presented with a neutral description of a classic Gettier-case, and he makes a judgment that's naturally expressed by saying (something like) 'Smith has a justified true belief without knowledge!'
PEGGY	Peggy's field-assistant provides her with a neutral description of a certain exotic bird in flight. Peggy judges that that was a goat-eating nutcracker.
ZINAR	Zinar is cooking, when Elif utters: 'ez birçi me.' (Kurmanji for 'I'm hungry'.) Zinar comes to believe that Elif said that she's hungry, and/or that she asked to be invited for dinner.



To a first approximation, a ‘neutral’ description (here: of a philosophical problem-case and token bird respectively) is one that doesn’t transparently entail a verdict on the question that the judgment in the case offers an answer to.<sup>12</sup> (Perhaps Bertrand reads one of the original case-descriptions in Gettier 1963. Perhaps Peggy’s assistant calls out something of the form: ‘bird with such-and-such crown stripe, and such-and-such wingspan, overhead!’) I take a ‘judgment’ to be an occurrent or episodic belief. Nothing here hinges on how exactly this distinction is understood (or on which—judgment or belief—is featured in a given case). A complication is however introduced by the fact that the agent in the first example (BERTRAND) might confront a manifestly *hypothetical* scenario. If so, his apparent judgment about it can’t be taken at face value—perhaps it’s not a genuine judgment, or its ‘real’ content isn’t what it seems to be.<sup>13</sup> But this complication will be bracketed. None of the substantial points I’ll be making turn on the simplifying assumption that Bertrand responds to a problem case he takes to be actual. Note also that the characterization of ZINAR is meant to leave open whether Zinar forms two distinct beliefs (e.g. one about what’s said, and one about what’s implicated) or just one belief (about what’s overall communicated). The case can be used for its intended purpose regardless.

To give a sense of the broader scope, here are a few more examples:

- Mae reads that the prisoners were tortured, and she judges that they were wronged.
- Vernon exclaims: ‘I finished the book!’ Clive comes to believe that Vernon just said that he completed his memoir (and perhaps: that Vernon completed his memoir).
- Polonius watches Ophelia do the dishes—again. He judges that Ophelia is depressed.

## 2.2.

All of the examples I’ve given are heavily under-specified and, with respect to many details, we can leave them that way. But some details matter.

First, I’d like the reader to suppose that, in each case, the salient judgment or belief is ‘relatively unreflective’—i.e. preceded by minimal conscious reflection or none (where that includes any deliberation or other movement of thought that’s conspicuous in the agent’s phenomenology). Second, suppose that, even in retrospect, the proximate etiology of the belief is largely ‘subjectively opaque’—i.e. that the agent is unable to tell, by ordinary introspective and/or self-reflective means, how it was formed, or why they came to hold it. (A bit more precisely: those means don’t suffice to identify



a complete or satisfying causal-psychological explanation of the belief. At best, they suffice to identify what looks like a *fragment* of an explanation.) Third, suppose that the epistemic ground of the belief is largely opaque as well—i.e. that the agent is unable to tell, by ordinary introspective and/or self-reflective means, for what reasons (if any) they came to hold it, or what (if anything) makes it rational for them to do so. (Again, more precisely: those means don't suffice to identify a complete or satisfying justification for it. At best, it delivers what looks like a fragment of a justification.)<sup>14</sup>

The starting-point of my inquiry is that there are cases, actual and/or in nearby worlds, that are relevantly like the cases outlined here, with the further specifications—and in which the salient judgments are *rational* or *justified*, to at least some degree and at least *pro tanto*. What explains that? How—in virtue of what—are the judgments justified? The agents themselves can't tell us; at any rate they can't tell us much. That's one reason why the hard cases are interesting to me.

On certain theories, the subjective opacity of even part of the epistemic ground of a belief precludes it being justified (on that ground). Adherents of such theories might reject my starting-point, but they're welcome to treat the arguments that follow as conditional for now.<sup>15</sup> On other theories, the question of what justifies the target judgments has a ready answer: they're justified—if they are—in virtue of being produced by reliable processes. Reliability-considerations do play a role in the account I eventually favor. But I think that the relationship between justification and reliability, here and in general, is more complex than the ready answer assumes.

It would presumably be a mistake to expect a unified and *comprehensive* answer to the question just asked—a detailed account of what justifies the judgments in (actual or nearby realizations of) all these cases, or even a significant subset of them. The cases are simply too heterogeneous for that to be feasible. (Better: some of the overt differences between them line up with well-entrenched and *prima facie* useful epistemological distinctions.) But I'm not aiming for an answer of that kind. What I want to explore is just whether there are any interesting generalizations to be drawn at this, relatively high, level of abstraction—generalizations that may in turn guide or constrain the first-order epistemology: that affect the overall shape of plausible accounts of what justifies beliefs otherwise importantly like Bertrand's, of what justifies beliefs otherwise importantly like Peggy's, etc.

The fact that the target judgments have a number of other notable features in common suggests that there may indeed be some such generalizations.

In brief: in each of the hard cases, (i) the target judgment is relatively unreflective; (ii) the judgment's proximate etiology, and its epistemic ground, are largely subjectively opaque—moreover, (iii) the judgment incurs (what I call) an 'r-commitment'; (iv) the dual asymmetric dependence test comes out positive; (v) it's very hard to see how the availability constraint and the structural requirement could both be satisfied; and (vi) the proximate

causal-psychological explanation of the judgment doesn't fit squarely on either side of the divide between the personal and sub-personal level of explanation (as *that* distinction is standardly conceived).<sup>16</sup>

I won't say more about (i)–(ii) here. (iii) stands in need of explanation, but isn't by itself controversial: r-commitments aren't distinctive of the hard cases. (The diagnosis I favor, a little more so; see §4.) (iv) gives the judgments a *prima facie* claim to being inferentially justified—a claim that's strengthened by the fact that it's unclear how they could be *non*-inferentially justified (see §3). But (v) counts against that suggestion. And (vi) points the way forward: towards a new framework with the resources to explain what makes the target judgments justified. Or so I argue (see §5 and, for the full proposal, Malmgren ms(b)).

### 3.

#### 3.1.

In some respects, the target judgments resemble canonical instances of non-inferentially justified beliefs—notably in that they're relatively unreflective, and their etiology and epistemic ground largely opaque. But it's unclear how to extend any promising extant model of non-inferential justification to the hard cases (and what a new model that fits them would look like).

I can't examine all the possibilities here, but let's canvass a couple.

For one thing, there are no perceptual states with appropriate contents in the hard cases—appropriate, that is, to justify the target judgments. (At least not on the widespread assumption that the evidential support-relation for perceptual non-inferential justification is identity.) The point doesn't turn on a restrictive view about content: even if perceptual contents can be sufficiently rich—e.g. contain concepts like knowledge or goat-eating nutcracker—the agents simply aren't in a position to entertain those contents in the right mode. Bertrand doesn't see or hear what Smith (or what someone/anyone playing the 'Smith-role') is doing, or indeed *not* doing, Peggy doesn't see or hear the bird, etc. So it doesn't seem like their judgments could be perceptually non-inferentially justified, whatever the best account of such justification might be.

It might be proposed that the cases contain *quasi*-perceptual states with appropriate contents: states of the type invoked on certain models of non-perceptual non-inferential justification—what's sometimes called 'intuitions' or 'intellectual seemings' (see e.g. Bealer 2002; Bonjour 1998). My general misgivings about this kind of view aren't original, and perhaps best expressed as a challenge—a challenge to make a persuasive case that there's a distinct psychological state, present in a large enough range of cases, that can do the requisite explanatory work (see e.g. Schechter forthcoming). The challenge

seems to me still open, but be that as it may: extant versions of the model clearly don't have the scope that's called for here. Some of its advocates might be willing to take a more permissive approach—that covers, say, cases like PEGGY and actual realizations of BERTRAND (as well as realizations featuring manifestly hypothetical problem-cases). But to my knowledge nobody has done so yet. And note the face-value tension with the restriction to non-perceptual—*a priori*—non-inferential justification.<sup>17</sup>

The cases could of course come apart in this regard; that's one of the dimensions along which they may well vary. The situation, then, would seem to be this: at best there's an argument available 'from the trenches'—from a suitable positive account, with the slightest initial appeal—for classifying *some* of the hard cases as cases of (non-perceptual) non-inferential justification. That still leaves the rest. I'd be content with that: I just want to understand them all better. But suppose, furthermore, that the framework I outline later is along the right lines—in the first instance in application to the cases that remain. We then have an argument from the trenches for a contrasting verdict, too (since that framework covers all of them): the verdict that the justification exemplified in the hard cases is not non-inferential, but either inferential (on an overhauled conception of that category) or neither. The relative overall merits of the competing positions, once developed in detail, would have to decide between them. For now, just note that it's also incumbent on the opponent—the advocate of the quasi-perceptual model—to explain away the dual asymmetric dependence in the cases her model *is* supposed to cover.

