

The Hard Problem of Responsibility

Victoria McGeer and Philip Pettit

INTRODUCTION

The background assumption in this paper is that holding agents responsible for their deeds is appropriate to the extent that they have the capacity to appreciate and act on the considerations that, by our more or less common lights, are relevant to what they ought to do or ought to have done.¹ We the folk hold an agent responsible insofar as we endorse reactive attitudes like resentment and gratitude in the case where the action affects us, indignation and approval in the case where it affects others (Strawson, P.F. 1962). And we endorse such attitudes insofar as we take two conditions to be fulfilled: first, the agent had the capacity to register the available options, with their respective pros and cons; and, second, the agent had the capacity to choose between the options on the basis of those reasons. To introduce a term of art, we take the agent to be responsive to the reasons relevant in their choice, where being responsive means having the capacity both to register and to act on those reasons.

This view takes reason-responsiveness to be sufficient for fitness to be held responsible, despite the fact that it need not involve a libertarian or contra-causal free will. And it also takes reason-responsiveness to be necessary for fitness to be held responsible. Thus it allows that some people may be exempt from responsibility for their actions insofar as they are temporarily or permanently out of their minds and deprived of responsiveness; they merit our taking what P. F. Strawson (1962) described as the objective as

¹ For some approaches on broadly these lines, see Fischer and Ravizza (1998); Pettit and Smith (1996); Scanlon (1998, 2008); Smith, Angela (2007, 2008); Smith, M. (2003); Vihvelin (2004); Wallace (1994); Watson (2004); Wolf (1987, 1990). For a useful, if critical overview see Clarke (2009) and for an overview of background work on dispositions see Cross (2012).

distinct from the participant stance. And it allows of course that people may be fully or partially excused insofar as they act out of ignorance, failing to register properly the options or considerations before them, or are subject to disabling pressures not to act appropriately. Exemption is due when reason-responsiveness fails to exist, excuse is due when the exercise of reason-responsiveness is hindered.²

Reason-responsiveness in this sense constitutes what we call a specific capacity to respond to the reasons relevant in a choice. If someone is subject to an exempting condition, then they lack the capacity to respond to reasons altogether. If they are not subject to an exempting condition, then they have the generic capacity to respond to the reasons, as we shall say. And if in addition they are not subject to an excusing factor of any kind, then they have the specific capacity—a capacity specific to the situation on hand—to respond to the relevant reasons. The specific capacity is just the generic capacity, operating in the absence of any excusing factors.

The standard reason-responsive view, which comes in different versions, offers an analysis of what it means for agents to possess, not just a generic capacity to respond to reasons, but the specific capacity to do so in a particular situation. In offering this analysis it solves the problem of identifying the feature that makes someone fit to be held responsible for what they do in those precise circumstances. But solving this problem—we call it the ‘soft’ problem of responsibility—only gives rise to a deeper problem that the standard view fails to address. This is the ‘hard’ problem of responsibility we discuss in this paper. In section 8.1, we give a statement of the hard problem as it arises on the standard reason-responsive picture. We sketch a revision to that picture in section 8.2. And in section 8.3 we show how the revised picture offers a way around the problem.

8.1 THE PROBLEM IN THE STANDARD PICTURE

According to the standard reason-responsive approach to responsibility, it is appropriate to blame someone—to hold them responsible in a negative sense—just when they have the specific capacity to respond to reasons in a given choice but still fail to exercise that capacity. In such a case we say or imply that the agent could have done otherwise—they had no excuse for failing to exercise or manifest that capacity—and the presence of the

² For further discussion of the distinction between exempting and excusing factors see Wallace (1994) and Gardner (2007). We note that the presence of excusing hindrances may not exonerate an agent if they are there because of a past failure, but we shall generally put aside the distinction between exonerating and non-exonerating excuses.

specific capacity is supposed to explain why this is so.³ Although the standard view also makes a claim about why praise is merited in the positive case where someone actually exercises the capacity to respond to reasons, we shall focus here on the negative scenario.

