

# On fundamental responsibility

Anna-Sara Malmgren

Stanford University

## Correspondence

Anna-Sara Malmgren, Stanford University.  
Email: malmgren@stanford.edu

## Abstract

Some psychological states—paradigmatically, beliefs and intentions—are rationally evaluable: they can be rational or irrational, justified or unjustified. Other states—e.g. sensations and gastrointestinal states—aren't: they're a-rational. On a familiar but hard-to-make-precise line of thought, at least part of what explains this difference is that we're somehow *responsible* for (having/being in) states of the former sort, in a way we're not for the others. But this responsibility can't be modeled on the responsibility we have for our (free, intentional) actions. So how should it be understood? In this paper I address that question. The overall shape of my answer is in line with tradition: I take the responsibility to be grounded in certain capacities for *reflection* and *control*. Answers in this family have recently been subjected to an interesting challenge. But the version I defend meets that challenge.

## 1 | INTRODUCTION

On a familiar picture, rational evaluability is grounded in a certain kind of responsibility: a responsibility we have for some, but not all, of the states we're in (and perhaps for some, but not all, of the things we do), that we don't have for the states of *other* agents, and that non-agents don't have for anything. Following Hieronymi (2014), let's call the relevant kind of responsibility 'fundamental', to distinguish it from the kind of responsibility that's partly explained by responsibility for something else—such as the responsibility we have for the foreseen consequences of our intentional actions (and arguably for those actions themselves).

Beliefs and intentions are paradigmatic of states for which we're fundamentally responsible. Certain complex reactive emotions also good candidates. (Perhaps decisions too, depending on their relation to intentions.<sup>1</sup>) Aches and urges, gastrointestinal and spatiotemporal states, the representational states manipulated in the early visual system, and the—mental and physical—states of other people, are

paradigmatic of states for which we aren't. I'm responsible, in this way, for believing that I'm over forty, for intending to stay out late, and (perhaps) for deciding to vote for a certain presidential candidate. But I'm not responsible in this way for *being* over forty, for having a headache the next morning, or for any of that candidate's ensuing political decisions or actions. At most I have derivative responsibility for some of their decisions or actions, and for my headache. I have none for my age.

What makes something an object of fundamental responsibility? I address that question below. But first, in §2, I briefly elaborate on the link between responsibility and rational evaluability—in particular as regards mental states. In §3, I sketch my proposal, and rehearse a recent challenge aimed at all proposals of the same broad type as mine. In §4–8, I clarify and defend the proposal, show that it meets the challenge, and argue that it explains fundamental responsibility.

## 2 | RATIONALITY AND RESPONSIBILITY

Our overall practice of rational evaluation includes *prima facie* appeals to deontic norms governing the management—formation, possession, and revision—of certain mental states, as well as to such norms governing actions. (Where, to a first approximation, deontic norms are norms stating permissions and/or obligations in a given domain of behavior, broadly conceived.) We commonly think or say things like: 'you should believe that you'll die within the next hundred years, but, absent further information, suspend judgment about exactly when', and 'the members of the Bush administration shouldn't have concluded that Iraq had WMDs, nor decided to launch an invasion, on the basis of such questionable evidence'. Taken at face value, claims like these tap genuine deontic norms—deontic norms that bear directly on what beliefs or intentions agents have, and on what grounds (e.g. Kornblith, 1983, 2001; McHugh, 2013).

The proper application of a deontic norm to an agent plausibly requires that the behavior the norm concerns—e.g. licenses, forbids, or dictates—is such that that agent can have responsibility for behavior of that type, in a sense of 'responsibility' yet to be made precise (e.g. Hieronymi, 2008, 2014; Nolfi, 2014). Indeed, as long as we distinguish between the *inapplicability* and *blameless violations* of a norm, we can say something stronger: it requires the behavior to be such that she *has* responsibility for behavior of that type. I'm blameless for being over forty, and for getting older each day—but not because I blamelessly violate, or for that matter heed, some deontic norms governing aging. I'm blameless because there are no such norms (at any rate: none that apply to me). And at least part of the reason there are no such norms is that I lack the requisite responsibility for aging. Likewise for any other category of 'behavior' that subsumes being over forty and getting older each day, which might figure in a candidate norm explaining my blamelessness.

What these appeals suggest, then, is that we're responsible for our beliefs and intentions in the way that makes intending and believing qualify as domains of behavior that can be governed by deontic norms. And they do so by virtue of suggesting that believing and intending *are* governed by such norms.<sup>2</sup> The fact that we take (what Strawson (1962) called) 'participant reactive attitudes' to each other and ourselves—in ways that accord with these appeals—lends further support to the suggestion (e.g. Boghossian, 2016; Hieronymi, 2006; Nolfi, 2015; Smithies, 2016).<sup>3</sup> E.g. we routinely praise or criticize each other, on account of our intentions and beliefs. We also recognize various forms of legitimate *excuse*, parallel to those we take to weaken or eliminate culpability for the violation of deontic norms governing actions.

The claim that a presumption of responsibility is implicit in our practice is of course consistent with the presumption being false. As it might be if, say, it entailed a solution to the free will problem that conflicts with our best scientific theories (Pereboom, 2001). Or if it turned out that we're only

responsible, in the requisite sense, for what's *voluntary*—in a way that mental states in particular can't be, or at least aren't actually (e.g. Alston, 1989; Williams, 1973). Having mentioned those options, however, I'll set them aside. A third—happier—prospect is that the presumption implies neither a radical conception of the will, nor an implausible form of voluntarism about the attitudes. Below I'll say something about what fundamental responsibility might be, within those constraints. I won't discuss many extant views in this family either.<sup>4</sup> (But I'll note points of contact and disagreement with some that are especially close to mine.)

### 3 | HIERONYMI'S CHALLENGE

I take fundamental responsibility to be partly grounded in what I call *direct deliberative leverage* (or 'direct leverage' for short).<sup>5</sup> We're only fundamentally responsible for states over which we have direct leverage. And, to anticipate, what it is for a state to be under our direct leverage is for it to be a potential unmediated outcome of deliberation—specifically: an outcome that's unmediated by any further attitudes or intentional actions—and to have that potential in virtue of the functional role that's characteristic of the type of state it is. (More on this in §4–6.)

What the notion of direct leverage is meant to capture is the elusive sense in which it's in or control or 'up to us', as agents, to believe and intend as we do—but not how old we are, how sharp a cut feels, or whether we're hungry. But it doesn't suffice for responsibility. Rather, we're fundamentally responsible for all and only *reflectively accessible* states over which we have direct leverage—where reflective accessibility is understood as accessibility to introspection and self-reflection, modestly conceived in turn. (See §7.)