### 3.2.

The dual asymmetric dependence test returns a positive result in the hard cases—provided we screen off over-determination, interfering demons, and other manifest noise. (This can be done without begging the question.) In other words: the target beliefs, in nearby realizations of the hard cases, asymmetrically depend—for their existence as well as for their justificatory status—on the agents' having certain other justified beliefs. The sketches below should convey the idea; in particular, what (other) beliefs I have in mind. (It's inconsequential that some of the candidate bases may break down in multiple beliefs.)

BERTRAND<sub>AD</sub>      If Bertrand weren't justified in believing that Smith *satisfies the case-description*, he wouldn't be justified in believing that Smith has a justified true belief without knowledge—but not vice versa.

PEGGY<sub>AD</sub>            If Peggy weren't justified in believing that *that* [bird/animal/object] *met her assistant's description*,

she wouldn't be justified in believing that that [bird/animal/object] was a goat-eating nutcracker—but not vice versa.

ZINAR<sub>AD</sub>

If Zinar weren't justified in believing *that Elif uttered 'ez birçî me'*, he wouldn't be justified in believing that Elif said that she's hungry, and/or that she'd like to be invited for dinner—but not vice versa.<sup>18</sup>

The reference to (the satisfaction of) descriptions, in the contents of Peggy's and Bertrand's candidate base beliefs, is just shorthand for whatever feature-cluster is attributed to the bird and to Smith, by the assistant and the author of thought-experiment respectively. There are various ways to fill out the details here, while retaining the intended structure of each case (as specified in §2.2). And for brevity I still write as if the target belief in BERTRAND is 'just what it seems to be', but the point made earlier applies equally to the candidate base belief(s).<sup>19</sup>

One might try to explain away the result in terms of enabled-on-enabler relations. Indeed, this seems to be the only competing diagnosis that's worth taking seriously (unless some crucial source of false positives has been overlooked). The possibility of spurious agents or mechanisms grounding the dependences can presumably be ignored, since the realizations of the hard cases that we're interested in are nearby or actual.

To illustrate the thought with a relatively uncontentious case: take the dependence that my writing this paper has on there being oxygen on Earth. Against a normal contrast class, and in otherwise normal circumstances, there being oxygen on Earth doesn't explain why (or how or that) I'm writing this paper. But I wouldn't be writing it unless there were in fact oxygen on Earth, and the converse is false.<sup>20</sup> If examples like this survive scrutiny, it looks like enabled-on-enabler dependences can indeed sustain asymmetric counterfactuals. So far so good.

What needs to be accounted for, in the hard cases, isn't just the asymmetric dependence of belief-on-belief, but of justification-on-justification. So the suggestion would be that Bertrand's believing that Smith satisfies the case-description enables—rather than explains—his believing that Smith has a justified true belief without knowledge, and Bertrand's (having) justification to believe the former proposition enables—rather than explains—his (having) justification to believe the latter. (Similarly: that Peggy's believing that that [bird/animal/object] meets her assistant's description enables—rather than explains—her believing that that's a goat-eating nutcracker, and her justification to believe the former enables—rather than explains—her justification to believe the latter; etc.)

But note, first of all, that it's *independently plausible* that there's an explanatory relation—specifically, a causal-psychological explanatory

relation—between the one belief and the other, in each of the hard cases. That is: that the target belief (e.g. that Smith has a justified true belief without knowledge, or that Elif said that she's hungry/wants to be invited for dinner) is causally explained at least partly by the belief on which it asymmetrically depends (here, that Smith meets the case description, and that Elif uttered 'ez birçi me'). The point can be further substantiated by appeal to characteristic marks of causal explainers and enablers respectively—e.g. the relative stability of explanations, across actual and counterfactual variation in enabling conditions, and the minimal predictive force of enablers, compared to that of explainers.<sup>21</sup>

Of course: that there's a causal-explanatory relation between the two beliefs doesn't yet show that the one is based on the other, even if the basing-relation entails a causal-explanatory relation of that kind. But it casts doubt on the claim that the one (the candidate base belief) enables *rather than explains* the other (the target belief). By the same token, it strengthens the appeal of the diagnosis in terms of an inferential dependence.

A fact or event can count as an enabler of a given effect under one (e.g. more specific) description, and as an explainer of the same effect under another (e.g. more general) description. My writing this paper is my doing something. That I narrowly escaped the bomb might explain why I'm doing something—doing anything at all—whereas, sadly, Bettie isn't. At the same time, it might enable—but not explain—my writing this paper (rather than writing another paper, or training for the Olympics). Perhaps it's also possible for a fact or event to enable an effect, under a fixed description, relative to one set of interests and concerns (or against one contrast class) while explaining the effect under that description relative to another. (That I narrowly escaped the bomb might enable—but not explain—my doing something, rather than still suffering from paralyzing shock.) Thus it might be argued that, in the hard cases, the candidate base beliefs do explain the target beliefs, but that they *also* enable them—perhaps under another description, or relative to other interests and concerns.

In the present dialectical context, however, that's a non-starter. Once we've granted that there's an explanatory relation between the target and candidate base beliefs, there's no *need* to invoke an enabled-on-enabler dependence to account for the asymmetric counterfactuals. There could still *be* some such relation—and there might be independent reasons to think so—but the point can no longer be used to explain *away* the asymmetric dependences. (The matter might simply be over-determined.) It also seems clear that the burden of proof is on the dissenter's shoulders: to specify the (other) descriptions of the target beliefs, or the (other) interests or concerns, that supposedly underwrite the claim that the candidate base beliefs function as enablers (as well as explainers) in the hard cases. And those descriptions or concerns had better be congenial to the inquiry into the epistemology of the target beliefs—equally or more so, than the descriptions or concerns that

underwrite the candidate base beliefs' explainer status. It's not at all clear what they would be.

The dissenter must moreover make good on the second half of their diagnosis: the claim that the agent's (having) justification for the target belief is enabled—rather than explained—by her (having) justification for the candidate base belief. How is this to be understood? The resting relation would seem to be explanatory too—but not causal-explanatory. Rather, it is (or implies) a constitutive or metaphysical 'part-to-whole' account of a certain normative phenomenon—viz., an agent's justification to believe a certain proposition (here, the content of the target belief). If that's right, the suggestion becomes that, say, Bertrand's justification to believe that Smith satisfies the case description enables—rather than *constitutively* explains—his justification to believe that Smith has a justified true belief without knowledge. Ditto for the other examples.

Most discussion of the enabler/explainer distinction appears in debates about causation and causal explanation, but we can arguably make sense of enablers relative to other explanation-types as well. Yet, it's much harder to find uncontentious examples—in particular: examples where both terms of the enabler-enabled relation are (even partly) *normative*.<sup>22</sup> Perhaps the most promising source is the literature on warrant-transmission-failure and easy knowledge (see e.g. Cohen 2002; Pryor 2004; Silins 2005; Wright 2004, 2007). It's striking, however, that all the potentially analogous cases to be found in that literature—cases that some 'conservatives' diagnose as cases where an agent's justification to believe *p* is enabled by her justification to believe *q*—seem to share certain broad structural features that the hard cases lack.<sup>23</sup> Relatedly, the conservative diagnosis is motivated by worries that don't generalize. In brief: it's a response to an apparent forced choice, between skepticism about justification or knowledge of a certain sort, or in a certain domain, and the acceptance of some narrowly circular ('boot-strapping') argument forms as good forms. In the hard cases, no such dilemma presents itself—or, more carefully: none featuring the sets of justified beliefs that are *prima facie* inferentially dependent on one another. We just can't get a parallel bind with intuitive pull off the ground here.

### 3.3.

Suppose, then, that we take the asymmetric counterfactuals in the hard cases at face value: as indicating inferential dependences. How do we assess the overall viability of that claim? Recall the seemingly natural move suggested in §1.3: consider whether the other requirements that comprise the standard conception are plausibly satisfied. (e) has already been addressed in passing. So the issue turns on (f) and (g): on whether the agent in each case—e.g. Bertrand, Peggy or Zinar—is (or would be) antecedently justified



in believing all the non-redundant premises of a good argument, from the identified premise(s) to the target conclusion.

To gain traction on that question, we first need a handle on what familiar forms (if any) the seeming enthymemes at play here are shorthand for: how to complete the ostensibly ‘missing steps’, without which there are no determinate familiar argument forms—let alone recognizably *good* such forms—instantiated in the hard cases at all. There are indefinitely many ways to fill the gaps, in principle (some of which correspond to good forms); that’s not the issue. Which of all these ways capture, or at least approximate, the full arguments that the agents in the intended realizations of these cases use or rely on—if they rely on something with more structure at all?