What does it mean to possess a specific reason-responsive capacity on the standard view? There are a number of analyses offered in existing reason-responsive approaches. While none of them on its own provides a solution to the hard problem of responsibility, as we shall see, the solution we develop for that problem presupposes some analysis. And so we begin by providing a sketch of the account that we like best.

According to our preferred account, someone has the specific capacity to respond to reasons in a given situation of choice, *S*, to the extent that in an open range of *S*-like situations where the relevant reasons are present and exempting and excusing conditions are absent, they generally register and act on those reasons.⁴ The agent responds to the reasons in most scenarios within that range; and it is that pattern of robust responsiveness that supports the ascription in the actual situation, *S*, of the specific capacity to respond as reasons require (Smith, M. 2003). What does it mean to say in the event of an actual failure to respond to the reasons, that the agent still had the specific capacity to respond, and could have done otherwise? It not only means, as some analyses have it, that the agent is such that in a suitably minimal variation on *S*, he or she would have registered and acted on the reasons. It means, more strongly, that the agent is such that he or she would have registered and acted on the reasons in most of the variations on *S* where the reasons remain present and exempting and excusing conditions absent.⁵

On this account, the specific capacity to respond to reasons in a given situation of choice is ~~of course~~ a disposition, in the presence of relevant

³ For purposes of this discussion, even the agent who is unknowingly blocked from taking ~~any course other than one~~—or even blocked from trying to take ~~any other course~~—can satisfy this clause (Frankfurt 1969). There will still be a difference between the agent who spontaneously acts in a certain way in light of the reasons and the agent who triggers the unrecognized blocker and so is forced to act in that way. Thus we can say of the agent who spontaneously acts in that way that they could have done otherwise: they could have not responded spontaneously in that manner. An alternative line on this issue is offered in Smith, M. (2003).

⁴ We call this our preferred account, while acknowledging that there is considerable variation in how the relevant range of *S*-like situations might be specified; see Vargas (2013: ch. 7) for a sensitive discussion of this issue.

⁵ There are different ways of modeling more or less robust or reliable responsiveness in a range of worlds: whether as full responsiveness in most worlds, likely responsiveness in all of the worlds, or a mix of the two. We do not mean to prejudge that issue, although we generally speak as if the first is appropriate. For discussion of a related question see Pettit (2015: App. II).

reasons, to register and act on those reasons.⁶ But, to signal a feature that is important in our solution to the hard problem, that disposition may take either of two forms, or may be a mix of both. It may consist in being such, here and now, as to register and act on whatever the reasons turn out to be. Or it may consist in being such, here and now, as to *adjust* in the presence of whatever the reasons turn out to be, ~~where that adjustment makes it possible~~ to act as the reasons require. In other words the disposition may be a once-for-all-inputs disposition, as when the agent is now fully ready to respond to the reasons; or it may be an input-by-input disposition, as when the agent is ready to become attuned to the reasons and, given the likelihood of attunement, ready to respond appropriately after its appearance.

An analogy may help make this difference clear. Consider what it means to have an *adaptive* immune system. Creatures with such a system are disposed to fight off infections but they may be disposed to do this in either of two different senses. They may have been exposed to a non-lethal form of the disease, in which case they will have developed the relevant antibodies and will be actively disposed—disposed in a once-for-all-inputs fashion—to fight it off: the antibodies will be there, waiting to get to work when suitably triggered. But even if they have not been exposed to the disease, and not developed this active immunity, the creatures will be disposed in a more remote, input-by-input sense to fight it off; by hypothesis, their adaptive immune system will start manufacturing antibodies in response to any form the disease may take. This passive immunity is not quite as effective as the active sort but it can provide a perfectly adequate level of protection. For all that our favoured account of the capacity to respond to reasons requires, it may assume the once-for-all-inputs form analogous to the first case, or the input-by-input form analogous to the second. Or of course it may consist in the first form for some inputs, the second for others.