Thus the picture I favor is a version of what Hieronymi (2014) calls 'the reflective view': one on which (fundamental) responsibility is grounded in a capacity for reflection or awareness, together with a capacity for discretion or control of a suitable kind. The version of the view that Hieronymi primarily engages—and uses as her foil—takes the capacity for control to *require*, or even partly consist in, reflective capacities (in turn typically glossed as certain capacities for attitude self-ascriptions, and for higher-order desires or endorsements).<sup>6</sup> That's not how I see the relationship. But it'll be useful to start by considering this version, and Hieronymi's criticism of it. Doing so helps pinpoint some important constraints.

Here's how she introduces the reflective view, and the intuitive motivation for it (2014, 4–5):

*We are, it seems, responsible for our intentional actions, if we're responsible for anything. Intentional action provides a kind of paradigm case of responsible activity. Intentional action also seems to involve, at least in its paradigm instances, a certain sort of 'having in mind.' In the paradigm cases, we act intentionally by first deciding what to do and then doing what we decided. [...] Our sense of control over our own actions thus involves both a certain kind of awareness—we have in mind what we intend to do—and a certain kind of voluntariness or discretion—we can decide to do whatever we think worth doing. It's very natural to think that this sort of control, the kind that, in its paradigm instances, involves both discretion and awareness, isn't only a ground for but also a condition on our responsibility for our intentional actions: that we're responsible because we enjoy such control, and that, if we lack it, we can't rightly be held responsible.*

To avoid confusion, I'll use 'Control' for the kind of sway described here, and whose paradigmatic manifestations—on this Hieronymi agrees with her opponent—involve "a certain kind of awareness

[...] and a certain kind of voluntariness or discretion". (Or, as we might say, 'control'—but control of a different, explanatorily more basic, kind, on pain of obvious narrow circularity in the characterization, or the launch of a regress.)

Hieronymi proceeds to argue that, natural as it may seem, this model can't be used to explain fundamental responsibility—specifically: that there's no plausible gloss on reflection/awareness and discretion/control, such that the responsibility we have for our beliefs and intentions is explained by our having Control over them. And the main objection, as I understand it, is this: however we analyze reflection, it's idle on its own. The (exercise of the) ability to identify and think about one's attitudes or actions—even to rationally, reliably or knowledgeably self-ascribe, evaluate or endorse them—can't, by itself, bring about any relevant psychological effects. So discretion is key. But we'd better not analyze discretion along the same lines that we just analyzed Control—or in terms of anything that requires or manifests Control—for the reason already outlined: doing so leads to a very tight circle, or an explanatory regress.<sup>7</sup>

Witness (*ibid.*, 8):

*[M]erely being able to reflect upon a thing doesn't provide one with control over that thing. If one is to control something of which one is aware, one must also be able to change that thing. [...] However, insofar as the reflective strategy secures our control over ourselves by appealing to the fact that we can reflect upon and change ourselves, it has, it seems, secured our control over ourselves by appeal to a self-directed action. But this won't do. If there was a question or problem about how or why we're responsible for our intentional actions, we can't answer it by appeal to a self-directed intentional action.*

Hieronymi considers two lines of response: first, that (on the right way of understanding the reflective view) fundamental responsibility is grounded, not in Control, but in a 'special, *sui generis*, reflective activity' that secures the requisite discretion or control. She then points out, fairly it seems, that "we're owed some account both of what that activity is and, crucially, why it, with whatever features it boasts, does the job of grounding or conditioning our responsibility" (*ibid.*, 9). The second response is to say that reflection is required for separate reasons—i.e. not because it secures control over our attitudes (on the problematic model of Control, or in some other way). Hieronymi asks, again fairly, *why* then reflection would be needed, and voices skepticism about the possibility of answering this question without falling back on the idea that responsibility (of any kind) requires Control, and Control requires reflection (*ibid.*, 25–6). But that's too quick. I think there's a rather straightforward rationale available—one that's independent in the way Hieronymi's challenge demands. We'll get to this in §7.

Now let me clarify how I think of discretion or control—or, in my preferred terminology, direct (deliberative) leverage.

## 4 | DELIBERATION AND LEVERAGE

I take deliberation to be a psychological process aimed at resolving a broadly practical question for the agent—paradigmatically, and loosely put, what if anything to do, or what if anything to believe, in the given circumstances (e.g. Shah & Velleman, 2005; Wallace, 2018). It involves the manipulation of at least some propositional attitudes—at a minimum: beliefs, suppositions or hypotheses about the facts or considerations which the agent takes to indicate that something is true or likely, or a good thing to do. And it requires an at least rudimentary ability to respond to those considerations in ways that constitute attempts to heed the fundamental norms that govern the deliberative pursuit at hand. I'll return to

what this comes to. (It can—but on my view needn't—involve “judgments about which considerations constitute rationalizing or justifying reasons for belief [or action]” (Nolfi, 2015, 49).) Deliberation is first-personal, it's typically prospective, and it proceeds on the assumption that an eventual answer to the deliberative question will translate—with phenomenological immediacy and effortlessness—into the realization of the resolved-on doxastic attitude, or of the intention to perform the resolved-on action.<sup>8</sup> And in normal cases, that's indeed what happens. Specifically: deliberation is normally efficacious, with respect to the implicated aspects of the agent's own psychology—in a way that requires no further intentional actions, or processing of attitudes, on her part.

Unless deliberation had this kind of traction, it would require the completion of a supertask. (Recall the worry about understanding fundamental responsibility on the model of Control.) But it only has that traction in a certain domain—a domain that, paradigmatically, includes (our own) beliefs and intentions. Simplifying a bit, my suggestion is that to have direct deliberative leverage over a mental state is just for the state to be in that domain.

On this picture, then, what direct leverage comes to is the ability to bring about a preferred psychological outcome—an outcome that sometimes amounts to a change—simply by engaging in deliberation. (Where the 'preferred' outcome is just the outcome that corresponds to one's answer to the deliberative question.) Some but not all mental states are normal unmediated results, or end-points, of deliberation—those that are, and only those, we have direct leverage over. This may sound flat-footed, but it's not ad hoc. The idea is that it's consequent on, or even part of, the distinctive functional role of certain attitudes that the kind of information-processing that constitutes deliberation is a potential partial proximate cause of states of that type.<sup>9</sup>

I'll soon say more about what I take such processing to involve. But the first—the present—point is that direct leverage is at heart an architectural matter: certain mental state types have a functional profile such that tokens of that type are normal unmediated outputs of deliberation. (Specifically: unmediated by further actions or attitudes.) These states are, constitutionally, hooked up in this way with the cognitive machinery deployed in deliberation. Other states aren't.<sup>10</sup> And, evidently, none of my mental states are hooked up that way with the cognitive machinery of *other* agents. This is the right result.