We certainly can’t ‘read off’ a standard argument form from the enthymeme, in any of the hard cases. Absent a fixed inference rule, a given set or sequence of propositions always underdetermines the form. But a set is sometimes rich enough that it has a natural interpretation, in context—or at least: a manageable range of such interpretations. E.g. it may contain an *easy-to-see structural entailment*, explicit *statistical information*, or a reference to some salient *explanatory relation*, that links the putative premise and conclusion propositions in a perspicuous way. In none of the hard cases is the so-far identified set of propositions this informative.

What other resources are there? Sometimes the agent of interest, or her local environment, supplies further clues. (Depending on the domain of inquiry: a particular agent and context—e.g. oneself here now, or Petronella in the woodshop last Tuesday—or the canonical agent and context for justification of a given type.) The clues might consist in various pieces of broadly empirical information about her, her circumstances, and behavior—e.g. what she read this morning, or what’s in her line of vision now. They might also include information about what (else) of seeming relevance goes on in the agent’s mind—e.g. what else she believes, hears or remembers, or what she takes to be a reason for which she did such-and-such—information furnished by ordinary introspective or self-reflective means (when she’s us), and by testimony (when she’s not).

These clues are clearly fallible: we self-deceive and confabulate, testifiers lie, and overt behavior is notoriously coarse. But I take it that the practice of consulting them *sometimes* delivers: that there are cases where we do learn more, by relying on such clues—enough, in fact, to make it reasonable in context to assimilate an ostensible enthymeme to one rather than another familiar argument form. At any rate I’m happy to grant that here.<sup>24</sup> (If it’s a mistake, what I argue below just has wider application.)

The problem is that the (intended realizations of the) hard cases don’t come with any pointers of this sort, either. In particular: ordinary introspection and self-reflection offers little guidance (since the target judgments are unreflective, and their etiology and grounds largely opaque). By those



means, the agent can at best identify—and perhaps articulate and share—what looks like the very beginning of an explanation of, and/or justification for, her belief.

The strategy I'll propose for making headway—to trace certain rational commitments associated with the target judgments—points to rather complex but determinate bodies of information, that can help turn the enthymemes into arguments of recognizably good standard forms (see §4). However, the resultant arguments all have premises the agents in the hard cases don't—perhaps can't—believe with antecedent justification. This suggests that the availability constraint is violated, hence that the target judgments aren't inferentially justified after all. The moral I eventually draw is that this requirement is too strict.<sup>25</sup> But we need to replace or supplant it; it's not an option to simply do without. (More on this in §5.)

## 4.

### 4.1.

To a first approximation, an 'r-commitment' is the defeasible rational commitment that a particular reason-based judgment (or action) incurs, to upholding a certain wider pattern—viz. to *responding in the same way*, in all *relevantly similar circumstances*. The failure to upholding that commitment is a rationality failure of a distinctive sort: the kind of failure we might suspect someone guilty of if, say, today they judge that it rained last night, ostensibly in response to noting that the streets are wet, but tomorrow they *don't* make that judgment, on noting that the streets are wet, although the two situations seem alike in all relevant respects.

What, exactly, the wider pattern is—which ways count as the same ways, and which similarities are relevant—varies from case to case, and can be very hard to articulate in detail. Perhaps all we're able to achieve, at least by current methods, are partial and/or approximate specifications of the patterns projected by most of our actions and beliefs (what I call their 'implicit generality'). It's unclear what to make of this difficulty. Relatedly, it's unclear what r-commitments *are*: how to understand their source, status, and their relationship to other commitments and demands. But it's arguably a hallmark of reason-based activity that it incurs commitments of this type.

There's a lot more to say about this.<sup>26</sup> But suffice it, for now, that *if* a certain tempting explanation of r-commitments is correct—viz. as particular manifestations of a standing *pro tanto* requirement to *apply reasons consistently*—then the wider pattern associated with a particular belief (or action) reflects a generality in some of the reasons on which that belief (action) is based. To the extent that the pattern can be traced, then, it's a window into features of our operative reasons—perhaps including features that we

have no other way of getting at. This is significant, even if we never manage to articulate the patterns in full detail.

The tempting explanation presupposes that reasons have a certain built-in-generality—at a minimum: that they apply to more than one case, or in more than one set of circumstances. Otherwise there would be no wider pattern to uphold: at least none to be explained by the mandate to apply reasons consistently. (The consistency would already be achieved—the pattern exhausted—by the original judgment/action.) The assumption that reasons have the requisite generality has been challenged (see e.g. Dancy 2004, 2013). And it no doubt needs to be spelled out in turn. (What *is it* for a reason to apply to more than one case?) But this isn't the place to defend or explicate it properly. I'll just note that, to my mind, the fact that the assumption is needed for the above explanation to work is one of the strongest arguments in favor of the assumption. That's because I know of no better, or even comparably good, explanation of this phenomenon. More carefully: no better *sketch* of an explanation—for what I've given is admittedly only a sketch. But it seems by far the most promising sketch in the offing.<sup>27</sup>

Next: we have everyday methods for tracing r-commitments, in our own case and others'—methods sometimes expressed with questions like: '*given that I  $\varphi$ -ed in  $C_1$ , should/may I  $\varphi$  in  $C_2$ ?*', '*what other situations call for  $\varphi$ -ing, in light of the fact that  $C_1$  does/did?*', or (on a certain reading) '*would you/I still have  $\varphi$ -ed, if  $C_2$  rather than  $C_1$  had obtained?*' Sometimes the answers to these questions are obvious; other times not at all. It can be a surprising discovery that one is r-committed to  $\varphi$ -ing (or to not  $\varphi$ -ing) in  $C_2$ . But if one probes conscientiously, and considers sufficiently many *prima facie* germane permutations of  $C_1$ , a pattern typically starts to emerge.

Behavior, past and present, can give clues as well—in the first instance to the shape of the associated dispositions (to conform to r-commitments). We have quotidian as well as more refined—broadly experimental—ways of getting at those dispositions.<sup>28</sup> As with all behavioral evidence, there's potential performance-error to contend with. And r-commitments can be difficult to tell apart from other commitments and constraints. It can also be hard to ignore defeating considerations, and alternate reasons for or against a certain response, in some variations of the original situation. That noise needs to somehow be screened off, when we ask (each other and ourselves) the probing questions—e.g. whether to  $\varphi$  in  $C_2$ , given that we  $\varphi$ -ed in  $C_1$ ,—with the distinctive aim of exposing, or get a fuller picture of, our r-commitments.

Interestingly, however, it seems we're generally able to do this, with some degree of success—at any rate if our manifest confidence levels are anything to go by. E.g. suppose I ask myself '*given that I judged Sven to be Scandinavian, on learning that he's Swedish, should I judge that Agnes is Scandinavian, too, on learning that she's Swedish?*' The (future actual and counterfactual) circumstances where I learn Agnes' nationality that are relevant to answering the question, as intended, clearly don't include ones

where the socio-geographical facts have changed—so as to boot Sweden from Scandinavia—or where I also find out that Agnes is a member of the local Scandinavian society. And I'm confident of this, upfront: that these circumstances are relevantly different overall. Likewise for an indefinite number of other variations on the original situation (and in many other, more complex, examples).<sup>29</sup>

Part of the rationale for the advance note of caution—that perhaps current methods at best deliver partial or approximate specifications—is that there are doubtless places where our confidence runs out: both with respect to the identification of germane variations, and of what, if anything, we're r-committed to doing in a given variation (of the original circumstance). Furthermore, the *mechanics* of the practice are unclear: it's puzzling how it could reliably track certain properties of our operative reasons (including some that it seems we can't otherwise access). A full theory of reasons—their metaphysics and epistemology—would have to account for this. But for now I'll proceed on the assumption that the right theory vindicates, rather than undermines, a tentative trust in the practice.

With the above caveats, then: suppose that the methods outlined here at least sometimes succeed, in revealing aspects of the patterns that our particular beliefs (and actions) r-commit us to upholding, and that, by systematizing the yields, we can arrive at rough-and-ready specifications of those patterns. Suppose, also, that the tempting explanation-sketch of such commitments is on the right track. This is enough to give a clear direction to the next stage of the inquiry.

#### 4.2.

What r-commitments do the target beliefs incur (in realistic realizations of the hard cases)?