Let us assume that someone is fit to be held responsible for a given choice if, in the sense explained, they have the specific capacity to respond to reasons in that choice. Suppose then that the person fails to respond to reasons and we hold them responsible, as the approach suggests we should do. Presumably the failure will be explicable in some fashion. But can the explanation allow us to blame the agent for that failure, as when we say or

⁶ Does the account avoid the so-called masking and mimicking problems associated with dispositions, on the broadly modal account given here (Cross 2012)? We think that it avoids mimicking problems insofar as it requires that the agent is categorically constituted so as to respond appropriately. And we think that it avoids masking problems insofar as it requires that the agent responds appropriately in an open range of possible scenarios; see Manley and Wasserman (2008).

imply ‘You could have done otherwise’?⁷ As we explore this question, we will be confronted with the hard problem of responsibility.

Some explanations of the failure would certainly allow us to blame the agent. We would blame the agent, for example, if we explained the failure by appealing to a trait like laziness or weakness of will. But these explanations are special. They allow us to condemn the failure that they explain only because we hold the agent responsible for the persistence of the trait in question; that trait is not, as we might put it, a brute factor. We have to think, in accordance with the reason-responsive approach, that the agent has the specific capacity to respond to reasons and overcome that trait. We must deny, for example, that the laziness or weakness of will is sourced in some pathology, or even some pattern in the past, that makes it impossible to overcome without serious therapy or biochemical intervention. If we thought that the trait was maintained in that way, we would treat it as an excusing factor.⁸

This observation shows that, even in a case where excuses are absent, we may explain the failure of an agent to respond to reasons by invoking a non-brute factor—i.e., a factor for whose persistence we hold the agent responsible, as with laziness and weakness of will. But how do we explain the agent’s independent failure to exercise the specific capacity to respond to reasons and resist or remove that trait? We might invoke a further non-brute factor to explain the failure to respond to the reasons *for* overcoming that trait. And while we might go on in the same vein, searching for ever deeper, non-brute explanations, it would make little sense to postulate an infinite regress of non-brute factors that account for the agent’s failure in the case at hand. The explanatory regress must end somewhere by invoking a brute factor.⁹

⁷ For a good discussion that identifies this general sort of problem, see Hieronymi (2007). The positive approach we adopt, however, is very different from hers; it is a version of the reason-responsive approach that she rejects.

⁸ Notice that on the approach taken here, an agent may have an excuse for failing to be reason-responsive on a given occasion without thereby being exonerated for her failure. While she drank to excess and is not now capable of responding to the reasons, e.g., we may nevertheless blame her for failing, at that earlier time, to respond to the reasons for remaining sober. Non-exonerating excuses invariably point to the presence of a specific capacity the agent failed to exercise at some previous time, whereas exonerating excuses do not indict the agent for any such prior failing.

⁹ By some accounts the fact that we have to reach a brute explanation at some level means that the explanation at the original level—the explanation that invokes laziness or weakness of will—is itself brute and does not allow us to hold the agent responsible for the action (e.g., Strawson, G. 1994). This regress argument would undermine the possibility of vindicating responsibility within a naturalistic, reason-responsive approach but for the moment we put it aside (we revisit it with our response in n. 22).

We return, then, to the question raised earlier. Does explaining the failure to respond to reasons by referencing a brute factor allow us to blame the agent for that failure? The reason-responsive approach is committed to thinking that it must allow this, if the agent had the specific capacity to respond to reasons at that level. And that commitment, as we shall see, is hard to maintain.