Many beliefs and intentions—even those of mature sophisticated agents with time on their hands—aren't generated through deliberation (as glossed above, and in further detail shortly). And it would be impossible for them all to be, on pain of another regress. But any one of our beliefs or intentions can be (or could have been) deliberately generated, at least in principle. Moreover, any one of them can be adjusted later—or strictly speaking terminated and replaced—through subsequent deliberation. Not so our metabolic and deep-seated affective states, visual and auditory experiences, states of other people, etc. ('At least in principle' is an unsatisfying place-holder, but some such hedge is needed to accommodate the various specific reasons, mundane and pathological, why deliberation may fail to occur.)

Perhaps there are possible creatures with direct leverage over a much wider, or just a very different, range of states. Perhaps there are also creatures who are incapable of subsequent effectual deliberation: once they form a belief or intention, in whatever way, it's fixed—in the sense that no amount of future deliberation can result in its removal or alteration. The only way to change it is by force (e.g. through surgery or drug intake). But that's irrelevant. We're not constituted like that, and states with the stubborn character just outlined (absent interference) would arguably not be beliefs or intentions at all. Either way: the possibility of such creatures is irrelevant, insofar as the deontic norms that govern our behavior—specifically, attitude management—are sensitive to the default general abilities and limitations of actual human beings. And it seems clear that they are. For one thing, beliefs and intentions are evaluated as rational/irrational relative to agents and contexts—and the context of evaluation sometimes differs, in ways that affect the verdict, from that in which an attitude was originally formed (or in

which it's been altered). This only makes sense, it seems, because we're able to reopen the deliberative question again and again—indeed, in principle indefinitely—and resettle it, when appropriate. If we weren't, perhaps the initial 'entry-rules' for the attitudes concerned would be way stricter than they actually are; likewise for the norms governing our strategies and habits as inquirers, and as makers of action-plans. (Conscientious agents might manage their habits of attitude-formation as we would our practical habits if all actions had the irrevocability of murder.) If the scope of our leverage was different, that too would likely be reflected in our norms, and reactive responses. (If it included headaches, but not beliefs about whales, there might be rational obligations directly bearing on our having headaches—but there aren't—and the norms governing doxastic attitudes would exempt whale-beliefs—and they don't.)

Direct leverage was characterized with reference to deliberation, so it's important to say more about that—in particular, to clarify that deliberation needn't recruit reflective capacities of the sorts that are required, to exercise discretion/control, on the 'classic' version of the reflective view that Hieronymi targets (§2).

## 5 | DELIBERATION AND NORMS

To repeat, I take deliberation to be a process aimed at resolving a question (canonically: what to believe/whether to believe that  $p$ , or what to do/whether to  $\varphi$ ), to engage at least some attitudes (e.g. beliefs about the considerations one takes to indicate that  $p$  is true, or that it would be good to  $\varphi$ ) and to normally be immediately efficacious. In particular, the preferred outcome (e.g. belief that  $p$ /intention to  $\varphi$ ) is normally realized without the mediation of further attitudes or actions. Normally it's also realized without mediation by moods or broadly sensory states (Davidson, 1963), or by the operations of external mechanisms or agents (Bishop, 1989). For present purposes it doesn't really matter what other processing goes on—e.g. at lower levels of explanation.

I also said that deliberation requires an at least rudimentary ability to respond to the (re)presented considerations—the considerations one takes to indicate that, say, something's true or good to do—and to respond in ways that constitute attempts to heed the fundamental norms that govern the deliberative pursuit at hand. These points in particular need clarification.

When trying to settle what doxastic attitude to take, the deliberative question is some variant of 'what to believe?'. When deliberating over actions, it's some variant of 'what to do?'. As is familiar, the kinds of considerations that have rational bearing on answering these questions differ in ostensibly deep ways. Notably only considerations pertaining to the truth (or falsity) of a proposition can count in favor of (against) candidate answers to the question what to believe. In particular, considerations to do with the desirability, or expected practical consequences, of answering it such-and-so can't make that answer rational (e.g. Hieronymi, 2008; Shah, 2003). But, of course, such considerations standardly support certain answers to the question what to do.

By 'the fundamental norms governing the deliberative pursuit at hand' I mean the most general norms or principles, whatever exactly they are, that explain this difference—or, better, the most general, deontic and/or non-deontic, norms that explain the more specific norms governing each pursuit: doxastic and (what I'll call) 'action-oriented' deliberation, respectively.<sup>11</sup>

The fundamental norms governing doxastic deliberation arguably include a norm to the effect that one should believe, and/or only believe, what's true (e.g. Boghossian, 2003; Wedgwood, 2002; Williams, 1973). In the case of action-oriented deliberation, they arguably include a norm to the effect that one should do, and/or only do, what's good or desirable (e.g. Davidson, 1978; Tenenbaum, 2007, 2018), and a norm to the effect that one should only intend to do what's possible—or even: what's in

one's unassisted power—to do (e.g. Baier, 1970; Velleman, 1997). Perhaps both kinds of deliberation are also governed by fundamental norms of structural coherence (e.g. Bratman, 1987; Broome, 1999, 2013). The exact formulation of all these norms is controversial, but we can bracket this.

In the first instance the above claims concern *rational* deliberation. But it's very plausible that the respective fundamental norms set limits to what counts as undertaking deliberation at all, and deliberation of one sort or another. Deliberation of any kind requires a minimum of *rationalizing* relationships between (attitude) contents, and the fundamental norms constrain what can rationalize what in each domain.<sup>12</sup>

To illustrate: take an agent—call her 'Ruby'—who's inclined to form an attitude with the content that *p* because she thinks it would be good if *p* were true (and/or she desires that *p*), in the face of considerations she confidently takes to indicate *p*'s falsity, and without countervailing suspicions or beliefs. Ruby doesn't count as deliberating about whether to believe that *p* (cf. Shah & Velleman, 2005, 502–6). At least she doesn't count as doing that in virtue of anything cited in *this* little story. (And we'd have to elaborate on it in quite fanciful ways to substantiate that she does.) Perhaps Ruby is deliberating about something else—say, about whether to hope that *p*. (Or she's not deliberating at all. Perhaps she's engaged in wishful thinking.) If, suppose, Ruby believes that her  $\varphi$ -ing would make it the case that *p*, she might in addition be deliberating about whether to  $\varphi$ . But if she also thinks that she's unable to  $\varphi$  without assistance, she's arguably not doing that either. At most she's deliberating about whether to *try*, or aim or aspire, to  $\varphi$  (Velleman, 1997, 31–3).