To start off: imagine variations of the original case studies with slightly altered local 'triggers', and consider how the respective agents would reasonably be expected to respond to those—given that they respond/ed such-and-so in BERTRAND, PEGGY and ZINAR. E.g. suppose that Bertrand reads a description of an otherwise identical problem-case where the protagonist is female (or a refugee), that Peggy is told of a bird with the same markers the next day (or by another trusted assistant), and that Zinar hears Sosin (or Ayça) rather than Elif utter 'ez birçi me.' It seems clear that the agents' original responses r-commit them to responding in the same (or corresponding) ways in each of these—slightly different but relevantly similar—circumstances: Bertrand to judging that [*female*] Smith has a justified true belief without knowledge, Peggy to judging that this [*other* bird/animal] is a goat-eating nutcracker, etc. If they made contrasting judgments, or threw up their hands, that would jar with their original responses—in a way that's best

explained as the violation of r-commitments. (E.g. it's not plausibly explained as the discord of believing logically inconsistent contents, or of failing to believe the obvious consequences of other things they believe. And the fact that even *suspense* would jar with their original responses shows that it won't help to invoke a more inclusive coherence constraint on contents—one that goes beyond logical consistency.)

To fill out the picture: consider next what, if anything, the agents may or should be doing in situations that depart from the original situation more drastically overall (e.g. variants of PEGGY where a comet is about to hit), and/or in *prima facie* relevant respects (e.g. variants of BERTRAND where the case-description stipulates that Smith has an independent route to knowledge, of ZINAR where Elif manifestly doesn't speak Kurmanji, etc.) In some, the agents are r-committed to responding in the same or corresponding ways; in others, to responding differently; in yet others, to nothing at all.

Eventually we arrive at approximate specifications of the wider patterns that the target judgments r-commit the agents to—perhaps something along the following lines:

- |                        |  |
|------------------------|--|
| BERTRAND <sub>IG</sub> | Necessarily, for any $x$ and $p$ , if $x$ competently but accidentally arrives at a true belief that $p$ , and $x$ has no undefeated defeaters and no other way to know (that) $p$ , $x$ has a justified true belief that $p$ but $x$ doesn't know $p$ .   |
| PEGGY <sub>IG</sub>    | If object $o$ were more similar, with respect to the most heavily weighted features $f \dots f_n$ , to the prototype for <u>goat-eating nutcracker</u> than to any other salient prototype, $o$ would be a goat-eating nutcracker.   |
| ZINAR <sub>IG</sub>    | If (normal, human) subject $x$ , competent with language $L$ , utters a declarative sentence in $L$ that in context $C$ means that $p$ , $x$ is, by making that utterance, saying that $p$ ; and if, in $C$ , the most relevant and accessible speech-act $x$ would be performing by saying that $p$ is a request that $q$ , $x$ is (by saying that $p$ in $C$ ) requesting that $q$ . |

Note that I'm not wedded to any of these approximate specifications in particular, nor to any particular ways in which to sharpen them up. If *anything even roughly in this ballpark* captures the implicit generality of the target judgments, it's hard to see how (f) and (g) could both be satisfied in the case studies. Nor do I assume that, for each example, there's just one pattern that's consistent with the target judgment being rational. (E.g. that there couldn't be a 'happy' nearby realization of PEGGY where the judgment's implicit generality is best approximated with appeal to the statistical information embodied in a prototype—roughly as above—and another such realization where it's best approximated in terms of a causal-explanatory

principle linking perceivable features to species, and/or where a strict conditional with a *ceteris paribus*-clause, rather than a subjunctive, captures the generality better.<sup>30</sup>)

How do we turn these approximations into arguments—preferably *good* arguments—from the previously identified premises to the target conclusions? A natural thought is to simply treat the above conditionals as major premises in suitable arguments of *modus ponens* form. Further auxiliaries are needed, of course. (In BERTRAND, roughly to the effect that if Smith (or any subject  $x$ ) satisfies the case-description then Smith (or  $x$ ) competently but accidentally forms a true belief that  $p$ , and that Smith has no undefeated defeaters—etc.; in PEGGY, roughly to the effect that that [object] is more similar with respect to  $f \dots f_n$  to the prototype for goat-eating nutcracker than to any other salient prototype, and that  $f \dots f_n$  are the most heavily weighted features for that prototype—etc.; in ZINAR, to the effect that Elif is competent with Kurmanji, that in the present context ‘*ex birçî me*’ means that the speaker is hungry—etc.) But we can in principle see how the story might go: how to construct a valid little *modus ponens* argument, for each of the case studies, granted a substantive conditional bridging premise of this sort.

The cost, however, is that the resultant arguments all have premises that, by (g), are unavailable to the agents: premises they don’t believe with antecedent justification (or even have antecedent justification to believe). The conditionals that aim to capture the *r*-commitments, as well as some of the auxiliaries, feature theoretical concepts they may not even possess—e.g. the (relevant) concept accessibility, prototype, and undefeated defeater. Either way, the agents may fail to believe the putative bridging premises. They certainly don’t *need* to believe them—with antecedent justification to boot—for the target judgments to be rational. And it is unclear whether such justification is within reach in all cases. (At least one possible route clearly fails to deliver *antecedent* justification: the route that involves using the contents of the judgment, and a range of verdicts about relevantly similar cases, as data best explained by the generalization above—e.g. BERTRAND<sub>IG</sub>, PEGGY<sub>IG</sub>—or some anticipated precisification of it.)<sup>31</sup>

The patterns projected by the judgments presumably don’t *have* to be specified as conditionals fit to serve as premises in *modus ponens* arguments. We might try to capture them as inference rules (or sets of rules) instead. To illustrate, here’s a ‘rule-ish’ restatement of BERTRAND<sub>IG</sub>.

BERTRAND<sub>IG-R</sub> If you judge that  $x$  competently but accidentally arrived at a true belief that  $p$ , and that  $x$  has no undefeated defeaters and no other way to know  $p$ —perhaps: and the question arises, and *you* have no relevant undefeated defeaters, etc.—then [you may/should] judge that  $x$  has a justified true belief that  $p$  but  $x$  doesn’t know  $p$ . (The rule is defined over any  $x$  and  $p$ .)

This, however, doesn't make a helpful difference: the outlined rule still manipulates contents that Bertrand doesn't plausibly believe with antecedent justification. So the resultant argument still has premises that are unavailable to him (*unless* the relevantly complete representation of that argument omits those premises; see §4.3). Besides, the argument doesn't seem to instantiate a good, familiar, form. Perhaps it instantiates a good *unfamiliar* form. Just note that that thought brings us back to the question of argument goodness, and the hard residual issue left hanging earlier: how to generalize from the paradigms, and/or what a governing principle that unifies them might look like.

#### 4.3.

The remaining variable is relevant completeness. An argument form can be represented in a variety of ways—e.g. at different levels of grain, and with different 'division of labor' between rules and premises—and some of that variation impacts what it takes to satisfy the availability constraint (because it affects what the essential premises of arguments with that form are). Which representation is, or range of representations are, relevantly complete—complete vis-à-vis the application of (g)? It turns out to be very difficult to give a principled answer to this question (too). Here I can't discuss it in detail; I'll just explain how it bears on our immediate concern.

Suppose we cast the approximations as substantive conditional bridging premises in arguments of *modus ponens* form—hence avoid taking a controversial stance on goodness. In stating this option above, I effectively assumed that relevantly complete representations of the arguments would explicitly feature those premises (along with a few others, to ensure structural validity). Of course that's the canonical way that arguments by *modus ponens* are represented, both formally and informally. But perhaps it's not felicitous here! Perhaps the major premise (and any potentially problematic auxiliaries) needn't in fact figure in a relevantly complete representations of the argument(s). Perhaps the *only* premise-propositions are those that the dual asymmetric dependence test uncovered already? Generalizing, the suggestion would be that the relevantly complete representation of an argument by *modus ponens* is *very* light on premises—indeed, on this view, they don't distinguish that form from any other candidate form the argument might exemplify. The inference rule (alone) does that.<sup>32</sup> This may sound too radical, but recall: relevant completeness is just completeness *relative to the demands of the standard availability constraint*.

It's still far from obvious that the radical-sounding (henceforth, 'premise-minimalist') suggestion is defensible. But the main point I want to make here is this: if that suggestion is correct, the availability constraint is arguably met in the hard cases—but it does very little explanatory work, of the sort that



it's meant to be doing. The constraint provides a criterion for when, and a partial explanation of how, an otherwise suitable proposition can play a certain contributory role in an agent's (inferential) justification (see also §5.1). It could still do *some* such work; viz., with respect to the remaining premise-propositions. But that leaves out a lot—notably: it leaves out the complex conditionals that approximate the implicit generality of the judgments, without which the arguments that constitute the (would-be) justifications don't instantiate *prima facie* good familiar forms.