What might count as brute factors that could explain, at one or another level, why an agent fails to exercise a specific capacity to respond to reasons? One is that although there was a high probability that the agent would respond appropriately in the sort of situation in question, it was just sheer chance that this failed to happen. And another is that a glitch of some kind—say, a neural misfiring—obstructed the operation of the agent's capacity, leading to the failure of response. What happened under either of these hypotheses was untypical of the agent, as we might put it, and did not reflect their real nature (Smith, M. 2003: 36). It is true, then, that the agent could have done otherwise but only in the sense in which this means that their acting as they did—their acting in a way that failed to respond to the reasons—was a fluke. The failure did not mean that they lacked the capacity to respond to relevant reasons: not the generic capacity, and not even the specific capacity in the situation at hand, since by hypothesis there was no excuse present to hinder the exercise of the generic capacity.

But if this is the explanation to be given for why a reason-responsive agent did not actually respond to reasons in a given choice, why should we hold the person responsible for having made that choice? That is the hard problem of responsibility. Spelling it out in greater detail, why do we get emotively engaged in maintaining that the agent could have done otherwise, as resentment and disapproval suggest that we typically do? And, even more important, why do we blame or condemn the agent rather than taking a detached, if disappointed view of the failure? We say: 'You could have done otherwise; you could have responded to the relevant reasons in your action'. But why should this remark not count as a disengaged comment on possibility? And why should it not communicate consolation rather than condemnation, conveying the message: 'This was not characteristic of your performance; it was just an unhappy chance or an unknown glitch that led you to do what you did'?

On the analysis given, 'you could have done otherwise' is true if and only if you are such that you would have done otherwise—you would have registered and acted on the reasons—in most of the variations on the situation where those reasons retained their force and excuses and exemptions were absent. But the fact that you would have done otherwise in those variant scenarios does not explain why we are likely to respond to your actual failure by investing, often emotively, in censure. On the contrary it suggests

that we should respond by offering consolation. One response to this argument might be that no analysis can be expected to save every connotation of what it analyses: this is the so-called 'paradox of analysis'. But the analysis given here does not just fail to save the condemning connotation of the remark, 'you could have done otherwise'. It supports a contrary connotation of consolation.¹⁰

The hard problem is nicely underlined with a parallel from outside the realm of human agency. Suppose that we ascribe to your race horse the generic capacity to run a mile under a minute and a half: that is, to run a mile at that champion pace in the absence of wet conditions that might excuse a failure. And suppose that conditions being perfect, we ascribe to the horse the specific capacity to achieve this result. Let the horse fail to run the mile under a minute and a half, then, and we will have little option but to suppose that the failure was due to chance or to a glitch of some kind. Thus, we may say that the horse would have run at a champion pace in most variations on the actual scenario that preserved the relevant factors. But in saying this, we will be offering a sort of consolation, communicating the message that despite the failure you still have a champion on your hands. Why wouldn't something similar be true in the analogous case where, despite your failure to respond to reasons in the actual situation, we say you would have done so in most variations that preserved those reasons? Why wouldn't this communicate a similar message of consolation and reassurance regarding your capacity to be suitably reason-responsive?

The hard problem of responsibility, as this parallel emphasizes, has two aspects. One is to explain why we should typically feel an emotive, resentful engagement with the agent who fails in a brutally explicable way to manifest a specific capacity to respond to reasons. And the other is to explain, in particular, why this engagement should take the form of blaming or condemning the agent for the failure.

One response to the problem might be to say that when we blame the agent, we are seeking to impose a penalty in order to improve the agent's performance in the future (see, for instance: Schlick 1939: ch. 7; Smart 1961). Although we think there is a grain of truth in this response, it is inadequate insofar as it fails to honour the phenomenology of what we do in blaming. When we tell you that you could have done otherwise, making clear that we condemn what you actually did, we target and indict a failure on your part in that particular choice; we do not intervene simply out of an explicit intention to change your general habits. True, we might give up

¹⁰ We thank Michael Smith for raising this concern.

blaming one another for the bad things we do, if we came to believe that blame had no effect on future action. But that is not to say that blaming is a brutally deterrent intervention.¹¹