There's plenty of middle ground here: agents can deliberate very badly. But there's a limit, at which one isn't even deliberating badly anymore—e.g. about what to believe or do, or whether to fear rather than hope that *p*. And, interestingly, that limit seems to be marked by the fundamental norms that also explain what counts as deliberating well in each domain (by way of grounding more specific norms setting out those conditions). This point fits nicely with the suggestion made earlier: that it's part of the functional roles distinctive of intentions and beliefs that they're normal unmediated outcomes of deliberation. (We can now refine that suggestion by indexing the role of belief to doxastic, and of intention to action-oriented, deliberation.<sup>13</sup>) And I take the point to be independently plausible.

Ruby, in the example above, doesn't respond to considerations in ways that constitute attempts to heed the fundamental norms that govern doxastic deliberation. (Perhaps she responds in ways that constitute attempts to heed those governing deliberation of another sort—perhaps not.) Someone who spends years reading up on a legal case, methodically assessing all the relevant evidence, and eventually coming to believe that which she concludes is best supported by that evidence, in part because she concludes it's thus supported, presumably satisfies this requirement (other things equal). But a lot of deliberation, both doxastic and action-oriented, is much less sophisticated. What does the limit case look like?

I suggest that, at the limit, deliberation is just the formation or revision of propositional attitudes—or the pursuit thereof (in cases without resolution)—on the basis of other attitudes that rationalize them. That is: attitude-management for *reasons*, supplied at the attitude level—e.g. as the representational content of (other) beliefs and/or desires of the agent.<sup>14</sup> Attitudes are arguably not the only mental states that can provide reasons—but attitudes surely can, and do. Specifically: beliefs with suitable contents provide reasons for other beliefs (and perhaps intentions/actions), and suitable belief-desire combinations provide reasons for intentions and actions. And, as I think of it, deliberation—and the corresponding species of folk-psychological explanation—always trades in structural relations between reasons of this sort: reasons provided by attitudes.

It's important that 'on the basis of', in this gloss on limit-case deliberation, doesn't designate a relation which involves further attitudes—as basing ('proper') does on some analyses of the basing relation. It designates a non-deviant proximate causal relation. It's notoriously difficult to explicate

the relevant notion of (non-)deviance, and I won't make a serious attempt at doing that here. But the restriction is needed to rule out both types of interference mentioned early in this section. And we can get somewhere towards an explication—of so-called 'basic' causal deviance—by noting that the kind of causation and explanation that's involved in limit-case deliberation is *inference-like* or *content-sensitive* (e.g. Bishop, 1989; Peacocke, 1979; Schlosser, 2007).<sup>15</sup> Content-sensitive causal explanations support counterfactuals that track certain similarities and differences in the contents of the representational states that constitute the relata. And (in the case of interest here) *which* similarities and differences make a difference—and what difference they make—depends in the first instance on what type of deliberation the agent is pursuing; e.g. whether it's action-oriented or doxastic. (More precisely, on my view: which pursuit is at issue depends on what attitude-inputs dispose the agent this way or that, and the counterfactuals reflect those dispositions.)

It's unclear whether the notion of content-sensitivity, properly explicated, suffices to isolate the kind of transition that's involved in limit-case deliberation. But that it's *required* for deliberation, even at the limit, helps illuminate what it takes to respond to considerations—now precisified as attitude-provided reasons—in ways that constitute attempts to heed one or another set of fundamental norms. To thus respond is for the transition to (or towards) the target attitude to be content-sensitive in a way that's sufficiently in line with the fundamental norms proprietary to a broad but determinate type of deliberation. The overall modal pattern must display enough conformance to, say, the truth-norm—as filtered through the specifics of various counterfactual scenarios that differ with respect to the contents of the input-attitudes (here, beliefs) of the actual, would-be deliberative, process. In none of these scenarios, I take it, does the process have to make, or even incline, the agent to believe what's in fact true (in those scenarios). But it must incline her to believe what's sanctioned, or what she takes to be sanctioned, by the derivative norms the truth-norm explains—e.g. believe what's made likely, and disbelieve what's made unlikely, by the contents of the input-beliefs—in sufficiently many of them. How many are sufficiently many is another hard question.<sup>16</sup> But the bar is presumably quite low. Recall that all I'm doing here is stating minimum requirements for a psychological process to count as deliberation.

I now want to explore the implications.

## 6 | LEVERAGE WITHOUT REFLECTION

Consider a small child at the top of the stairs, facing the choice between staying and descending. Suppose that, prior to settling which action to perform, she visually or otherwise imagines a situation where she descends and falls, that this makes her believe that if she descends she will (or might or is likely to) fall, and that she really doesn't want to fall. Suppose also that this makes her disposed to intend to stay, and other things equal form that intention. If what I said in §5 is along the right lines, this child may well have engaged in (action-oriented) deliberation: deliberation of a primitive sort, but deliberation nonetheless. All we have to add, or read into the case, is that the transition to (towards) intention is relevantly unmediated, and that it's content-sensitive in a way that's sufficiently in line with the fundamental norms governing action-oriented deliberation. We *don't* have to add that the child imagines more than one hypothetical outcome, that she weighs any conflicting or competing considerations in thought, or that she thinks about any such considerations as counting for or against—as reasons for/against—taking either course of action. (She may or may not have the conceptual repertoire needed to do so.)<sup>17</sup> What makes the process one of deliberation, albeit primitive, is that she manifests at least a rudimentary ability to respond to the considerations she *does* entertain, in ways that constitute attempts to conform to the fundamental norms governing action-oriented deliberation.<sup>18</sup>

Similarly for doxastic deliberation. Suppose the child comes to believe that the front door slammed shut—perhaps because she heard a certain noise—and that this inclines her to judge (and other things equal judge) that mom’s home. This, too, could be deliberation: we just need to add that the transition to (towards) the resultant judgment is relevantly unmediated, and that it displays the right sort of content-sensitivity. We *don’t* need to add that she ponders any contrasting scenarios in thought—e.g. alternate explanations of the noise, or of the door’s slamming shut—that she pauses to screen for defeating factors, or that she thinks of the door’s slamming shut as a (or her) reason to believe that her mother is home. Note also that, in neither version of the example, need we suppose that the process involves any beliefs of hers about what mental states she’s in or why, or about what she’s doing or why. It seems perfectly possible to engage in effectual deliberation, of either type, without recruiting any such beliefs.

In my view, then, limit-case deliberation doesn’t require (what Peacocke calls) ‘second-tier’ thought: “thought [e.g. belief] about relations of support, evidence, or consequence between contents” (1996, 129–30). Ditto for mental state self-ascriptions, and for beliefs or other attitudes concerning the process itself or its aim. E.g. the agent needn’t think of herself as settling, or trying to settle, what to believe or how to act (*pace* Shah & Velleman, 2005). She just has to try to do it. Sometimes that’s easy enough (as it might be for the child above, or an adult in equivalent circumstances). Perhaps she needn’t even have any metacognitive concepts, and she certainly needn’t apply such concepts to herself—reliably, rationally, or at all—to undertake deliberation at the limit.