Even if those approximations are too crude (or plain wrong with respect to some details) it should now be clear that there's a problem here. The proposed strategy delivers initial regimentations of the enthymemes that introduce contents the agents in the hard cases don't believe with antecedent justification—unless the premise-minimalist suggestion about relevant completeness is correct. That suggestion may or may not be viable, all-things-considered. But there's a sense in which it doesn't really matter. Suppose that we accept the (premise-minimalist) suggestion, and conclude that (g) is met in the hard cases. It's then incumbent on us to explain the *difference* between those cases and 'unfortunate' ones that are identical as regards (the availability of) the premises in the minimal set—i.e. between an agent who has the requisite epistemic and psychological relationship to the members of that set, but for whom this is the *entire* argument that allegedly constitutes her justification, and an agent whose justification-constituting argument also includes premises corresponding to the implicit generality of her judgment. (Or, given a rule-ish approximation of the pattern: between an agent who uses that rule, and one who jumps straight to the conclusion from the minimal premise-set.) The premises in the minimal set don't comprise a recognizably good argument, as they stand, but they do when conjoined with the additional contents the strategy I've suggested turns up. However: by the standard availability constraint, some of those contents are unavailable in the cases of interest. The premise-minimalist wants to say that the constraint is satisfied—fine; but she owes us a story about *the rest of the argument*.

Alternatively: suppose we opt for richer representations, and conclude that (g) is violated in the hard cases. The premise-minimalist can say that the target judgments are inferentially justified—provided she fills out the standard conception with some or other supplementary requirement, that accounts for the difference just highlighted (and that's plausibly met in these cases). We, on the other hand, can't—at least not without replacing, or radically revising, the availability constraint. But *either way* a new principle is needed: to explain how the additional contents contribute to the agents' justifications. Either way this is a formidable challenge—i.e. whether those justifications are in the end rightly classified as inferential or non-inferential (or neither).

In the final section I give a preview of my own attempt at meeting it.



## 5.

## 5.1.

The availability constraint answers a certain selection problem—'why  $x$ , rather than  $y$  or  $z$  (or nothing at all)?'—and it does so in a principled way, while giving a partial explanation of how that answer could be correct: how it's possible for a proposition with the right formal properties to play a certain contributory role in a given agent's inferential justification: the role for auxiliaries specified in (f). (Available to 'combine' with  $q$  to evidentially support  $p$  for her.) The proposal previewed here is in the first instance intended to supplement, not replace, this constraint. It articulates *another* explanatory selection criterion: one that handles the propositions—or more broadly: structured contents—that capture the implicit generality of the target judgments, whatever exactly those contents are. And that, in its final formulation, is neutral between renditions of that generality as premises or (combinations of premises and) rules—thus avoids at least that aspect of the problem of relevant completeness.<sup>33</sup> The full proposal and defense of it doesn't fit in this paper, but the core ideas are outlined below.

There's more than one way for a proposition to be available in the sense that the structural requirement is after. The standard availability constraint governs one such way; let's call it '*principal* availability'. Perhaps there's a type of justification all of whose propositional constituents have this status. How *much* there is partly depends on how the questions of argument goodness and relevant completeness are answered. (Given a permissive view of goodness, and premise-minimalism about completeness, there might be a lot. Given answers at the opposite extreme, there might be none.) In each of the hard cases at least one proposition is principally available to the agent: the content of the candidate base belief. But—to recapitulate—there's no good familiar inference rule that sanctions a transition straight from that belief to the target judgment; equivalently: the candidate base and target contents don't, as they stand, constitute an argument of a good familiar form. At most they're enthymematic of some such argument. (Barring premise-minimalism.) There are in principle indefinitely many ways to flesh out these enthymemes, some of which yield good forms. But what we're interested in is the shape of the justifications that the agents in (actual/nearby realizations of) these cases in fact possess—not just justifications they could possess, or possess in other contexts, or that some other (e.g. highly idealized) agents do or could possess. So we need a method of initial regimentation that draws on pertinent facts *about those agents*: facts consistent with what else is true of them (as given by the initial stipulations, general limitations of human psychology and physiology, etc.) The method we've pursued is to trace the implicit generality of the target judgments—granting both the difficulty of approaching precise specifications, and gaps in our understanding of the mechanics of the

practice. But if our specification attempts are at least somewhat successful, and the diagnosis of that generality in terms of reasons is right, we now have a (well-motivated) way in.

The ‘missing steps’ the strategy indicates aren’t principally available to the agents. But they’re arguably *collaterally* available—to a first approximation: encoded in certain inaccessible, but inferentially integrated, causally operative mental states (or mechanisms). Not just by further stipulation: it’s independently plausible that the generation and sustenance of belief in us often involves states with this functional profile. This is a broadly empirical claim, but there’s a growing body of evidence to support it. (Better: it’s a presupposition of a lot of recent, fruitful, psychological research.) And it’s plausible, albeit more controversial, that the generalizations that capture the judgments’ implicit generality are collaterally available to the agents in these cases.<sup>34</sup> It’s of course also controversial that this has normative import: that merely collaterally available information can contribute to the (constitutive) explanation of the rationality of belief. I’ll offer the beginning of a defense of that claim below, after clarifying the key notions.

Let’s say that, for doxastic (inferential) justification,  $r$  is collaterally available to  $S$ , relative to  $p$ , iff it’s the content of a sufficiently *inferentially integrated*, but *reflectively inaccessible*, *mind-to-world-directed* representational state—and/or a corresponding processing-mechanism—that figures in the proximate causal explanation of  $S$ ’s belief that  $p$ , on the basis of believing that  $q$ . (This is still rough, but it’ll do for now.<sup>35</sup>)

What’s inferential integration? By way of initial gloss, it’s a tendency to enter into inference-like proximate causal transitions with a variety of beliefs and/or other attitudes of the agent; where an inference-like (causal) transition, in turn, is a transition that’s sensitive to the contents of the states—here, mental states—that constitute the relata (see Evans 1982; Davies 1989; Stich 1978). Sensitive how? One suggestion is that the transition exploits certain structural or formal relations between contents—but it’s unclear how to demarcate the range of relations that’s at issue *here*.<sup>36</sup> Presumably it overlaps in part with the range that instantiate the inferential evidential support-relation, but it’s not plausible that the two coincide. (E.g. it should be possible for inference-like transitions to embody bad arguments, in our sense.) Fortunately, we don’t need to settle this. All the picture I’m about to sketch requires is that the range *includes* implication-relations, broadly understood—relations that correspond to arguments that are truth-preserving, necessarily or contingently—and that inference-like transitions in a normal human mind often exploit such relations.

Note that ‘inferential integration’ is defined in terms of certain interactions with the agent’s propositional *attitudes*: it says nothing about such interactions with states of other types. And it has two distinct aspects, that I’ll call ‘top-down’ and ‘bottom-up’ integration respectively (cf. Stich 1978, 507–510)—roughly: bottom-up integration is a measure of the range of

attitudes to which the state under consideration tends to *lead*, via inference-like proximate causal transitions. Top-down integration is a measure of the range of attitudes that tend to lead to *it*, via such transitions (or to *changes* in it, depending on how these states are individuated).

Inferential integration clearly comes in degrees, but there is supposed to be a salient threshold, below which a state no longer counts as integrated (but as ‘insular’ or ‘encapsulated’). Paradigms include certain representational states featured in classical computational theories of early vision (Marr 1982), face-recognition (Bruce & Young 1986), and grammatical competence (Chomsky 1976; Fodor 1983). The attitudes themselves are canonical of the other category: they’re over the threshold. Another feature that arguably distinguishes these paradigms from one another is that the former aren’t, but the latter are, consciously or reflectively accessible to the agent.<sup>37</sup>

But with one qualification (below), it’s extremely plausibly that those features come apart; specifically, that many of our higher cognitive capacities involve computations over states that are integrated, but not accessible, in the way (or ways) that the attitudes are. (E.g. most going accounts of quick-and-dirty concept-categorization, different as they may be in other ways, posit psychologically real bodies of information that are inaccessible but arguably best understood as integrated.<sup>38</sup> The literature on dual system- and process-models provide other illustrations.<sup>39</sup> Likewise for the psychology of speech-comprehension, and mind-reading more generally.<sup>40</sup>) The qualification concerns the threshold(s) implicit in the characterization of inferential integration: perhaps there’s a strict but credible reading on which nothing other than the attitudes qualify as integrated. The point I want to make, however, doesn’t turn on the exact location of the threshold; only its precise formulation does. Suppose we adopt the strict reading. Then the way to put the point is to say that, below the cutoff imparted by that reading, there’s something *close* to integration—a *notable* tendency to enter into *near-enough* transitions of the specified sort with *some* range of attitudes—that various mental states that aren’t reflectively accessible exhibit. As with the inferential/non-inferential justification distinction, there seems to be a spectrum, where the traditional view suggested a clean dichotomy. (I conjecture that this isn’t an accident.)

Now let me say why it matters.

