

Explaining (away) the epistemic condition on moral responsibility¹

Gunnar Björnsson, Umeå University

Abstract: It is clear that lack of awareness of the consequences of an action can undermine moral responsibility and blame for these consequences. But when and how it does so is controversial. Sometimes an agent believing that the outcome might occur is excused because it seemed unlikely to her, and sometimes an agent having no idea that it would occur is nevertheless to blame. A low or zero degree of belief might seem to excuse unless the agent “should have known better”, but it is unclear how to spell out this normative condition.

This chapter combines (a) an independently motivated account of responsibility, blame, and credit as grounded in a normal explanatory relation between agential qualities and objects of responsibility with (b) the familiar Strawsonian idea that *moral* blame and credit depend on the agent’s quality of will. The resulting *explanatory quality of will* condition on moral responsibility is then further motivated by being shown to account for the effects on moral blame and credit of justifications, as well as of excuses and undermined control in cases not involving ignorance.

The explanatory quality of will condition is finally applied to cases involving various degrees of lack of awareness. Though this condition itself involves no awareness requirement, it is shown how it accounts for the degrees to which lack of awareness can excuse. It is also explained how lack of awareness fails as an excuse exactly when the agent should have known better and can be blamed for not doing so.

Ignorance or lack of awareness can undermine moral responsibility, moral blame, and moral credit. But *how* and *when* it does so is controversial. In some cases, its undermining effect is clear enough:

Sad Clown: Suffering from the heat, Vera gently opens the door of the waiting room to get some fresh air. Unbeknownst to her, a clown was balancing precariously on a tall stilt just outside the door and is pushed out of balance. The clown falls and suffers a severe concussion.

If Vera had known about the clown when opening the door, she would have been responsible and to blame for his fall. But she is not, because she did not foresee, and had no reason to suspect the effects of her action.²

¹ This paper has benefited from comments from audiences at the University of Gothenburg, Umeå University, and VU Amsterdam, written comments from Phil Robichaud, Jan Willem Wieland, Mark Zimmerman, Derk Pereboom, Kalle Grill, and Samuel Murray, and several rewarding discussions with George Sher. Work on this paper was funded by Riksbankens Jubileumsfond and the Swedish Research Council.

² “Moral responsibility”, as understood here, signifies the kind of responsibility in virtue of which agents deserve moral blame (moral guilt, indignation) or credit (moral gratitude) for bad or good objects of

Though agents' blamelessness in cases like *Sad Clown* are often explained with reference to the "epistemic" or "knowledge" condition on responsibility, what excuses here might not be lack of *knowledge* that something bad would happen. Vera could be to blame for the outcome even if she had *truly but unjustifiably believed* that the clown would be hurt (Peels 2014: 493–4). Moreover, what is required is not true belief that something *would* happen: if Vera had believed that she *might* hurt someone by opening the door, she might still have been to blame. But what excuses is not just lack of true belief that something might happen. If Vera believed that there was a small but *insignificant* risk, she might still be blameless in taking the risk and causing the outcome.

What excuses, then, might seem to be lack of belief that there is a *significant* risk for the negative outcome, a risk that would motivate further precautions. However, lack of such belief does not always excuse, at least not completely. Consider:

Knockout: Leaving the room, Victor pushes the door open quickly and with great force, inadvertently knocking unconscious the person just about to open the door from the busy corridor outside. At the moment of action, it didn't cross Victor's mind that opening the door in that way might hurt someone, though he would have realized this if the question had come up.

Badgering: Badgering his interlocutor with questions, Victor fails to pick up readily available signs of severe distress in her face and to adjust his behavior. His deeply insensitive interrogation, happening at an unfortunate time, causes her great discomfort.

Victor never considered the possibility that he might knock someone unconscious by opening the door, or cause great discomfort by his insistent questions. Still, most would be inclined to assign at least *some* blame to Victor for the effects (though certainly less than if he had brought these effects about knowingly).

One way to explain why Victor is responsible for the unfortunate effects in spite of his lack of awareness is to say that *he should have been aware*. However, "should" is ambiguous. In saying that Victor "should have been aware", one might just be saying that it would have been better if he had been. That sort of merely evaluative "should" seems too weak here, as it fails to indicate any moral fault of Victor's that would ground moral blame. On a relevant reading, by contrast, it would follow that Victor is *culpable* for not being aware of what he should have been aware of.³ But now a worry arises. Not only is it clear that lack of awareness undermines culpability in many cases. Awareness is also required for the sort of control that seems paradigmatic for agency—conscious control. Given this one might naturally conclude that responsibility in cases like *Knockout* and *Badgering* requires that the agent had the opportunity to become aware of the relevant risks *but consciously decided to remain ignorant*. If this right, then if Victor is to blame for what happens in *Knockout* and *Badgering*, the blameworthiness must be

responsibility (actions, omissions, outcomes). This is at least partly stipulative, as other relations too can make claims to that label. For example, one might think that Vera's causal responsibility for the clown's fall is a form of moral responsibility as it arguably grounds a special moral obligation to help the clown.

³ For an early discussion, see e.g. H. Smith 1983. For a recent discussion of should-have-knowns that do not ground culpability, see Goldberg 2015.

traced back to a prior conscious decision to take a certain risk for which the agent is directly responsible.⁴ That view, however, might seem to suggest that we are very rarely responsible in cases like Victor's, as it seems that we very rarely decide such things.⁵

Ideally, a full account of moral responsibility and the importance of awareness would help us understand under what conditions and to what extent lack of awareness excuses, but also the relevant *should have known* and the extent to which it might require prior awareness. The purpose of this chapter is exactly to provide such an account, and to explain how it clarifies the relevance of ignorance and awareness for moral responsibility. (For a very different, revisionary, response to various puzzles concerning the mitigating role of ignorance, see Peter Graham's contribution to this volume, where he denies that we are ever fundamentally blameworthy to such things as outcomes of actions.)