This is important, since I glossed direct leverage—which I take to partly ground fundamental responsibility—in terms of deliberation. I’ve now clarified that deliberation, as I understand it, needn’t involve attitude self-ascriptions or second-tier thought. Thus direct leverage doesn’t require the kind of reflective capacities that it (or the corresponding power: discretion/control) requires on the classic version of the reflective view.

A lot of deliberation *does* engage such capacities. Agents who have the capacity for second-tier thought often exercise it, and there’s a distinct and natural use for such thought in deliberation. (E.g. in the weighing of conflicting evidence, and in comparative assessments of alternate hypotheses/outcomes.) Likewise for self-ascriptions, and for partial rationalizing explanations of our actions and attitudes. Mature human deliberation—in particular *subsequent* deliberation—routinely involves self-ascriptions of mental states, and of operative reasons: reasons for which attitudes are held, and actions performed. If some want to reserve ‘deliberation’ for processes more complex than that of the child described above—e.g. for belief and intention management for attitude-provided reasons that recruits these capacities too—be my guest. I’m not wed to the term. But to minimize confusion I’ll use ‘*reflective* deliberation’ for deliberation that does.<sup>19</sup>

Importantly, the more sophisticated type of attitude-management requires—and simply builds upon—limit-case deliberation. The ability to respond to attitude-provided reasons in ways that constitute attempts to heed a suitable set of fundamental norms is present in rudimentary form already in the child, and manifest in her process. That ability is just exploited further in more complex attempts at resolving what to do or believe. If, say, the child does have the concept of evidence, and judges that the door’s slamming shut is (good) evidence that mom’s home, this may strengthen her disposition—incline her even more—to believe that her mom is home. But the move from that judgment, together with the belief that the door slammed shut, is just another relevantly unmediated content-sensitive transition of the appropriate (roughly truth-directed) sort. And she could in principle settle the deliberative question without making or drawing on that judgment: the unmediated causal efficacy—her direct leverage—doesn’t *depend* on it. That’s the crucial point.

To recapitulate: beliefs and intentions are rationally evaluable because we’re fundamentally responsible for (having/being in) states of that kind. And we’re fundamentally responsible, I’ve argued, in part because we have direct leverage over such states: we’re able to manage—form, revise, or terminate

—our beliefs and intentions simply through a process of deliberation (at varying degrees of sophistication). More specifically, those states are normal unmediated outcomes of deliberation—doxastic and action-oriented, respectively—in virtue of the functional roles that constitute states of each kind. That’s why we’re responsible for our *particular* beliefs and intentions, no matter their specific etiology (or means of sustenance). And it’s independently plausible that this mode of creation and change is part of the distinctive profile of these states.<sup>20</sup>

## 7 | WHY BE REFLECTIVE?

Direct leverage isn’t enough. Small children can engage in effectual deliberation, at least of the primitive sort; likewise for higher animals (insofar as they have attitudes). But small children and most animals lack the kind of responsibility that grounds rational evaluability. The notion of direct leverage is meant to capture the sense in which the management of certain attitudes is ‘up to us’, although—unlike the execution of free intentional actions—it isn’t voluntary. But there are further prerequisites for the proper application of deontic norms (than that the behavior they govern be up to us).

One might suggest that what makes the difference is the capacity for second-tier thought. But it’s not entirely clear why, by itself, it would.<sup>21</sup> The ability to (say) make judgments to the effect that such-and-such is a reason to  $\varphi$ , that this evidence outweighs that, or that a given claim implies another, might make one *better* at deliberation—in that it makes one better at conforming to the governing norms. (At least if the ability is somewhat reliable.) But how is that, in turn, supposed to explain fundamental responsibility?

What makes the difference, I suggest, are our *introspective* and *self-reflective* capacities—understood as at least moderately reliable capacities for attitude self-ascriptions, and for rationalizing explanations of our actions and attitudes. (Second-tier thought is needed to frame such explanations.) I’ll clarify how I think of these capacities; then I’ll motivate the suggestion.

In brief: I take introspection to be an epistemic and psychological route to self-ascriptions of (present and recently past) mental states.<sup>22</sup> The route is fallible but ‘peculiar’ (Byrne, 2005)—only, albeit perhaps contingently, available to the first person. And it affords kind of justification (and knowledge) that at least appears to have an interesting degree of independence from evidence pertaining to the agent’s current overt behavior. We can leave open to what extent it really is independent.<sup>23</sup> We can also leave open how introspection fares against other, non-peculiar, routes to mental state self-ascriptions. I’m not assuming that it’s more reliable—more likely to yield true beliefs (or for that matter justification or knowledge)—than all other extant routes, certainly not across the board. I do however assume that it’s reliable *enough*, in a certain domain, to vindicate its standard role in reflective deliberation, and in ‘egocentric’ folk-psychological explanation and prediction. That domain includes intentions, beliefs, certain emotions and experiences. It excludes, e.g., aches and urges, the representational states processed in early vision, complex character traits, and (trivially) all states of other people.

I take self-reflection, next, to be a route to rationalizing explanations of our own mental states—notably attitudes—and actions. It affords beliefs about why (or on what grounds/for what reasons) one has a certain attitude or does/did a certain thing: beliefs that are sometimes justified (or constitute knowledge). To introspect, one must have at least some metacognitive concepts and skills—e.g. the ability to classify something as a belief that  $p$ , and as one’s own. To self-reflect one must, in addition, be able able to engage in some second-tier thought—e.g. classify something as a reason to believe that  $p$ —and to construct little explanatory hypotheses in terms of reasons (or kin: grounds/motives, etc.). Beyond this, we can leave open what self-reflection requires. We can also leave open to what extent it’s unified along other dimensions: e.g. whether there are limits to the kinds of evidence or methods

it recruits (and if so, what those limits are). Certain default methods seem to be peculiar in their own right—e.g. the self-scrutiny at play when I ask myself why I believe that  $p$ , and an apparently viable partial answer comes to mind with phenomenological immediacy. Others don't—e.g. interpretation aided by a therapist's testimony. How reliable self-reflection is presumably varies quite considerably, depending on which sub-capacities and methods are deployed, what other capabilities (and limitations) the agent has, what state/action is at issue, etc. But I take it that, like introspection, self-reflection tends to be reliable enough, in a certain domain—a domain that includes beliefs and intentions—to vindicate its standard role in reflective deliberation, and egocentric folk-psychological explanation.

The argument below is compatible with a variety of precisifications of these characterizations (as well as with bolder views) of self-reflection and introspection. The key claim is just that we—normal human agents of a certain maturity—have capacities for self-ascriptions and rationalizing explanations of mental states that are at least passably reliable in a given domain. As we'll see, this helps explain why we're fundamentally responsible for some states in that domain—viz., those we have direct leverage over.