## 5.2.

Certain mental states are rationally evaluable—paradigmatically, perhaps exclusively, the propositional attitudes. Only, perhaps all, such states that have a mind-to-world direction of fit are subject to norms of distinctively *epistemic* (rational) evaluation. That class includes beliefs, arguably some emotions—e.g. certain fears, regrets, angers—and perhaps conditional

beliefs (assuming that those aren't reducible to attitudes of other types). It's *because* they're rationally evaluable, that states like these must themselves be justified to be sources of justification. What *makes* them thus evaluable, in turn, is that they're both reflectively accessible—accessible to introspection and self-reflection—and directly revisable: not simply by will but in response to sufficient good evidence, and/or the appearance of such. (Borrowing a term from Hieronymi 2014: we have *discretion* over them.) Since they are reflectively accessible, they can be 'summoned' for scrutiny, in light of new or standing evidence, as the agent deliberates about what to believe—or (say) what to fear or regret, or whether to fear or regret anything in the circumstances. And since they are directly revisable, such deliberation isn't *pointless*: it can have prompt effects—e.g. lead to revision or rejection of the attitude—and it typically does so without further effort on the agent's part.<sup>41</sup>

Interference mechanisms—or simply or lack of time or interest—may prevent a state like this from ever being scrutinized, revised, etc. Incompetence, ignorance or distraction might lead to assessments or revisions that are poor. But there's nothing in the *nature* of beliefs (and other mental states of this type) to prevent their being subject to effectual deliberation. It's consequent on—perhaps even part of—their distinctive functional profile that, absent interference, they're both reflectively accessible and directly revisable. And that, plainly put, is why we can't 'get away' with basing (other) beliefs, or intentions to act, on states of this type unless they're themselves justified. (Note that interference-mechanisms at best exculpates—e.g. a belief that hasn't been properly revised, because of interference, still doesn't count as rational.)

States that lack one or both of these properties—reflective accessibility and direct revisability—are 'beyond justification and un-justification': incapable of being either justified or unjustified. But some such states can arguably still be *sources* of justification. Perceptual experiences are accessible but not revisable; despite that, they can provide justification—perhaps, at least in part, because of their distinctive phenomenology. (For now, just note that most parties to the debate about perceptual justification agree that experiences can justify, without being justified in turn.<sup>42</sup>)

More controversially: inaccessible states that are inferentially integrated—in particular *top-down* integrated—to a non-negligible degree can also play a justificatory role (even though they're themselves a-rational). That's because integration expands the reach of our rational agency—by providing it with 'backdoor' access to our beliefs and other attitudes—and, in doing so, increases the efficiency with which we execute certain projects whose pursuits are constitutive of that agency. Notably, the project whose aim is to 'get things right'; to a rough first approximation: to believe truths rather than falsities, on topics that matter (and/or should matter) to us.<sup>43</sup> This makes having inferentially integrated states—more generally: integrated minds—preferable, from our perspective as rational agents.

A rational agent could try to get things right in any number of ways, but in so far as there's a kind of trying that's built into her 'job description', *qua* rational agent, I suggest that it's something like this: it's trying (to get things right) by the specifically first-personal methods of assessment and direct revision that we normally all come equipped with—the methods by which (at least to date) we can only manage our own beliefs and other attitudes, not those of others. In other words: the basic procedures that exploit introspection and self-reflection, and that are involved in critical deliberation of precisely the kind that grounds the need for beliefs—and other rationally evaluable mental states—to be justified, if they're themselves to be sources of justification. (When I say the agent's 'deliberate efforts to get things right' I mean *those* efforts.)

To close, then: I'm suggesting that there's an epistemically significant difference between mental states that exhibit a robust insensitivity to changes in the agent's attitudes, and states that don't. The latter tend to be at least *indirectly* responsive to her agency, and her deliberate efforts to get things right. Eventual progress on that score may trickle down, and bring about revisions among her inaccessible but (top-down) integrated states—revisions that may in turn feed back into her overall attitude system in ways that advance the goal. Eventual *regress* may trickle down as well; that's the inevitable flipside. These feedback-loops serve to facilitate her effort—whichever way that effort's headed, as it were. Whether she in fact makes progress—how successful she is with respect to the goal—depends on many other factors. But an important part of being a rational agent is that one *tries* to do so, by the default first-personal methods mentioned above.

In contrast: her deliberate efforts don't have any effect, direct or indirect, on states that are thoroughly (top-down) insular. Of course there *are* ways to affect such states; viz. by various forms of brute force. But there's no smooth intra-mental path—no preprogrammed extension of her default methods—by which to do it. In this way, then, inferential insularity curbs the efficiency of her deliberate efforts to get things right. Integration increases that efficiency in inverse proportion. And *that*, in briefest outline, is why a merely collaterally available content can play an auxiliary role in an agent's (inferential) justification. Sufficiently integrated, inaccessible mental states—with the right direction-of-fit and suitable contents—can confer justification on belief, although they're not themselves justified.

There's more to say about all of this, but that's a task for another occasion.<sup>44</sup>

## Notes

1. Nothing that I argue here turns on whether the kind of justification the hard cases exemplify is in the end best classified as a kind of its own—a kind that's

neither inferential nor inferential—or as a distinctive sub-variety of inferential justification (more inclusively conceived than tradition has it). The sketch in §5 is compatible with the latter choice. But whether that choice is ultimately viable depends on what (else) we want the proffered categories to do. (See Malmgren ms(b) for more discussion.)

2. And I take it that the subject-matter of the theory of justification includes the practices of such agents: that at least one of its aims is to describe what *we* do (if and) when we do things right, epistemically speaking—what it takes for agents like us to be justified, justified in certain ways, etc.
3. See e.g. Audi (1986, 2011); Bonjour (1985); Boghossian (2003, 2014); Coady (1992); Fumerton (1985, 2006); Ginet (2005); Goldman (2008); Huemer (2002); Leite (2011); Markie (2005); Neta (2013); Pollock & Cruz (1999); Pryor (2005); Tucker (2012, 2014); Wedgwood (2012).
4. On one version of coherentism, resting and basing can indeed be symmetric (when things go well). But this can be accommodated by adding provisos to (c) and (d) permitting cases where the resting/basing relation is sufficiently indirect in one direction, or by denying that resting/basing is transitive.
5. Suppose Bonnie is inferentially justified in believing *p*, partly on the basis of *q*. On the standard conception (this entails that) *q* stands to *p* in an inferential evidential support-relation—alone or together with other propositions *r*... *r<sub>n</sub>*, that are available to Bonnie. Equivalently: that there's a good inference rule from (believing) *q* to (believing) *p*—a rule that licenses a transition straight from *q* to *p*, or a transition from *q* and *r*... *r<sub>n</sub>* to *p*, where *r*... *r<sub>n</sub>* are available to Bonnie. Or, again equivalently: that there's a good argument from *q* to *p*—either a direct argument, or an argument with auxiliary premises, *r*... *r<sub>n</sub>*, where *r*... *r<sub>n</sub>* are available to her. (These are just different ways of stating the structural requirement.)
6. E.g. the claim that S's justification to believe *p* embodies an argument—in the richer sense—of form F can be rephrased as the claim that it embodies an argument—in the thinner sense—that's correctly evaluated by F standards.
7. There are many complicating factors—e.g. the impact of defeaters and higher-order norms (see e.g. Bach 1984; Harman 1984, 1986; Schechter 2013).
8. Perhaps there's a parallel constraint on the relationship between agents and premises like that. Perhaps there are also good arguments with premises that only should/may be intended, desired, feared, or hoped. (This, too, is inconsequential here.)
9. When used below, the simplification should always be understood as shorthand for the more precise conjunctive claim stated here.
10. It's plausible that belief-on-belief asymmetric dependence is sustained by the explanatory relation involved in basing, and that the justification-on-justification dependence is (at least partly) grounded in an asymmetry in the pattern of undercutting defeat (cf. Pryor 2005, 183). So perhaps we can also 'test' for that set of marks. This—more discriminating—test returns the same result in the hard cases, while avoiding the noise of possible over-determination. (See Malmgren ms(b).)
11. I've chosen to work with specific examples, rather than already-familiar categories (of judgment or belief) because most, if not all, such categories—semantic,



epistemic, and psychological—seem to crosscut the category I would like to draw attention to. Perhaps this turns out to be wrong in the end, but I want to keep the possibility open. I also keep open—and am prepared to take on board—that some of the examples I’m working with don’t in fact (have relevant realizations that) instantiate that category.