To provide a deep understanding of the importance of awareness for moral responsibility and moral blameworthiness, the account we seek should be more than an *ad hoc* collection of conditions meant to capture intuitions about cases involving various degrees of ignorance. Ideally, it should be grounded in an independently motivated understanding of responsibility and blameworthiness. To that end, section 1 first outlines an account of agential responsibility in general according to which responsibility requires a certain explanatory connection between relevant agential qualities and objects of responsibility, and then characterizes moral responsibility, blame and credit in particular in terms of the involvement of agent's "quality of will". Section 2 then clarifies and further motivates some of the core features of this account by applying it to a variety of cases not involving ignorance or lack of awareness of putative objects of blame. Section 3, finally, applies the account to cases that do involve such lack of awareness, explaining how it provides an understanding of why and when awareness matters for blameworthiness, and makes sense of the intuition that ignorance fails to excuse when the agent *should* have known.

One central consequence of the account is that although lack of awareness often undermines moral responsibility, such responsibility does not necessarily require awareness. But another upshot is that the responsibility-undermining effects of lacking awareness are straightforward consequences of the core condition of moral responsibility: when ignorance excuses, it does so because it breaks that explanatory connection between the agent's quality of will and the putative object of responsibility. In this sense, there is no *fundamental* or *independent* epistemic condition on moral responsibility.

This conclusion, though in line with broadly Strawsonian accounts of the excusing role of ignorance,⁶ runs contrary to what is sometimes assumed, namely that we need to add

⁴ For discussions of tracing conditions on moral responsibility, see e.g. Vargas 2005; Fischer and Tognazzini 2009; King 2012.

⁵ See e.g. Zimmerman 1997; Rosen 2003; 2004; Levy 2009; cf. Montmarquet 1995. The problem becomes particularly pronounced if one thinks that the awareness required is awareness of the wrongness of the action in question, rather than awareness of the non-moral properties that make it wrong.

⁶ See e.g. Strawson 1962; Wallace 1994; Arpaly 2003, e.g. 159–162; Harman 2011; Vargas 2013, ch. 8; Arpaly and Schroeder 2014; Mason 2015.

epistemic requirements to some independently identifiable control, freedom, or ownership condition on responsibility. Here, for example, are John Fischer and Mark Ravizza:

We shall assume that our analysis [of the control condition on moral responsibility] can be supplemented with a component that would attend to the epistemic condition on responsibility (1998: 13–14, n. 21).

Similarly, Derk Pereboom’s way of understanding free will and moral responsibility seems to presuppose that an epistemic condition can be added to an independent ownership condition:

For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary (2014: 2).

As a final example, Michael McKenna (2012, Ch. 1) suggests that the control and epistemic conditions need to be supplemented by a quality of will condition. If the view outlined in this chapter is correct, however, responsibility and blame involves one unified condition—one making reference to the quality of will of the agent. This condition accounts for the clear cases normally handled by control, epistemic and quality of will conditions, while providing independently motivated guidance when we want to understand cases like *Knockout* and *Badgering*.

1. The explanatory quality of will condition on moral responsibility

In this section, I first outline a general account of the kind of responsibility that grounds agent-directed blame and credit, whether moral or non-moral. This general account is then applied to moral responsibility specifically, guided by the familiar idea that moral blame and credit are sensitive to the agent’s quality of will. The upshot is a simple but powerful theory of the sort of responsibility that grounds moral blame and credit specifically, providing unified explanations of a variety of familiar facts about moral responsibility, including the role and strength of various non-epistemic forms of excuses. In the next section, this account is then applied to questions about how lack of awareness undermines responsibility.

In general, it makes sense to blame people (morally or otherwise) for some of their qualities and behaviors, but not others: perhaps for how they dress, treat their children, solve a theoretical puzzle, or build a new kitchen cabinet, but not for how they fall to the floor when knocked unconscious; perhaps for their clumsiness, ignorance, and lack of respect, but not for their date of birth, the color of their skin, or their need for sleep. (Likewise for credit or praise, though I will mostly focus on the negative.)

Looking at what unifies the wide variety of phenomena for which (agent-directed) blame makes sense and distinguishes them from others, blame seems appropriate only when implicating agential aspects that are relevantly *responsive* to blame: aspects that can at least in principle improve in response to recognition of fault (though not necessarily by the agent blamed on the occasion; we unhesitatingly blame people whose opportunities for improvement have passed).

For blame to make sense, the agential aspect in question must constitute a *fault*—a state of the agent falling enough short of some ideal that it is a proper subject of criticism. Moreover, the object of blame—what the agent is blamed for—must itself be bad, and explained by the fault in question. For example, I can be blamed for the sorry state of a kitchen cabinet if it is the result of some fault of mine (lacking craftsmanship, or laziness), but not if it is explained by the unavailability of proper building materials or tools, or someone else’s vandalism, or if I had nothing whatsoever to do with the kitchen cabinet. In addition, the explanation has to be *normal* in some relevant sense. Suppose that I am just about to finish a decent kitchen cabinet in spite of my poor craftsmanship, when some spiteful person secretly changes the dimensions of some of the parts because they think that someone with my laughable skills doesn’t deserve a nice cabinet. Then the end result is a bad cabinet, and it is bad in part because of my poor craftsmanship—it is what triggered the sabotage. Still, since the explanation is abnormal, I am not to blame for the cabinet’s sorry state.