According to this inadequate response to the hard problem, ascribing capacity in the presence of failure, and so blaming the agent for failing to exercise the capacity, is an exercise in incentivizing the agent: providing a motive for the agent not to fail like that in the future. In this paper we argue that ascribing capacity in the presence of failure—and indeed more generally—is rather an exercise in capacitating the agent: supporting and reinforcing the capacity ascribed. We hope that the considerations to be outlined in the remainder of this paper will provide ground for accepting this view of what the ascription of a reason-responsive capacity involves, as well as providing a more satisfying justification for our reactively infused practice of holding one another responsible for what we do.¹²

In the next section, we begin our positive account by arguing that the standard picture of reason-responsiveness is importantly incomplete and in need of a critical revision. Building on our revision, we then address the hard problem of responsibility in the third section, showing how it can be legitimate to blame an agent for the failure to manifest a specific capacity to respond to the reasons relevant in a given choice.

8.2 FROM THE STANDARD TO THE REVISED PICTURE OF REASON-RESPONSIVENESS

8.2.1 Revising the Standard Picture

On the received story about the specific capacity to respond to reasons, it consists in a two-place relation. You are responsive to relevant reasons in a given choice insofar as there is a relationship between you on the one side and the reasons on the other. Excuses being absent, you tend to appreciate the reasons relevant to the choice and to act as they require.

We propose to expand this picture, beginning from an assumption that no one is likely to challenge. This is that your being disposed to appreciate and act on relevant considerations in various choices—your having the generic and often specific capacity to respond to relevant reasons—is not

¹¹ This point was originally made by P.F. Strawson (1962) and is extensively discussed in McGeer (2014, 2012).

¹² We have each gestured in previous writings at a general line of the sort developed here. Thus see McGeer (2008b, 2012, 2014) and Pettit (2001, 2007); Pettit and Smith (1996).

just the result of your own perception and appreciation of the factors that weigh for and against certain options. It is bound to be due in good part to your having heeded the judgment and advice of others, and benefitted from their encouragement. In developing a reason-responsive capacity, you learned from exchanges with others, in particular others who were ready to reason with you, to draw attention to purportedly relevant factors, to listen to your reactions to their observations, to look for a common point of view from which to reconcile differences, and so on.

This developmental assumption, as we may call it, need not bear on past history alone. Just as it is plausible that you depended in the past on interaction with advisors for achieving responsiveness to reason, so it is plausible that you continue to rely on such interactions for maintaining that capacity. It may be essential for your continued responsiveness to reasons that you pay attention to others, seeking their advice or gauging their reactions, in order to remain attuned to the demands of reason. If power corrupts, as age-old wisdom has it, that may be because too much power would make you inattentive or indifferent to others and deprive you of the check and balance that they can provide.

To introduce this developmental assumption, whether in its historical or continuing version, is not yet to make any revision to the standard reason-responsive picture. It is to postulate that individuals have a reason-responsive capacity in virtue of a past or continuing history of interaction with other reasoners, other advisors. But it is not to say anything new about the nature of the capacity in itself. This can be taken to consist, as in the standard view, in a standing sensitivity to the reasons themselves: a cognitive attunement that makes their demands salient and a motivational attunement that gives those demands an affective force.

At this point, however, we introduce an extra element to the reason-responsive picture. Given that you have depended, and continue to depend, on the guidance of others for your responsiveness to reasons, it is inevitable that you should regard those others in more than an instrumental light. You are bound to care about what they advise you to do in a given choice, about what you think they would advise were they in a position to do so, and about what they are later likely to think you should have done, assuming they come to learn of your decision. Or at least this is inevitable in the case of others whom you take to be well-informed and not mistaken in the judgment you ascribe to them. You would display a striking cognitive dissonance if you treated others as checks on whether you are being properly responsive to reasons, as the developmental assumption registers, and yet did not care about the judgments they pass, whether as advisors at the time of choice or as later examiners of the choice. Thus, in authorizing others as potential advisors, you are willy-nilly led to authorize them in yet

a deeper way: as an audience whose judgments of your performance, voiced or unvoiced, matter to you in their own right.