In general, *non-culpable ignorance of relevant facts*—or better, non-culpable false beliefs (or lack of true beliefs) about such facts—can exculpate: excuse agents from blameworthiness, when violating deontic norms, and undercut their praiseworthiness when complying (e.g. Anscombe, 1963; Smith, 1983).<sup>24</sup> *Systematic* (non-culpable) ignorance can systematically exculpate—indeed, *sufficiently* systematic ignorance can even render norms inapplicable in the first place. (On one understanding of that term, it can 'exempt' the agent: undercut any demands a norm would otherwise have placed on her, or permissions it would have granted.<sup>25</sup>) And, in the domain under consideration, the relevant facts—ignorance of which can excuse or exempt—includes facts about which attitudes one has, and why. I'll explain these points in turn.

Suppose that I step on a ladybug on the way across the lawn, killing it, but that I didn't see it (and had no salient reason to think it was there, and/or had overriding reason to look elsewhere). My ignorance here could presumably exculpate, even though (let's grant) it was wrong to kill it. Next suppose that I'm *blind*, and that that's why I didn't see it.<sup>26</sup> I then have a standing excuse from blameworthiness for killing ladybugs by stepping on them (although perhaps not for walking places I take to be ladybug-heavy). And I have an excuse on this occasion that derives from my standing excuse. Third—a bit more fantastically—suppose that ladybugs were imperceptible to humans, left no observable trace, and that entomologists hadn't yet inferred their existence, but that they were as prevalent and vulnerable as they actually are. Here, it would arguably not even be wrong—for me, or humans more generally—to kill ladybugs (as opposed to bad, for them and perhaps simpliciter). That is, we wouldn't just be blameless for doing so: our actions wouldn't be governed by norms prohibiting killing ladybugs in the first place (in this, or perhaps any other, way). That's because, for no fault of ours, we'd be sufficiently systematically unable to tell when we did or didn't undertake behavior of the target type.

It doesn't help to add that, say, there's a causal mechanism that makes our feet twitch whenever they descend towards a ladybug, prompting enough hesitation for it to get away—thus ensuring regular conformance to the norm.<sup>27</sup> For a deontic norm to govern our behavior, in a given domain, it seems we must be able to *register at the attitude-level* when we do/don't exhibit that behavior (under a description that's germane to the directives of the norm; more on this shortly). And unless there's at least a modicum of *reliability* to the ability, we may still have a standing excuse (as we might, in this case, if we were intermittently, but very rarely, able to perceive ladybugs or correctly infer their existence and whereabouts). The question *why* this is so will have to be postponed for now. But that it's so is suggested by examples like that of the blind ladybug-stepper, and of the (more severely limited) agents without any means of tracking ladybugs. What the agents in these cases are missing isn't, say, the appropriate motivational dispositions, or voluntary control over their movements: we can assume they

all have that. Nor is it the odd ladybug-sighting, or a reliable twitch-causing mechanism: that wouldn't help. What's missing is a reliable way to judge—or otherwise detect at the attitude-level—whether and when they're engaging in the would-be prohibited behavior.

Stepping on, and killing, ladybugs are fully-fledged intentional actions, but that's inessential: it seems clear that, with respect to this issue, what goes for (responsibility for) actions goes for states. And, I suggest, it goes for fundamental as well as non-fundamental responsibility. If an agent is non-culpably ignorant of being in a certain state—e.g. standing on a ladybug, carrying explosives, or fearing success—she's excused. If she's systematically non-culpably ignorant, of whether and when she's in states of that type, she's systematically excused: she either has a standing excuse, or (given sufficiently robust ignorance) an exemption.

However, this is decidedly *not* a normal human agent's epistemic situation vis-à-vis her own intentions and beliefs. We can be blamelessly ignorant on occasion (e.g. of what we intend or believe). Perhaps some of us are even systematically blamelessly ignorant, due to persistent—e.g. neuropathological—interference. (Depending on the details, the right thing to say about such agents may well be that they're not responsible for the affected attitudes.) But *without* interference, we're not blamelessly systematically ignorant on this score—certainly not systematically *enough*, to preclude the management of beliefs and intentions from being subject to deontic norms altogether. We're normally able to tell, with at least a modicum of reliability, whether and when we believe or intend such-and-such. At least once we've acquired the basic metacognitive tools needed to actuate and utilize our default route to attitude self-ascriptions (viz. introspection).

Now consider a final version of the case. Suppose we could indeed perceive ladybugs, and reliably tell when our feet hit and squashed them, but that we (non-culpably, systematically) failed to understand that this caused or constituted killings—e.g. because of blameless false beliefs pertaining to insect-survival. Arguably such ignorance too could excuse or exempt us from culpability for killing them. That's because we'd (still) be unable to track our behavior under a mode of presentation that's germane to the directives of the salient norm: one that captures its normatively relevant—here, 'bad-making'—aspect or features.<sup>28</sup>

For another example: suppose I live in a house whose previous inhabitants, unbeknownst to me, were brutally displaced to make room for me. If my ignorance is blameless, my living there is too, for I fail to recognize what I'm doing under a description that's germane to the demands of the norms that (let's grant) make it wrong. Similarly, if, unbeknownst to me, the regime has planted landmines all around the property to prevent my being ousted. Here, the normatively relevant features, of my living in the house, are causal-explanatory (triggering and sustaining respectively). But we can add that, me unawares, my living there is very likely to prolong the war. That's a relevant (foreseeable) effect. What features—and features of what broad kinds—are normatively relevant depends on what norms are salient.

It should be clear where this is going. A significant class of deontic norms governing attitude-management trade in (actual and potential) *operative reasons*—the reasons-for-which we believe such-and-such/in-light-of-which we (intend to) act such-and-so. (Or equivalently for present purposes: in grounds or motives.) Introspection, as I understand it, is a sufficiently reliable route to intentions and beliefs—individuated in the canonical way: by attitude-type and content. An agent could conceivably have access to that but be completely unable to identify the reasons for which she holds (or forms/discards/revises) her beliefs or intentions. Assuming her predicament is blameless, as well as systematic enough—e.g. because she lacks the requisite concepts—such an agent would, it seems, be exempt from culpability for violating any norms in this class. (Similarly with respect to praiseworthiness for complying.) In short: her attitude-management wouldn't be governed by these norms, because she'd be unable to reliably identify her attitudes under descriptions that are germane to their directives.

However, normal human agents aren't in this situation (either). Absent persistent interference, we're not sufficiently systematically non-culpably ignorant of why—for what reasons/on what grounds—we intend and believe as we do. At least we're not, once we've acquired the additional tools needed to self-reflect.