Agential aspects that are open to criticism and praise can be ordered in evaluative dimensions: craftsmanship (understood as level of skill) ordered from the poorest to the greatest, mathematical understanding from non-existing to superhuman, regard for others from abysmal to saintly, and so forth. Corresponding to those aspects are kinds of events (mental episodes, actions, omissions, outcomes) that are normal upshots of such aspects and can be ordered in corresponding evaluative dimensions: craftsmanship (understood as quality of products) or mathematical reasoning ranging from the poorest to the most excellent and social behavior ranging from the least to the most respectful. When we blame and credit we do so against an understanding of how evaluatively ordered agential aspects explain correspondingly ordered events in typical or normal ways.⁷

The suggestion, then, is that responsibility of the sort that grounds blame (or credit) should be understood in terms of the following *explanatory* condition:

EXPLANATORY RESPONSIBILITY: For X to be responsible for Y is for Y to fall in an evaluative dimension and be explained in a normal way by a corresponding aspect of X⁸.

Notice that on the proposed model, agential aspects that ground responsibility can themselves be objects of responsibility if they are explained in normal ways by other agential aspects. However, there is no general requirement that agents are responsible for the relevant agential aspects. Notice also that EXPLANATORY RESPONSIBILITY itself allows that objects of responsibility can be good or bad in virtue of the agential qualities explaining them, allowing

⁷ To say that a certain explanation of events by agential aspects is “normal” is not to say that when such an aspect occurs the aspect most often explains the event in this way, but rather to say that *when* such an aspect does explain that sort of event, it often does it in this way (cf. Millikan’s 1984 notion of a (capital N) Normal explanation). We might also want to add that the relevant cases should be ones that the practice is designed to track: common but unknown explanatory connections do not seem to ground responsibility. These remarks still leave open exactly how the normalcy requirement should be understood, but I take considerations of abnormal cases to make it clear enough that there is some such requirement. A longer discussion would go beyond the constraints of this contribution.

⁸ This suggestion can be seen as generalizing the proposal of Björnsson and Persson 2012.

for praise for failed efforts to bring about something good, and blame for recklessly risking bad outcomes that never materialized. Finally, notice that this account has no additional clause meant to handle so-called *indirect* responsibility. It is natural to think the drunk driver can be responsible for an accident over which she had no control at the time in virtue of being responsible for some earlier event (getting drunk, getting behind the wheel). But such cases are straightforwardly handled by EXPLANATORY RESPONSIBILITY, as the bad outcome can be explained in normal ways by an earlier bad state of the driver, such as her insufficient concern for risks for potential victims (cf. King 2012).

Whether X deserves blame or credit for Y will next depend on whether Y is bad or good and on what states can be properly demanded of X:

EXPLANATORY BLAME (CREDIT): X deserves blame (credit) for Y if and only if Y is bad (good) and is explained in a normal way by X's corresponding aspect falling below (above) what could be properly demanded of X.

What can be properly demanded of X will depend on a variety of factors, factors that might in turn depend on what domain of evaluation we are concerned with, and perhaps the context of evaluation. Particularly relevant factors might include individual capacities—we often demand less of a child than a fully developed adult, and less of beginners than accomplished participants in the relevant practice—as well as difficulty in attaining and maintaining certain qualities under the circumstances.

What agential states can be properly demanded determines not only whether someone deserves blame or credit for something, but also the *magnitude* of what is deserved. The carpenter producing an excellent kitchen cabinet deserves more credit if she succeeded in spite of sparse materials and little time than if she worked without such constraints (other things being equal): in the former case, success required the agent to *further exceed* what can be demanded in terms of skill and dedication. Similarly, a carpenter producing a somewhat defective cabinet under difficult circumstances deserves less blame than one working under ideal circumstances, as failing under such circumstances requires *less of a shortcoming* in skill and determination. Generalizing this, and adding that the goodness or badness of the object of credit or blame matters, we can say:

DEGREE FROM DEVIATION: The degree of blame (credit) X deserves for Y depends on the value of the outcome and on how much of a deviation of the agential aspect from what can be properly demanded is required in the (normal) explanation of the object of blame (credit).

Notice that the degree of blame or credit depends on the degree of deviation *required to explain the object of responsibility*. If Joan and Jane make equally good kitchen cabinets, the fact that Joan is the better craftsman does not mean that she deserves more credit for hers, as Joan's superior skills might not be needed to explain her success. Notice also that it depends on the degree of deviation *from what can be properly demanded*, which, as before, might depend on a variety of

factors, including the capacities of the agent and the situation: if Joan operates under worse conditions, she might deserve more credit even Jane’s cabinet was equally good.⁹

Thus far, the account is designed to cover blame and credit generally, whether the evaluative dimensions are ones of craftsmanship, mathematics, sports, rationality, or morality. Turning to *moral* blame and credit in particular, I will follow the tradition of Strawson’s (1962) “Freedom and Resentment” that takes moral credit and blame to relate to the “quality of will” of the agent.¹⁰ I understand the quality of will of an agent as the extent to which the agent cares about how well things go compared to what can be properly morally demanded of her.

To care about something in some regard in the relevant sense is to be in a state particularly disposing one to react and act in ways conducive to its going well with it in that regard: to pay attention to information that might be relevant for how well it goes with it and to take the fact that some action would make it go better as reason for that action (i.e. to be motivated to make it go better).¹¹

The normal way in which morally ideal caring explains morally good events is simply that it leads agents to detect cases where they can act so as to promote or not obstruct such events and moves them to adjust their behavior accordingly even when it is costly. The normal way in which lack of such caring explains morally bad events is that agents lacking it ignores evidence of such cases and to fail to be motivated to adjust their behavior when they do recognize the opportunity, because they care more about other things and so act on those, or care about morally bad things. When someone cares as could be reasonably morally demanded of them but not more, that will normally explain why morally neutral routes of action are taken rather than the morally exemplary or awful.