This implies that when you fail to act as the reasons require, you are bound to accept that you fail to act as others authorized by you would recommend that you act. You are bound to accept that you fail not just by your own lights; you fail by their lights too. The others with whom you share relevant standards will inevitably form expectations of you: that you will conform with the requirements of the relevant standards, as they see them, or at least that you will reason with them as to why the requirements are not as they think. Moreover, the existence of these expectations is bound to be a matter of shared awareness between you and them: each of you will be aware of this, each aware that each is aware, and so on (Lewis 1969). Hence, you must recognize that failing to abide by these standards also means not living up to the express or presumptive expectations of your audience, and in that sense disappointing them.

The developmental assumption makes clear why disappointing others in this way should also matter to you, and indeed matter inherently. Insofar as you rely on interaction with others for registering and acting on reasons, you show a concern for their judgment; hence, you are bound to care about how you stand in their estimation if you fail to accord with the relevant standards. You will show yourself either to have been insincere in endorsing certain shared standards as to what reasons are relevant, obtuse in understanding what those reasons require of you, or irresolute in acting on the reasons endorsed. You will put yourself in a position where, by ordinary standards, you are bound to feel shame—or at least embarrassment—at the putative flaws in you that your failure implies.

But, of course, there are also instrumental considerations why disappointing others should matter to you. If you fail to live up to shared standards, then you must look like a very bad bet as a partner with whom to enter into common projects, reciprocal contracts or the community of friendship. You must face the prospect of a gentle, or perhaps not so gentle, ostracism. And that alone ought to give you a concern for doing well in the eyes of the others around you—in particular, those whose general standards and judgments you share—that is distinct from the concern to do well as such.

It is plausible that we are all disposed by nature to care about the opinion that others form of us; evidence of not caring is likely to seem like an indicator of psychopathy or some other disorder. But even if nature did not dispose us to care about not being shamed before others—and to care more generally about enjoying the esteem of others—the observations made in this discussion ought to show that it would be irrational not to have a concern for how you are viewed by them. Indifference to others would sit uneasily with authorizing them as guides in your reasoning and it would

jeopardize the prospect of benefitting from their community and cooperation (Brennan and Pettit 2004).¹³

8.2.2 The Significance of the Revision

The extra factor introduced under our revised picture means that as a reason-responsive agent you can be expected, not only to be moved by the reasons you confront in making a choice, but also by the audience, actual or prospective, that you confront. This means that there is a variable, unrecognized in the standard picture, that may help to cue you to reasons in choosing between the options before you in any choice. This variable is not the standing force exercised by the reasons themselves, assuming that you are cognitively and motivationally attuned to those considerations. Rather it is the situational force exercised by the expectations to which you are held by others, given you naturally desire to live up to those expectations, commanding a secure position in the opinion that others hold of you.

In the standard picture, your responsiveness to reasons is wholly the product of your standing sensitivity to reasons: your disposition, nurtured in past and continuing interaction with others, to appreciate and act on those relevant reasons in any instance of choice where no potentially excusing factors get in the way. But on the revision we are now proposing, your responsiveness to reasons may be the product in part of quite a different sort of sensitivity as well. This we describe as your sensitivity to others, in the role of actual or prospective audience, rather than simply to reasons; in particular, your sensitivity to the opinion that they are likely to form about you, given how you think and act.