12. It’s neutral on the distribution of the ‘test-properties’, in the terminology of Malmgren (2013(b)).
13. See e.g. Häggkvist (2017); Ichikawa & Jarvis (2009); Malmgren (2011); Williamson (2007).
14. I’m not relying on any particular substantive theory of introspection or self-reflection. I take it that we have a distinctive epistemic route to the content and type of our own attitudes (and perhaps some experiences and moods): a route that’s fallible, but that others lack, that often yields relatively unreflective judgments (as defined above), and that at least seems to have an interesting degree of independence from routes that exploit current behavioral evidence. That route is what I call ‘introspection’. (It’s arguably *peculiar* but not *privileged*, in the terminology of Byrne 2005.) It involves at least some second-order cognition—e.g. whatever is required to (rightly or wrongly) classify a state of oneself as a belief that *p*. ‘Self-reflection’, as I use it, involves more—e.g. whatever is required to classify something as a reason why or for which one believes *p*, or as a cause of that belief—and it presupposes the ability to introspect. It may also draw on memory outputs, but we can stay neutral on the details of that interaction.
15. This constraint is further discussed in Malmgren ms(b).
16. The explanation might straddle that distinction in two different ways (depending on how exactly the distinction is glossed). First, and least controversially, the explanation is ‘mixed’: it trades in mental states *some but not all* of which are canonical personal-level states—viz. attitudes (beliefs or judgments). Second, it trades in states that are *reflectively inaccessible* but significantly *inferentially integrated* (and/or in the corresponding mechanisms); see §5. On one influential characterization of the personal/sub-personal divide—derived from Stich (1978) (cf. Evans, G. 1982; Davies 1989)—the second claim implies that the explanation breaches the divide in more than one way. (For competing characterizations, see e.g. Dennett 1969; Drayson 2012, 2014; McDowell 1994; Kriegel 2012. For more discussion, see Malmgren ms(b).)
17. More on this—and other options—in Malmgren ms(b).
18. The other (examples of) hard cases are also naturally read in ways that make the test come out positive. E.g. if Mae hadn’t been justified in believing *that the prisoners were tortured*, she wouldn’t be justified in believing that the prisoners were wronged, but not *vice versa*; if Polonius weren’t justified in believing *that Ophelia is doing the dishes—again*, he wouldn’t be justified in judging she’s depressed, but not *vice versa*; etc.
19. My view is that both are (disguised) beliefs in metaphysical possibility claims (Malmgren 2011, §2.3).
20. The usual way to elaborate on this, in turn, is to say that an enabler makes it possible for an explainer to yield the given effect, or explain what it explains. (See e.g. Cheng & Novick 1991; Lombard 1990; Plato *Phaedo* 98e; Woodward 2011; Yablo 2003.) What ‘making possible’ comes to can vary, it seems: the enabler can



be an individually necessary condition, as in this example, but it needn't be—in other cases, it's an INUS condition (Mackie 1965). Perhaps it can also be an INUS condition for the explainer, but not the effect (see Lombard *ibid.*, §VI).

21. See e.g. Bermudez (1995); Malmgren (2013(a), §4.1); Woodward (2011).
22. E.g. cases where an agent's having a priori justification is enabled by the possession of certain concepts, or certain experiences, are importantly different in this regard.
23. More on this, too, in Malmgren ms(b).
24. But note that such clues alone—including introspectively accessible information—never settle whether a given argument is instantiated in a way that's relevantly complete.
25. More carefully: too strict unless the traditional (inferential/non-inferential) dichotomy is abandoned, and replaced with multiple categories.
26. See e.g. Malmgren (2011, §2.3).
27. There are phenomena in the vicinity—other commitments, associated with specific actions/beliefs—that *are* plausibly explained in other ways; e.g. commitments that derive from obligations to avoid simultaneously believing logically inconsistent contents, to endorse the obvious deductive consequences of one's current beliefs, and to act in ways that reflect one's considered views on how to act. (Perhaps there are also commitments that derive from our competence with certain concepts.) But these are, precisely, *other* commitments—with other projection patterns—and their explanations aren't plausible *competitors* to the explanation I've outlined: they're explanations of something else.
28. I'm not suggesting that the first-person exercise, and the testimonial route that piggybacks on it, dovetails perfectly with any extant experimental methods. But some such methods—e.g. in psycho-linguistics, the study of mind-reading, folk-biology and concept categorization—reveal behavioral patterns that are plausibly interpreted as manifesting dispositions to conform to r-commitments.
29. Including cases where the original response is mistaken. Then the exercise may end in retraction of that response. (Sometimes that's precisely how the mistake is discovered.)
30. All the overall argument requires is that there are some actual/nearby realizations where the target beliefs are rational, but incur r-commitments that cause trouble for the standard conception. That much seems extremely plausible. (It doesn't matter how exactly the trouble is caused, or if it's caused in the same way across all realizations of the same case.)
31. In some of the cases (e.g. BERTRAND, ZINAR) this may be the only fundamental route within reach. The agent—e.g. Bertrand—might of course take that route independently of any particular verdict. So any particular judgment of his, about the distribution of the test-properties in one or another Gettier-case, might in principle be (antecedently) justified that way. But then let our question be about the *other* judgments: in virtue of what are *they* justified? This is an additional stipulation—but it's harmless since if none of his judgments are justified some other way then none of them are justified at all.
32. Brandom (2000) seems to endorse something like this view.
33. It leaves open whether a content is encoded in a state and/or processing-mechanism (relative to a given agent, belief and explanation). In that way, then,

- the proposal isn't hostage to empirical fortune, and it honors the observation that a difference in the division of labor between rules and premises, in the specification of an argument form, doesn't *per se* make an epistemic difference.
34. Both of these claims are substantiated further in Malmgren ms(b).
  35. A mechanism can't be inferentially integrated, but it can be more or less *automatic*, and non-automaticity of a certain sort plays a role similar to integration in the rational agent's cognitive economy. See Malmgren ms(b). (There I also discuss collateral availability for propositional justification.)
  36. See e.g. Kornblith (2012); Orlandi (2014); Richard (forthcoming); Siegel (2017); Sperber & Wilson (1986). Stich (1978) thought that the relevant content-properties supervene on syntactic properties, and that only the latter are causally efficacious. If ampliative (and analytic) strength isn't reducible to logical validity, this restricts the range considerably. (Quilty-Dunn & Mandelbaum 2018 boldly bank on the reducibility. But see e.g. Goodman 1955; Titelbaum 2010.)
  37. The traditional conception of the personal/sub-personal distinction mentioned in fn. 16 seizes on this.
  38. It seems by far the best way to make sense of certain typical changes in our categorization dispositions. (See e.g. Ashby & Maddox 2005; Carey 1988, 2011.)
  39. See e.g. Fugelsang & Thompson (2003); Schroyens, Schaeken & Handley (2003); Stanovich & West (2000). But here the possibility of interference, and disagreements between 'parallel-process' and 'default-interventionist' models, complicate matters (cf. Evans, J. 2007; Lyons 2016).
  40. See e.g. Carston (2002); Gopnik & Meltzoff (1997); Recanati (2002); Sperber & Wilson (1986, 2002).
  41. The picture I'm giving is meant to be neutral between different accounts of what exactly this involves: what it takes to have discretion (see e.g. Frankfurt 1971, 1988; Hieronymy 2014; Nolfi 2015).
  42. Siegel (2017) is a notable exception.
  43. Other constitutive projects may include that whose aim it is to increase *understanding* on topics that matter and/or should matter to us.
  44. Many thanks to Patrick Greenough, Cory Juhl, Jack Lyons, Ram Neta, Nico Orlandi, and Mike Raven for helpful comments on an earlier draft.

## Bibliography

- Ashby, F. & Maddox, W. 2005. "Human category learning." *Annual Review of Psychology* 56, 149–178.
- Audi, R. 1986. "Belief, Reason and Inference." *Philosophical Topics* 14, 27–65.
- . 2011. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*, 3rd. edition. New York: Routledge.
- Bach, K. 1984. "Default Reasoning: Jumping to Conclusions and Knowing When to Think Twice." *Pacific Philosophical Quarterly* 65, 37–58.
- Bealer, G. 2002: "Modal Epistemology and the Rationalist Renaissance." In *Conceivability and Possibility*, eds. Gendler, T. & Hawthorne, J. Oxford: Oxford University Press.
- Bermudez, J. 1995. "Syntax, Semantics and Levels of Explanation", *Philosophical Quarterly* 45, 361–367.
- Boghossian, P. 2003. "Blind Reasoning." *Aristotelian Society Supplementary Volume* 77, 225–248
- . 2008. "Epistemic Rules." *The Journal of Philosophy* 105, 475–500.