Applying the quality of will notion of moral aspects of agents to EXPLANATORY BLAME (CREDIT), we get:

MORAL EXPLANATORY BLAME (CREDIT): X deserves moral blame (credit) for Y if and only if Y is morally bad (good) and is explained in a normal way by X’s quality of will falling below (above) what could be properly morally demanded of X.¹²

⁹ A further complication, set aside here, is that the function from deviation to blame might itself be sensitive to such factors.

¹⁰ Some of the most sophisticated and influential such accounts have been formulated in terms of the quality of will *revealed* in action, rather than in terms of an explanatory relation, as when Nomy Arpaly talks about someone’s *proving* himself lacking in moral concern or about acts *speaking well or ill* of the agent (2003: 161–2) It seems to me that thinking of the relation as evidential fails to get at the heart of the matter: ill will can be revealed in ways that do not ground blame, and wrongful acts that could not in principle be revealed—negligently, instantly and irreversibly wiping out all cognizant life, say—would not be necessarily blameless. But the mistake is understandable: if we know that something is required for the explanation of Y, we will typically be able to infer its existence from Y. (In the latest version of Arpaly’s view (Arpaly and Schroeder 2014, ch. 7), the explanatory relation of realization provides the substance to seemingly evidential talk of “manifestation”.)

¹¹ Caring so understood is a certain kind of responsiveness to reasons (cf. Fischer and Ravizza 1998). For a related view of caring and its relation to responsibility, see Sripada Forthcoming.

¹² For discussion of this kind of understanding, see Björnsson 2011; Björnsson and Persson 2012; 2013; Björnsson 2014a. In the literature, Nomy Arpaly and Timothy Schroeder’s recent (2014, ch. 7) account is a

A few things are worth pointing out before we begin applying this idea:

First, what can be properly demanded might be a relative level of caring: to care about respecting the basic rights of others more than one cares about doing one's job, say. Consequently, one might be to blame for how one treats others even if one does care about them (cf. Baron 2014).

Second, what can be properly morally demanded of an agent presumably depends on the agent's *capacity* to care, which might in turn depend on both situational and internal psychological factors.¹³

Third, MORAL EXPLANATORY BLAME (CREDIT) allows for blame and credit for what is out-of-character. Even if someone generally cares about others to an extent that can be properly demanded, she might on occasion care less than this. Sometimes such deviations of caring will be due to situational factors that also lower what can be properly demanded (perhaps she just witnessed a fatal accident, or is blamelessly under the influence). But when they are not, they might ground blame. (*Mutatis mutandis* for credit.)

Fourth, in taking moral blameworthiness to depend on a substandard quality of will, MORAL EXPLANATORY BLAME (CREDIT) takes a controversial stand on a class of cases discussed in the literature on the mitigating role of ignorance. These are cases where someone fails to remember something important and seem to blame for it even though it is stipulated that this is not for lack of proper concern with values at stake (see Sher 2009; Clarke 2014). As I argue elsewhere, however, the blame properly attributed in such cases is not *moral* blame, but rather a form of *skill* blame (Björnsson manuscript).

close relative to MORAL EXPLANATORY BLAME (CREDIT) for the special case of responsibility for action. One difference is that their account does not invoke any notion of what quality of will can be properly demanded of an agent; another is that, like most people in the literature, they understand blameworthiness in terms of wrong action. On my preferred view, however, the *primary* notion of wrongness is itself defined in terms of blameworthiness: some behavior is wrong insofar as you would be to blame for doing it. (Other notions can be defined in terms of this. For example, we might say that an action is *objectively* wrong insofar as one would be to blame for it if one knew all the relevant facts, and that an action of kind K is *pro tanto* objectively wrong insofar as one is blameworthy for performing actions of kind K unless one has a complete excuse or special justification for doing it.)

Given MORAL EXPLANATORY BLAME (CREDIT), requirements to care about certain matters are the ultimate foundations of moral responsibility; elsewhere, I have argued that they are also central to obligations (Björnsson 2014a) and forward-looking or normative responsibilities (Björnsson and Brülde forthcoming). One might sensibly worry, though, that moral credit- or blameworthiness can be grounded in cognitive, conative, or affective states other than states of (not) caring (I thank Michael Zimmerman for pushing me on this). Given more space to respond to this, I would argue that when we correctly take such other factors to be relevant for moral blame and credit, it is because they are indicative or constitutive of sub- or superstandard caring.

¹³ Plausibly, the capacity to care about specifically moral reasons is lacked by most non-human animals, held only to a limited degree by small children and those suffering from some serious mental illnesses, but often present in the merely foolish (thus answering Chris Franklin's (2013: 481) question why "insanity exempt one from responsibility, while foolishness does not"). For discussion of the vexed issue of what such capacities to care are relevant, see e.g. Shoemaker 2013; Vargas 2013.

Finally, DEGREE FROM DEVIATION applies to *moral* blame and credit in just the same way as to other forms.