When an audience is actually present and aware of what you are doing, perhaps even deliberating with you about the act, it is quite clear why you should care, not just about doing well, but about doing well in their eyes. Equally when others are likely to discover and assess what you did after the fact, it is clear why you should care about doing well in their eyes, not just about doing well as such. Indeed, for the record, we think it is plausible, even when others are not aware or particularly likely to become aware of

¹³ Perhaps the most powerful statement is from Adam Smith (1982: 116): ‘Nature, when she formed man for society, endowed him with an original desire to please, and an original aversion to offend his brethren. She taught him to feel pleasure in their favourable, and pain in their unfavourable regard. She rendered their approbation most flattering and most agreeable to him for its own sake; and their disapprobation most mortifying and most offensive’.

how you choose, that you will still care about what they would think if they ever were to become aware.¹⁴

Let us assume, then, that in a given choice you are not only sensitive to reasons but also sensitive to audience: sensitive to the expectations that others hold of you, at least when they actually know what you do or are likely to learn about it (for simplicity, we put aside the case where they might but probably won't learn about it). The sensitivity to reasons constitutes a disposition, as we have seen, to appreciate the reasons and to act as the reasons require. We think of it as a standing disposition, no doubt formed over a long history of interaction with others, that you bring to a choice. But how does the sensitivity to audience figure in the picture? How does it act as a force that is independent from the sensitivity to reasons itself?

We suggest that it constitutes a disposition to become even more strongly disposed to appreciate and act on the reasons that are relevant in the choice. It is a higher-order disposition that works on a case-by-case basis, serving in the presence of an actual or prospective audience to reinforce the lower-order, standing disposition to be sensitive to reasons. Suppose you bring to a choice a sensitivity to reasons of a certain strength, *S*, where the strength of a disposition is determined by the probability it puts in place that under a relevant scenario or stimulus you will respond to reasons. The idea is that your sensitivity to audience in that choice may reinforce your sensitivity to reasons by making you more attentive, more careful, more motivated to track the reasons that there are, at least for the duration of the choice. It may increase the strength of that disposition so that your ultimate responsiveness to reasons is of strength, *S*-plus, not just *S*.

Why not think that the two sensitivities work in tandem, so that you have a dual concern: on the one side, to respond as reasons require; on the other, to respond in a way that does not disappoint others? The answer is, in brief, because the expectation on the part of others that you are required to meet if you are to win standing in their eyes is the expectation that you will be responsive to reasons (Brennan and Pettit 2004: ch. 1). Let it seem to others that you are responsive to them for the sake of being responsive to them, not that you are responsive to reasons, and you will not achieve the

¹⁴ This concern may be sourced in a certain kind of risk-aversion. It will often be within the power of others, should they wish, to make themselves aware of how you choose in any given case and you will be exposed to that power insofar as you do not act up to shared standards. However, the concern may also become simply a matter of cognitive-affective habit. For an argument as to why we are naturally concerned about possibilities, however unlikely, that it is in the power of others to realize, see Seligman et al (2013). For a recent research perspective in which such possible audiences would have a natural place see Pettit (2015).

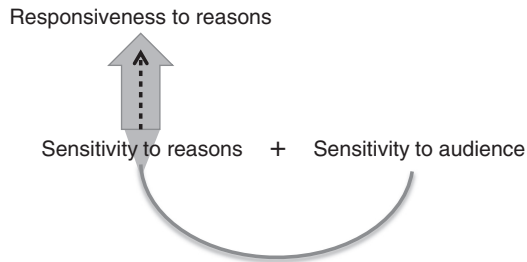


Figure 8.1 Determinants of reason-responsiveness.

standing in their eyes that is naturally appealing: ‘nothing is so unimpressive as behaviour designed to impress’ (Elster 1983: 66). And so, inevitably, you must let your sensitivity to others express itself in a heightened sensitivity to the reasons that you and they both take to be relevant. Or at least you must do so in your dealings with others who, by your lights, are worthy of being authorized as advisors.¹⁵

The upshot of this picture is that your responsiveness to reasons in a given choice, your specific capacity to act as they require, is a function of the standing sensitivity to reasons that you bring to that choice and the situational strengthening of that sensitivity under the impact of your sensitivity to an authorized audience. We might picture the play of these factors on the lines of Figure 8.1.

In this picture, responsiveness to reasons is the resultant of the two sensitivities, where one of those sensitivities normally operates via the other sensitivity.