## 8 | CONCLUSION

I've proposed a version of the reflective view that answers Hieronymi's challenge. Fundamental responsibility is grounded in direct deliberative leverage, as glossed above, and in certain reflective capacities: introspection and self-reflection. But the role of these capacities isn't to secure direct leverage. They just explain why we're not sufficiently systematically blamelessly ignorant of certain facts (about what attitudes we have, and why) to prevent certain central norms (bearing on what attitudes we have, and why) from getting a foothold.<sup>29</sup>

## ENDNOTES

<sup>1</sup> I'll write as if decisions are acts of intention-formation, and judgments acts of belief-formation. (But nothing turns on this.) And I take intentions to be mental states in their own right, explanatory of actions.

<sup>2</sup> This point is neutral on whether heeding the deontic norms governing belief suffices for epistemic justification. It's also largely neutral on how these norms relate to norms of inquiry/'evidential conduct' (Audi, 2008). (Only 'largely' since it does presuppose that they don't reduce to norms of this type.)

<sup>3</sup> To coherently take a reactive attitude to someone, for doing something, one must hold them responsible for doing that thing. This isn't to say that to be responsible *just is* to be a target, or an appropriate target, of reactive attitudes—although it's compatible with that sort of view. It's what one would expect if reactive attitudes are only apt with respect to behaviors that are governed by deontic norms.

<sup>4</sup> For one thing I'll completely bypass 'character-based' accounts (e.g. Hobart, 1934).

<sup>5</sup> In Malmgren (2018, §5), I used 'direct revisability'. But it's not ideal to use 'revision' for both the formation and subsequent alteration of attitudes. (In particular if, as seems plausible, alteration is best understood as the termination and replacement of one state with another.) 'Leverage' isn't ideal either, but it'll have to do. The more commonly used 'control' carries connotations of voluntariness which don't seem to get a grip, at the right point in the explanation, when it comes to attitude-management. But what I call 'direct leverage' is close to what Hieronymi (2006) calls 'evaluative control', McHugh (2017) calls 'attitudinal control', Moran (2001) calls 'deliberative control', and Nolfi (2014) calls 'doxastic control'.

<sup>6</sup> See, e.g., Frankfurt (1971, 1982), Korsgaard (1996), McDowell (1994), and Scanlon (1998).

<sup>7</sup> See also Kornblith (2012, esp. Ch. 3).

<sup>8</sup> The disparity here is superficial. We typically settle what to intend by settling what to do. But likewise, we typically settle what to believe by settling what's true. Indeed that's arguably the *only* way to settle that question (at least rationally)—i.e. in the deliberative context, it's necessarily 'transparent' to the question what's true. It's unclear whether the corresponding holds for the question what to intend (e.g. Hieronymi, 2006, §5). But that that question *can* be, and typically is, transparent to the question what to do shouldn't be controversial.

<sup>9</sup> See Nolfi (2014, 2015) for a suggestion in a similar spirit (on stronger assumptions about deliberation).

<sup>10</sup> E.g. contrast the corresponding—highly implausible—claim about pitch-perception, gastroenteritis, sexual desire, or vertigo.

<sup>11</sup> To be distinguished from 'detached' deliberation about what to intend (e.g. Kavka, 1983; cf. fn. 8). For simplicity I bypass complications peculiar to such deliberation.

<sup>12</sup> If I understand him correctly, Wright makes a very similar point in his (2012, §3.1). (What Wright calls 'basic inference' is close to what I call 'limit-case deliberation'.)

- <sup>13</sup> Cf. Shah and Velleman (2005, 502). (On my picture, the order of explanation is reversed.)
- <sup>14</sup> Where reasons are understood non-factively—e.g. as propositions.
- <sup>15</sup> See also Fischer and Ravizza (1998) and McHugh (2013, 2017) on ‘reason-responsiveness’.
- <sup>16</sup> One that’s not usefully addressed at this level of abstraction—e.g. without more detailed formulations of the truth-norm, and other candidate (fundamental/derivative) norms.
- <sup>17</sup> It’s of course controversial what it takes to have, and exercise, concepts. But note that a transition can be relevantly content-sensitive without the agent satisfying the ‘generality constraint’ (Evans, 1982) vis-à-vis the concept reason (or kin) (cf. Lavin, 2011, §4).
- <sup>18</sup> Contrast the models of deliberation found in (e.g.) Arpaly and Schroeder (2012), Broome (1999), McHugh (2014, 2017), Nolfi (2014, 2015), and Pettit (2010).
- <sup>19</sup> In my (2018), I called it ‘critical deliberation’.
- <sup>20</sup> Perhaps not an *indispensable* part. I take attitudes to be richly dispositional states, constituted by clusters of causal-functional roles (Lewis, 1972). But breakdown in this regard is certainly among the features that cast doubt on something’s status as a belief.
- <sup>21</sup> Except on a certain reductive approach to responsibility, which I lack the space to discuss further here (cf. fn. 3). See, e.g., Nolfi (2015), Strawson (1962), and Smithies (2016).
- <sup>22</sup> I also use ‘introspection’ for the capacity to take/use this route. (Correspondingly for ‘self-reflection’.)
- <sup>23</sup> Likewise for the extent to which it’s independent of evidence of *past* behavior (cf. Lawlor, 2009).
- <sup>24</sup> That it’s not strictly lack of *knowledge* that exculpates is stressed in, e.g., Harman (2011). Much of the surrounding debate concerns what it takes for ignorance to be non-culpable (e.g. Levy, 2009; Mason, 2015; Rosen, 2003, 2004). My argument doesn’t turn on controversial views on this.
- <sup>25</sup> More on exemptions in my *ms.*, Ch 6.
- <sup>26</sup> See Boghossian (2019) for a similar point and example.
- <sup>27</sup> McHugh (2013, 2014) seems to miss this, in his attempt to analyze fundamental responsibility in terms of reason-responsiveness alone. Reason-responsiveness, as McHugh understands it, is fairly close to what I call ‘content-sensitivity’ (in line with fundamental norms). But as such it’s at best part of the explanation of the sense in which the target attitudes are up to us.
- <sup>28</sup> This presupposes that normative ignorance can exculpate, but only in the weak sense that normative ignorance due to ignorance of normatively relevant facts can do so.
- <sup>29</sup> Many thanks to Michael Bratman, Pamela Hieronymi, David Hills, Lisa Miracchi, Ram Neta, Declan Smithies, and Sergio Tenenbaum for useful feedback on a previous draft. Thanks also to the audience at the 2<sup>nd</sup> KBNS workshop, Stirling Dec 2018.