2. Non-epistemic applications: justification, excuses, control

MORAL EXPLANATORY BLAME (CREDIT) encapsulates two general ideas: that moral blame and credit depend on the quality of will of the agent, and that blame and credit generally are directed at the normal bad or good upshots of an agent's fault and excellence, respectively. Accordingly, it tells us that X's moral blame and credit for Y are undermined when:

- (a) nothing about X explains Y,
- (b) X cared appropriately about morally relevant aspects (in which case blame is undermined) or didn't care more than can be reasonably demanded (in which case credit is undermined),
- (c) X has the quality of will that could ground blame / credit, but it does not explain Y, and
- (d) X has the relevant quality of will that could ground blame / credit, but it does not explain Y in a normal way.

The basic suggestion, broadly familiar from proponents of quality of will accounts of blameworthiness and developed in the next section, will be that ignorance typically undermines blame by breaking the required explanatory connection between quality of will and object of responsibility, giving us situations of type (b), (c) or (d). In this section, we will consider various ways in which blame is and is not undermined, focusing on cases where agents are aware of morally relevant aspects.¹⁴

Begin with cases where, although Y occurred because of something the agent did or did not do, the agent's action or omission was *justified*. Perhaps when Paula pushes some people brusquely to the side to catch the subway train, she is not to blame for their inconvenience, as she needed to urgently get life-saving medication to a patient two stops away. One might want to say that Paula would then be blameless because she *did nothing wrong*. That might seem plausible, but it is important to notice that the action might not be *objectively* right. Perhaps the patient expires before Paula gets there, or perhaps someone else has unexpectedly provided the medication. In that case, no objective outcome justified brusquely pushing people around, but Paula might still be blameless if her action was right *given the evidence available to her*.¹⁵

¹⁴ Franklin (2013) suggests that what excuses invoke are factors undermining reasonable opportunity, and thinks that this gives excuses their normative significance (cf. Nelkin's (2014) suggestion that difficulty affects blame- and praiseworthiness by affecting quality of opportunity). On the current view, that significance is provided by the relation of being explained in a normal fashion by the agent's quality of will.

¹⁵ I say 'might', for two reasons. First, as discussed in the next section, she might be to blame for not having recognized (and acted on) better evidence. Second, even if she did what was permissible given the evidence, she might have done so for wrong reasons, out of impatience say (see below). Doing so might itself be morally bad, and so ground blame.

However, the most straightforward way of understanding the latter form of rightness is exactly in terms of whether the agent acted on the evidence as one would if having the appropriate quality of will.¹⁶ If so, the explanation of Paula's blamelessness in terms of the permissibility of her action is not significantly different from that provided by MORAL EXPLANATORY BLAME (CREDIT): although Paula pushed some people brusquely, what explained this was not a substandard will.¹⁷

Depending on circumstances, justification might only provide partial mitigation of blame. If Paula had a less urgent reason to catch the train—perhaps the patient's condition was not life threatening but only extremely painful—she might still be to blame, as she might have cared too much about her own mission and too little about the people blocking her way. But she would seem to be less blameworthy than she would have been if she had pushed people to the side just to save herself from spending 10 minutes underground waiting for the next train. This seems well accounted for given DEGREE FROM DEVIATION. To push people to the side to save someone from 10 minutes of excruciating pain might require some lack of regard for others, but not as much as doing so to save oneself from waiting. Paula the Samaritan thus deserves less blame for pushing people brusquely to the side on her way to the train than does Paula the Impatient.¹⁸

As generally with DEGREE FROM DEVIATION, the degree of blame deserved for something does not depend on how much the agent falls below what can be properly demanded of her, only on to how much of a deviation is required to explain the object of blame. If Paula and Pete both push people to the side to catch a train to save themselves from waiting and Pete has much less regard for others than Paula, then they might still be equally to blame for this particular act: the degree to which one must fall short of proper regard to do such a thing might be the same for both.¹⁹ (A complication here is that the moral value of one's action is at least partly determined by the quality of will that is involved in explaining the action; something is an act of recklessness, for example, in virtue of being the upshot of insufficient concern with risks. Even if a certain lack of regard is required to explain Paula's and Pete's actions, it could be that Pete nevertheless acts from a deeper lack of regard or even from ill will, and so performs an action that is in this sense worse. For the purpose of this case, I am assuming that Pete's deeper lack of regard is never engaged.)

Turn next from justification to excuses appealing to external *force*, *physical constraints*, *psychological conditions*, or *threats*. When someone is physically forced to do something, what she does will typically not be explained in a normal way by her quality of will. If physical force is not completely overwhelming but resisting it is difficult and costly, the resulting action might be explained by the agent's quality of will, but difficulty and cost might decrease blameworthiness along the lines just discussed in relation to justification: when doing good is

¹⁶ Cf. Björnsson 2014a: 114–16.

¹⁷ Cf. n. 12.

¹⁸ Cf. McKenna 2012, Ch. 3.

¹⁹ Cf. Arpaly and Schroeder's (2014: 188–89) response to Markovits 2010.

difficult, less of a substandard will is required to explain the agent's failure.²⁰ Similarly, bad actions performed under conditions of exhaustion, agitation or inebriation might be fully or partially excused because bad action under such conditions might not be explained in a normal way by substandard caring. Such actions might not reflect what the agent generally cares about, and levels of care for properly demanded for such conditions specifically might be lower than as the conditions might make higher levels harder to uphold or achieve.²¹ Likewise for threats, which might fully or partially *justify* an otherwise bad action if performing it in response to the threatened consequences would require less or no substandard caring, and fully or partially *excuse* the action if putting the agent in an abnormal psychological condition where normal concerns are not at work and where the same levels of concern cannot be properly demanded.