- . 2014. “What is Inference?” *Philosophical Studies* 169, 1–18.
- Bonjour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.
- . 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Brandom, R. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge: Harvard University Press.
- Bruce, V. & Young, A. 1986. “Understanding Face Recognition.” *British Journal of Psychology* 77, 305–327.
- Byrne, A. 2005. “Introspection.” *Philosophical Topics* 33, 79–104.
- Carey, S. 1988. “Conceptual Differences Between Children and Adults.” *Mind & Language* 3, 167–181.
- . 2011. *The Origin of Concepts*. Oxford: Oxford University Press.
- Carston, R. 2002. “Linguistic Meaning, Communicated Meaning and Cognitive Pragmatics.” *Mind & Language* 17, 127–148.
- Cheng, P. & Novick, L. 1991. “Causes Versus Enabling Conditions.” *Cognition* 40, 83–120.
- Chomsky, N. 1976. “On the Nature of Language.” *Annals of the New York Academy*, 46–57.
- Coady, C. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.
- Cohen, S. 2002. “Basic Knowledge and the Problem of Easy Knowledge.” *Philosophy and Phenomenological Research* 65, 309–329.
- Dancy, J. 2004. *Ethics without Principles*, Oxford: Oxford University Press.
- . 2013. “Moral Particularism.” *The Stanford Encyclopedia of Philosophy*, ed. Zalta, E. <https://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>
- Davidson, D. 1986. “A Coherence Theory of Truth and Knowledge.” In *Truth and Interpretation*, eds. LePore, E. Oxford: Blackwell.
- Davies, M. 1989. “Tacit Knowledge and Sub-Doxastic States.” In *Reflections on Chomsky*, ed. George, A. Blackwell.
- Dennett, D. 1969. *Content and Consciousness*. London: Routledge.
- Dogramaci, S. 2013. “Intuitions for Inferences.” *Philosophical Studies* 165, 371–399.
- Drayson, Z. 2012. “The Uses and Abuses of the Personal/Subpersonal Distinction.” *Philosophical Perspectives* 26, 1–18.
- . 2014. “The Personal/Subpersonal Distinction.” *Philosophy Compass* 9, 338–346.
- Evans, G. 1982. *The Varieties of Reference*, Oxford University Press.
- Evans, J. 2007. “On the Resolution of Conflict in Dual-Process Theories of Reasoning.” *Thinking & Reasoning* 13, 321–329.
- Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Frankfurt, H. 1971. “Freedom of the Will and the Concept of a Person.” *The Journal of Philosophy* 68, 5–20.
- . 1988. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.
- Fugelsang, J. & Thompson, V. 2003. “A Dual-Process Model of Belief and Evidence Interactions in Causal Reasoning.” *Memory and Cognition* 31, 800–815.
- Fumerton, R. 1985. *Metaepistemology and Skepticism*. Lanham, MA: Rowman & Littlefield.
- . 2006. *Epistemology*. Wiley-Blackwell.
- Gettier, E. 1963. “Is Justified True Belief Knowledge?” *Analysis* 23, 121–123.
- Gopnik, A. & Meltzoff, A. 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
- Ginet, C. 2005. “Infinetism is Not the Answer to the Regress Problem.” In *Contemporary Debates in Epistemology*, eds. Steup, M. & Sosa, E. Blackwell.
- Goldman, A. 2008. “Immediate Justification and Process Reliabilism.” In *Epistemology: New Essays*, ed. Smith Q. Oxford: Oxford University Press.
- Goodman, N. 1955. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.
- Harman, G. 1965. “The Inference to the Best Explanation.” *Philosophical Review* 74, 88–95.
- . 1973. *Thought*. Princeton: Princeton University Press.
- . 1984. “Logic and Reasoning.” *Synthese* 60, 107–127.

- . 1986. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press/Bradford.
- Hieronymi, P. 2014. "Reflection and Responsibility." *Philosophy and Public Affairs* 42, 3–41.
- Huemer, M. 2002. "Fumerton's Principle of Inferential Justification." *Journal of Philosophical Research* 27, 329–340.
- Häggqvist, S. 2017. "Thought Experiments, Disagreements, and Formalization." *Topoi*, 1–10.
- Ichikawa, J. & Jarvis, B. 2009. "Thought Experiment Intuitions and Truth in Fiction." *Philosophical Studies* 142, 221–46.
- Kornblith, H. 2012. *On Reflection*. Oxford: Oxford University Press.
- Kriegel, U. 2012. "Personal-level Representation." *Protosociology* 28, 77–114.
- Leite, A. 2011. "Immediate Warrant, Epistemic Responsibility, and Moorean Dogmatism." In *Reasons for Belief*, eds. Reisner, A. & Steglich-Petersen, A. Cambridge: Cambridge University Press.
- Lombard, L. 1990. "Causes, Enablers and the Counterfactual Analysis." *Philosophical Studies* 59: 195–211
- Lycan, W. 1988. *Judgment and Justification*. New York, NY: Cambridge University Press.
- Lyons, J. 2016. "Unconscious Evidence." *Philosophical Issues* 26, 243–262.
- Mackie, J. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2, 245–264.
- Malmgren, A-S. 2011. "Rationalism and the Content of Intuitive Judgements." *Mind* 120, 263–327.
- . 2013(a). "A Priori Testimony Revisited." In *The A Priori in Philosophy*, eds. Casullo, A. & Thurow, J. Oxford: Oxford University Press.
- . 2013(b). "Review of *Philosophy Without Intuitions* by Herman Cappelen." *Notre Dame Philosophical Reviews*.
- . ms(a). "Availability, Goodness, and Argument Structure."
- . ms(b). *Understanding Inference*.
- Markie, P. 2005. "The Mystery of Direct Perceptual Justification." *Philosophical Studies* 126: 347–373.
- Marr, D. 1982. *Vision*. New York: Henry Holt.
- McDowell, J. 1994. *Mind and World*. Cambridge: Harvard University Press.
- Neta, R. 2013. "What is an Inference?" *Philosophical Issues* 23, 388–407.
- Nolfi, K. 2015. "Which Mental States Are Rationally Evaluable, And Why?" *Philosophical Issues* 25: 41–63.
- Orlandi, N. 2014. *The Innocent Eye*. New York: Oxford University Press.
- Pollock, J. & Cruz, J. 1999. *Contemporary Theories of Knowledge*, 2nd edition. Rowman & Littlefield.
- Pryor, J. 2005. "There is Immediate Justification." In *Contemporary Debates in Epistemology*, eds. Steup, M. & Sosa, E., Malden. Cambridge MA: Blackwell.
- . 2012. "When Warrant Transmits." In *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*, ed. Coliva, A. Oxford: Oxford University Press.
- Quilty-Dunn, J. & Mandelbaum, E. 2018. "Inferential Transitions." *Australasian Journal of Philosophy* 96, 532–547.
- Recanati, F. 2002. "Does Linguistic Communication Rest on Inference?" *Mind & Language* 17: 105–126.
- Richard, M. forthcoming. "Is Reasoning a Form of Agency?"
- Schechter, J. 2013. "Rational Self-Doubt and the Failure of Closure." *Philosophical Studies* 163: 429–452.
- . forthcoming. "Internalism and Entitlement to Rules and Methods."
- Schroyens, S., Schaeken, W. & Handley, S. 2003. "In Search of Counter-Examples: Deductive Rationality in Human Reasoning." *Quarterly Journal in Experimental Psychology* 56, 1129–1145.
- Siegel, S. 2017. *The Rationality of Perception*. Oxford: Oxford University Press.
- Silins, N. 2005. "Transmission Failure Failure." *Philosophical Studies* 126, 71–102.

- Sperber, D. & Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- . 2002. “Pragmatics, Modularity and Mind-Reading.” *Mind & Language* 17, 3–23.
- Stanovich K. & West, R. 2000. “Individual Differences in Reasoning: Implications for the Rationality Debate?” *Behavioral and Brain Sciences* 23, 645–726.
- Stich, S. 1978. “Beliefs and Sub-Doxastic States.” *Philosophy of Science* 5, 499–518.
- Titelbaum, M. 2010. “Not Enough There There: Evidence, Reasons, and Language Independence.” *Philosophical Perspectives* 24, 477–528.
- Tucker, C. 2012. “Movin’ on Up: Higher-Level Requirements and Inferential Justification.” *Philosophical Studies* 157, 323–340.
- . 2014. “On What Inferentially Justifies What: The Vices of Reliabilism and Proper Functionalism.” *Synthese* 191, 3311–3328.
- Wedgwood, R. 2012. “Justifying Inference.” *Synthese* 189, 1–23.
- Williamson, T. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.
- Woodward, J. 2011. “Causes, Conditions, and the Pragmatics of Causal Explanation.” In *Philosophy of Science Matters: The Philosophy of Peter Achinstein*, ed. G. Morgan. Oxford: Oxford University Press.
- Wright, C. 2002. “Anti-Skeptics Simple and Subtle: G.E. Moore and John McDowell.” *Philosophy and Phenomenological Research* 65, 330–348.
- . 2004. “Warrant for Nothing (and Foundations for Free)?” *Supplement to Aristotelian Society Proceedings* 78, 167–212.
- . 2007. “The Perils of Dogmatism.” In *Themes from G.E. Moore*, eds. Nuccetelli, S. & Seay, G. Oxford: Oxford University Press.
- . 2014. “Comment on Paul Boghossian, ‘What is Inference?’” *Philosophical Studies* 169, 1–11.
- Yablo, S. 2003. “Causal Relevance.” *Philosophical Issues* 13, 316–328.