## REFERENCES

- Alston, W. (1989). *Epistemic justification*. Ithaca: Cornell University Press.
- Anscombe, G. E. M. (1963). *Intention*. Oxford: Blackwell.
- Arpaly, N., & Schroeder, T. (2012). Deliberation and acting for reasons. *Philosophical Review*, *121*, 209–239. <https://doi.org/10.1215/00318108-1539089>
- Audi, R. (2008). The ethics of belief: Doxastic self-control and intellectual virtue. *Synthese*, *161*, 403–418. <https://doi.org/10.1007/s11229-006-9092-2>
- Baier, A. (1970). Act and intent. *Journal of Philosophy*, *67*, 648–658. <https://doi.org/10.2307/2024585>
- Bishop, J. (1989). *Natural agency: An essay on the causal theory of action*. New York: Cambridge University Press.
- Boghossian, P. (2003). The normativity of content. *Philosophical Issues*, *13*, 31–45. <https://doi.org/10.1111/1533-6077.00003>
- Boghossian, P. (2016). Reasoning and reflection: A reply to Kornblith. *Analysis Reviews*, *76*, 41–54. <https://doi.org/10.1093/analys/anv031>

- Boghossian, P. (2019). Inference, agency, and responsibility. In M. Balcerak-Jackson & B. Balcerak-Jackson (Eds.), *Reasoning: New essays on theoretical and practical thinking* (pp. 101–128). Oxford: Oxford University Press.
- Bratman, M. (1987). *Intentions, plans and practical reason*. Cambridge: Harvard University Press.
- Broome, J. (1999). Normative requirements. *Ratio*, 12, 398–419. <https://doi.org/10.1111/1467-9329.00101>
- Broome, J. (2013). *Rationality through reasoning*. Malden, MA: Wiley-Blackwell.
- Byrne, A. (2005). Introspection. *Philosophical Topics*, 33, 79–104. <https://doi.org/10.5840/philtopics20053312>
- Davidson, D. (1963). Actions, reasons and causes. *The Journal of Philosophy*, 60, 685–700.
- Davidson, D. (1978). Intending. *Philosophy of History and Action*, 11, 41–60.
- Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68, 5–20.
- Frankfurt, H. (1982). The importance of what we care about. *Synthese*, 53, 257–272.
- Harman, E. (2011). Does moral ignorance exculpate? *Ratio*, 24, 443–468.
- Hieronymi, P. (2006). Controlling attitudes. *Pacific Philosophical Quarterly*, 87, 45–74.
- Hieronymi, P. (2008). Responsibility for believing. *Synthese*, 161, 357–373.
- Hieronymi, P. (2014). Reflection and responsibility. *Philosophy and Public Affairs*, 42, 3–41.
- Hobart, R. E. (1934). Free will as involving determination and inconceivable without it. *Mind*, 43, 1–27.
- Kavka, G. (1983). The toxin puzzle. *Analysis*, 43, 33–36.
- Kornblith, H. (1983). Justified belief and epistemically responsible action. *Philosophical Review*, 92, 33–48.
- Kornblith, H. (2001). Epistemic obligation and the possibility of internalism. In A. Fairweather & L. Zagzebski (Eds.), *Virtue epistemology* (pp. 231–248). Oxford: Oxford University Press.
- Kornblith, H. (2012). *On reflection*. Oxford: Oxford University Press.
- Korsgaard, C. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.
- Lavin, D. (2011). Problems of intellectualism: Raz on reason and its objects. *Jurisprudence*, 2, 367–378.
- Lawlor, K. (2009). Knowing what one wants. *Philosophy and Phenomenological Research*, 79, 47–75.
- Levy, N. (2009). Culpable ignorance and moral responsibility: A reply to FitzPatrick. *Ethics*, 119, 729–741.
- Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249–258.
- Malmgren, A-S. (2018). Varieties of inference? *Philosophical Issues*, 28, 221–254.
- Mason, E. (2015). Moral ignorance and blameworthiness. *Philosophical Studies*, 172, 3037–3057.
- McDowell, J. (1994). *Mind and world*. Cambridge: Harvard University Press.
- McHugh, C. (2013). Epistemic responsibility and doxastic agency. *Philosophical Issues*, 23, 132–157.
- McHugh, C. (2014). Exercising doxastic freedom. *Philosophy and Phenomenological Research*, 88, 1–37.
- McHugh, C. (2017). Attitudinal control. *Synthese*, 194, 2745–2762.
- Moran, R. (2001). *Authority and estrangement*. Princeton: Princeton University Press.
- Nolfi, K. (2014). Why is epistemic evaluation prescriptive? *Inquiry*, 57, 97–121.
- Nolfi, K. (2015). Which mental states are rationally evaluable, and why? *Philosophical Issues*, 25, 41–63.
- Peacocke, C. (1979). Deviant causal chains. *Midwest Studies in Philosophy*, 4, 123–155.
- Peacocke, C. (1996). Entitlement, self-knowledge, and conceptual redeployment. *Proceedings of the Aristotelian Society*, 96, 117–158.
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
- Pettit, P. (2010). Deliberation and decision. In C. Sandis & T. O'Connor (Eds.), *A companion to the philosophy of action* (pp. 252–258). Malden, MA: Wiley-Blackwell.
- Rosen, G. (2003). Culpability and ignorance. *Proceedings of the Aristotelian Society*, 103, 61–84.
- Rosen, G. (2004). Skepticism about moral responsibility. *Philosophical Perspectives*, 18, 295–313.
- Scanlon, T. M. (1998). *What we owe to each other*. Cambridge: Harvard University Press.
- Schlosser, M. (2007). Causal deviance reconsidered. *Analysis*, 67, 186–194.
- Shah, N. (2003). How truth governs belief. *Philosophical Review*, 112, 447–482.
- Shah, S., & Velleman, D. (2005). Doxastic deliberation. *Philosophical Review*, 114, 497–534.
- Smith, H. (1983). Culpable ignorance. *Philosophical Review*, 92, 543–571.
- Smithies, D. (2016). Reflection on: *On reflection*. *Analysis Reviews*, 76, 86–94.
- Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 1–25.

- Tenenbaum, S. (2007). *Appearances of the good: An essay on the nature of practical reason*. Cambridge: Cambridge University Press.
- Tenenbaum, S. (2018). The guise of the guise of the bad. *Ethical Theory and Moral Practice*, 21, 5–20.
- Velleman, D. (1997). How to share an intention. *Philosophy and Phenomenological Research*, 57, 29–50.
- Wallace, J. (2018). Practical reason. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Spring, 2018. Retrieved from <https://plato.stanford.edu/archives/spr2018/entries/practical-reason/>
- Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 36, 267–297.
- Williams, B. (1973). *Problems of the self*. Cambridge: Cambridge University Press.
- Wright, C. (2012). Comment on Paul Boghossian, 'The nature of inference'. *Philosophical Studies*, 169, 27–37.

**How to cite this article:** Malmgren A-S. On fundamental responsibility. *Philosophical Issues*. 2019;29:198–213. <https://doi.org/10.1111/phils.12148>