It is easy to come up with a variety of further ways in which explanatory connections between quality of will and putative objects of responsibility might break down: muscle spasms, brain hemorrhage, random errors of reasoning, or standard philosophical illustrations of invasive manipulation, neuroscientific intervention, and involuntary hypnotism. These and previous cases where factors excuse by changing the normal explanatory connection to the agent's quality of will can be intuitively understood as cases of diminished or lost *control* over the putative object of responsibility, and one might think that this is a better way of understanding them. On closer inspection, however, the control in question is one that essentially involves the connection to the agent's quality of will. To bring this out, consider a case where the agent knew that she could prevent a certain bad outcome at no significant cost to herself, but where she is nevertheless not to blame for not preventing it:

The Catch: There was only one way in which you could prevent the deadly explosion, and you knew it. If you were to press the button, you would prevent the explosion, *but only if you would not be acting for ultimately moral reasons, out of concern or respect for potential victims.* (Those meddling neuroscientists were at it again, monitoring your deliberation and tracking your motivation). Moreover, you were able to press the button, and able to prevent the deaths by doing so for non-moral or immoral reasons: perhaps to listen to the somewhat interesting sound emitted by the button, or to save the explosive device for an even deadlier occasion later on. As it happened, however, you did not at the moment care about those other things. So you did not press the button.

In this case, you do not seem to be to blame for not preventing the explosion or for the resulting deaths. Suppose first that you did care as can be properly morally demanded about the deadly effects of the explosion. Then you could not act based on this care to prevent the

²⁰ "Why", Chris Franklin (2013: 481) asks, "does physical constraint excuse one from responsibility, while forgetfulness or laziness does not? His answer is that only the former deprives us of the right sort of reasonable opportunities to avoid the object of blame. The answer offered here is instead that the latter are often upshots of substandard quality of will rather than factors blocking normal explanatory connections between quality of will and action.

²¹ As Dana Nelkin (2014: 8–9) points out, such cases seem problematic for Arpaly and Schroder's (2014) account.

explosion. Though you could bring yourself to be interested in the sound emitted by the button, say, if you were to do so for moral reasons and it lead you to press the button, this would not have prevented the explosion—that's the catch. Furthermore, your failure to press the button for entirely non-moral reasons hardly qualifies as a moral fault capable of grounding moral blame for the outcome. Moreover, all this remains true even if you did *not* in fact care about preventing the deadly effects.²²

What *The Catch* brings out is that even when an agent has perfect knowledge about consequences and control over which consequences to bring about, moral responsibility requires that the consequences are properly related to—explained by—the agent's quality of will. Once we spell out that requirement, however, as I have done in MORAL EXPLANATORY BLAME (CREDIT), there is no need for any *further* control requirement: the explanatory quality of will requirement, properly spelled out, already provides the relevant condition of *moral control*, we might say. In the next section, we will see how the same condition also accounts for the varying effects of ignorance on blameworthiness.

3. The explanatory quality of will condition and lack of awareness

Thus far I have introduced and motivated an account of when agents deserve moral blame and praise, and have done so based on cases where agents are aware of putative objects of responsibility. Elsewhere, I have argued that attention to the explanatory component of moral responsibility helps us understand a variety of otherwise puzzling features of responsibility and responsibility attributions, including shared responsibility and various recurring skeptical worries about moral responsibility.²³ Others, such as Nomy Arpaly and Michael McKenna, have done much to elucidate notions of quality of will and their relation to responsibility, blameworthiness and praiseworthiness.²⁴ The overall idea thus has considerably independent support and substance quite independently of its application to cases involving lack of awareness, to which we now turn.

Some of these applications are clear enough, and familiar from the literature.²⁵ Start with the observation that, like Vera in *Sad Clown*, an agent might be excused for bringing about or letting happen some bad event when she lacked the belief that she would significantly risk doing so. Lacking such a belief, her caring or not caring about that sort of risk would typically not explain why the bad event took place. Given MORAL EXPLANATORY BLAME (CREDIT), she would thus not be to blame for it. By the same token, she would *not* have been excused by lack of knowledge of the outcome if she had nevertheless believed that she was taking a significant risk that it would occur. An explanation of why she did not take further precautions under such circumstances would need to invoke a substandard concern for the safety of others.

²² Perhaps you could be blamed for not caring (if you are responsible for that), or for not trying (in vain, alas) to find a way to prevent the outcome. But that's significantly different from being blamed for the death of the victims.

²³ Björnsson 2011; Björnsson and Persson 2012; 2013; Björnsson 2014b; Björnsson and Pereboom 2015.

²⁴ Arpaly 2003; 2006; McKenna 2012; Arpaly and Schroeder 2014.

²⁵ See e.g. Arpaly 2003: 160–62; A. M. Smith 2005; H. M. Smith 2011.

While explaining how lack of belief about risks can excuse, MORAL EXPLANATORY BLAME (CREDIT) also straightforwardly explains how Victor might be responsible for the outcomes in *Knockout*, where he quickly and forcefully opens a door and knocks unconscious the person about to open it from the outside, and *Badgering*, where his continued questions cause great discomfort. In both cases there was information available in memory or perception based on which Victor could have realized that he might be putting others in danger or distress, and we might naturally conclude that his failure to do so and adjust his behavior accordingly is explained by a lack of appropriate concern for such outcomes.

Admittedly, this account does not explicitly make reference to our natural justification of blaming Victor in spite of his lack of awareness of the dangers, namely that *he should have realized* that he might be hurting others. But this justification follows straightforwardly from what has already been said together with MORAL EXPLANATORY BLAME (CREDIT). Victor's failing to notice the risks of hurting others is in itself something bad, and thus something for which he might be to blame if it is normally explained by his failure to care appropriately. Moreover, on the proposed account, his failure to notice *is* normally explained by the fact that he does not care sufficiently about not hurting others, as he would have picked up on the information available if he had cared. Given this, it follows that he should have realized that he might be hurting others, in a culpability-entailing sense: it was bad that he did not realize this, and he is to blame for not doing so. Moreover, the facts in virtue of which this is true are indeed facts justifying the conclusion that Victor is responsible for the outcomes.

As these cases illustrate, MORAL BLAME (CREDIT) does not draw a sharp line between cases where agents are and are not aware of the badness of acts and outcomes. Other views take direct blameworthiness to accrue only to cases clear-eyed akrasia, where the agent know at the time is wrong,²⁶ or to cases where we are aware of morally relevant features of actions. But these views are both revisionary—contrary to ordinary attributions of blame—and hard to motivate. Consider six kinds of cases where an agent is uniquely placed to save a person from severe agony but fails to do so:

- (1) X knows that not helping is morally wrong because she is uniquely placed to help and can do so at very small personal cost, but she nevertheless refrains because she cares little about the plight of others or about morality.
- (2) X knows that she is uniquely placed to help but refrains from helping because she cares little about the plight of others and is committed to a form of egoistic morality according to which her small personal benefit justifies not helping.
- (3) X knows that she is uniquely placed to help but cares too little about morality generally to conclude that the action is morally wrong, and too little about the plight of others to help.
- (4) X has all the evidence available to conclude that she is uniquely placed to help and that it would be wrong not to, but cares too little about the plight of others or about morality generally to draw those conclusions.

²⁶ See e.g. Zimmerman 1997; Rosen 2003; 2004; Levy 2009; cf. Wallace's Wallace 1994 focus on the quality of intentions. For critical discussion, see e.g. Arpaly 2003; Harman 2011; Mason 2015.

- (5) X had signs available that would have led anyone caring about the plight of others and about moral matters to look up, take in the situation, reflect, realize that she is uniquely placed to help at little personal cost, conclude that she ought to do so, and jump to action.
- (6) X had no such signs available.

We would ordinarily be ready to blame the agent morally in cases (1) through (5) but not in (6), and MORAL EXPLANATORY BLAME (CREDIT) provides a unified account of this phenomenon: in the first five cases, failure to care sufficiently about morally significant features explains, in normal ways, why the agent fails to prevent the suffering. Moreover, in conjunction with DEGREE FROM DEVIATION, MORAL EXPLANATORY BLAME (CREDIT) straightforwardly explains why blameworthiness tends to diminish as we move down the list. It requires a more radically substandard quality of will to refrain from saving a person from severe agony at little personal cost when one is aware that this is what one is doing and aware that it is wrong than to ignore signs suggesting more or less vaguely that something about one's action might be problematic (cf. Arpaly 2003: 160–62).

MORAL EXPLANATORY BLAME (CREDIT) draws the sharpest line at a natural place: between (5) and (6), i.e. between cases where it seems natural to assign at least some degree of blame, and cases where no blame seems appropriate. If we required blameworthiness to be grounded in clear-eyed *akrasia*, we would draw a sharp line between (1) on the one hand and (2) and (3) on the other, and if we required conscious awareness of the morally problematic features of the action, we would draw one between (3) on the one hand and (4) and (5) on the other. But neither of these lines correspond to radical discontinuities in our ordinary attributions of responsibility. Accounts postulating such discontinuities need strong justification, especially given the availability of an account making sense of not only our ordinary sense of continuity, but also of the distribution of blame and credit more generally.²⁷

4. Concluding remarks

In this chapter, I have motivated a certain understanding of responsibility, blameworthiness and praiseworthiness in general and, based on this, an understanding of specifically moral responsibility, blameworthiness and praiseworthiness. Central to this understanding is the idea that responsibility requires a normal explanatory link between the object of responsibility and a corresponding state in the agent. In the moral case, that state is the agent's quality of will. I have further explained how this sort of account captures various ways in which blameworthiness might be undermined in cases not involving ignorance or lack of awareness, also in cases where the agent could have consciously chosen to prevent the putative object of blame, and how it can plausibly explain attributions of degrees of blame. In the final section, I

²⁷ I do not want to deny the intuitive pull of some of the arguments offered by Zimmerman, Rosen, Levy and others, and one might well feel that something important has been lost by way of control when there is no conscious awareness of the situation, and so think that a revisionary account is better. A full defense of MORAL EXPLANATORY BLAME (CREDIT) against discontinuity views requires addressing these arguments, and the intuitive importance of the conscious perspective. I do so in Björnsson manuscript.

have indicated how this sort of account provides plausible explanations of how varying degrees of awareness corresponds to varying degrees of blameworthiness.

If the account is correct, there is no fundamental epistemic condition on moral responsibility. Lack of awareness often does undermine responsibility, but it does not always do so, and when it does, it does so by undermining a much more general explanatory condition on responsibility, a condition also accounting for phenomena standardly handled with appeal to control or freedom conditions of moral responsibility. The so-called epistemic condition on moral responsibility has thus been explained away: the circumscribed relevance of ignorance and lack of awareness for moral responsibility is better explained with reference to the explanatory condition on responsibility.

Obviously, a brief chapter will have to leave many problems unexplored. There is much more to say about what constitutes relevant qualities of will, or moral blame, and about what qualities of will can be reasonably demanded in various context. Still, the overall picture of responsibility offered by quality of will approach to moral blame, and the particular version offered by MORAL EXPLANATORY BLAME (CREDIT), strikes me as a promising one, capable of accounting for natural attributions of moral blame in cases displaying various degrees awareness and ignorance without ad hoc epistemic conditions.

Bibliography

- Arpaly, Nomy 2003: *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.
- Arpaly, Nomy 2006: *Merit, Meaning, and Human Bondage: An Essay on Free Will*. Princeton, NJ: Princeton University Press.
- Arpaly, Nomy and Schroeder, Timothy 2014: *In Praise of Desire*. Oxford University Press.
- Baron, Marcia 2014: ‘Culpability, Excuse, and the ‘Ill Will’ Condition’. *Aristotelian Society Supplementary Volume*, 88, pp. 91-109.
- Björnsson, Gunnar 2011: ‘Joint Responsibility without Individual Control: Applying the Explanation Hypothesis’. In *Moral Responsibility: Beyond Free Will and Determinism*. Vincent, Nicole, van de Poel, Ibo and van den Hoven, Jeroen (eds) Dordrecht: Springer pp. 181–99.
- Björnsson, Gunnar 2014a: ‘Essentially Shared Obligations’. *Midwest Studies in Philosophy*, 38, pp. 103–20.
- Björnsson, Gunnar 2014b: ‘Incompatibilism and ‘Bypassed’ Agency’. In *Surrounding Free Will*. Mele, Alfred (ed) New York: Oxford University Press pp. 95–122.
- Björnsson, Gunnar manuscript: *Explaining Away Epistemic Skepticism About Culpability*.
- Björnsson, Gunnar and Brülde, Bengt forthcoming: ‘Normative Responsibilities: Structure and Sources’. In *Parental Responsibility in the Context of Neuroscience and Genetics*. Hens, Kristien, Horstkötter, Dorothee and Cutas, Daniela (eds) Dordrecht: Springer.
- Björnsson, Gunnar and Pereboom, Derk 2015: ‘Traditional and Experimental Approaches to Free Will and Moral Responsibility’. In *Companion to Experimental Philosophy*. Sytsma, Justin and Buckwalter, Wesley (eds) Blackwell.
- Björnsson, Gunnar and Persson, Karl 2012: ‘The Explanatory Component of Moral Responsibility’. *Noûs*, 46, pp. 326–54.
- Björnsson, Gunnar and Persson, Karl 2013: ‘A Unified Empirical Account of Responsibility Judgments’. *Philosophy and Phenomenological Research*, 87, pp. 611–39.

- Clarke, Randolph 2014: 'Negligent Action and Unwitting Omission'. In *Surrounding Free Will*. New York: Oxford University Press pp. 298–317.
- Fischer, John Martin and Ravizza, Mark 1998: *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge U. P.
- Fischer, John Martin and Tognazzini, Neal A. 2009: 'The Truth About Tracing'. *Noûs*, 43, pp. 531–56.
- Franklin, Christopher Evan 2013: 'A Theory of the Normative Force of Pleas'. *Philosophical Studies*, 163, pp. 479–502.
- Goldberg, Sanford C 2015: 'Should Have Known'. *Synthese*, pp. 1–32.
- Harman, Elizabeth 2011: 'Does Moral Ignorance Exculpate?'. *Ratio*, 24, pp. 443–68.
- King, Matt 2012: 'Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility'. *European Journal of Philosophy*, pp. n/a–n/a.
- Levy, Neil 2009: 'Culpable Ignorance and Moral Responsibility: A Reply to Fitzpatrick'. *Ethics*, 119, pp. 729–41.
- Markovits, Julia 2010: 'Acting for the Right Reasons'. *Philosophical Review*, 119, pp. 201–42.
- Mason, Elinor 2015: 'Moral Ignorance and Blameworthiness'. *Philosophical Studies*, pp. 1–21.
- McKenna, Michael 2012: *Conversation and Responsibility*. New York: Oxford University Press.
- Millikan, Ruth Garrett 1984: *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
- Montmarquet, James A 1995: 'Culpable Ignorance and Excuses'. *Philosophical Studies*, 80, pp. 41–49.
- Nelkin, Dana Kay 2014: 'Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness'. *Noûs*.
- Peels, Rik 2014: 'What Kind of Ignorance Excuses? Two Neglected Issues'. *The Philosophical Quarterly*, 64, pp. 478–96.
- Pereboom, Derk 2014: *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Rosen, Gideon 2003: 'Culpability and Ignorance'. *Proceedings of the Aristotelian Society*, 103, pp. 61–84.
- Rosen, Gideon 2004: 'Skepticism About Moral Responsibility'. *Philosophical Perspectives*, 18, pp. 295–313.
- Sher, George 2009: *Who Knew?* Oxford U. P.
- Shoemaker, David 2013: 'Qualities of Will'. *Social Philosophy and Policy*, 30, pp. 95–120.
- Smith, Angela 2005: 'Responsibility for Attitudes: Activity and Passivity in Mental Life'. *Ethics*, 115, pp. 236–71.
- Smith, Holly 1983: 'Culpable Ignorance'. *The Philosophical Review*, 92, pp. 543–71.
- Smith, Holly M 2011: 'Non-Tracing Cases of Culpable Ignorance'. *Criminal Law and Philosophy*, 5, pp. 115–46.
- Sripada, Chandra Forthcoming: 'Self-Expression: A Deep Self Theory of Moral Responsibility'. *Philosophical Studies*.
- Strawson, Peter F. 1962: 'Freedom and Resentment'. *Proceedings of the British Academy*, 48, pp. 187–211.
- Vargas, Manuel 2005: 'The Trouble with Tracing'. *Midwest Studies in Philosophy*, 29, pp. 269–91.
- Vargas, Manuel 2013: *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press.
- Wallace, R. Jay 1994: *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard U. P.
- Zimmerman, Michael J. 1997: 'Moral Responsibility and Ignorance'. *Ethics*, 107, pp. 410–26.