If your responsiveness to reasons in a given choice is a function of two forces, ~~of course~~, then it becomes possible ~~that~~ your responsiveness ~~may~~ result from different combinations of those forces. The two sensitivities may combine in different measures to produce responsiveness and any degree of responsiveness may be realized via any of a range of equivalent combinations. In some combinations sensitivity to reasons will be high and

¹⁵ There may be people who are so powerful and intimidating, of course, that despite not being worthy of authorization—despite not being sensitive to the reasons you take to be relevant—they play an inhibiting role in your life. At the limit, they may even lead you to pretend to be sensitive to the considerations that move them and to pretend to go along with them as you might go along with an authorized audience. We assume in the present context that this will be the exception, not the rule. We concentrate on the significance for your responsiveness to reasons of a sensitivity, not to others in general, but to others whom you are disposed to authorize.

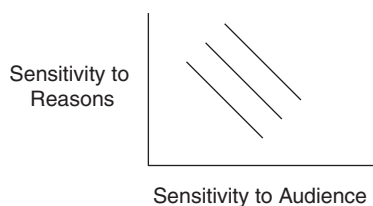


Figure 8.2 Equivalence curves in reason-responsiveness.

sensitivity to audience low, in others the reverse will hold, and in still others the factors will be more or less equal.

Thus we might represent your responsiveness to reasons in a given choice or type of choice on a graph in which the vertical axis depicts sensitivity to reasons, the horizontal sensitivity to audience; see Figure 8.2. Take a given combination as a point on the graph and now connect up those points that yield the same degree of responsiveness. Such points will connect in an equivalence curve in the space of responsiveness. We may characterize any of a range of curves, each corresponding to a different degree of responsiveness: the curves to the upper right will represent higher degrees of responsiveness, those to the lower left lower degrees, in analogy with the familiar picture of indifference curves in the space of utility. One of those curves will presumably characterize your degree of responsiveness to reasons in the choice in question. And on the curve that identifies you, there will be a point that represents your particular combination of sensitivities in the relevant choice.¹⁶

This analysis of the two components that may figure in your responsiveness to reasons has interesting implications for what it means to say that you have the capacity to respond to reasons. We mentioned earlier that, on the account adopted here, you are able to respond to reasons in a situation, *S*, to the extent that you are such as to respond to them in most *S*-like situations where the reasons remain present and excuses and exemptions absent. But what we must now notice is that when I take you to be responsive, it may be that I do not credit you with a very reliable, standing capacity to respond to reasons. I may take you to be suitably responsive—to have the required capacity—only in the actual or foreseen presence of the audience that I and perhaps others constitute. It may even

¹⁶ We draw the curves as straight parallel lines, although there is good reason to think that the marginal sensitivity to audience of the virtuous—i.e., those who score high on the vertical axis—is going to be less than the marginal sensitivity to audience of those who are unvirtuous—i.e., those who score low on that axis.

be that when I take you to be an audience worth authorizing—an audience capable of registering relevant reasons—I do so on the assumption that you are yourself performing up to scratch only because of the audience that I and others provide for you. Authorization and sensitivity may be reciprocal and mutually reinforcing.

We mentioned earlier that the specific capacity to respond to relevant reasons in a situation, *S*, may assume a once-for-all-inputs or an input-by-input form: it may involve a readiness, just as you are in yourself, to register and act on the reasons; or it may involve a readiness to adjust in the presence of those reasons and, after adjustment, to register and act on them. To the extent that your responsiveness to reasons is a function of your sensitivity to audience, it will assume the input-by-input rather than the once-for-all-inputs form. You will depend on being attuned to the reasons present in the situation by your sensitivity to audience and you will be ready to register and act on the reasons only when that attunement has materialized. Your capacity to respond to the reasons, then, will be fixed in place, not just by how you are in yourself, but by the audience-exposed environment in which you operate; it will have an ecological character.¹⁷

8.2.3 The Traditional Sense in the Revision

However unfamiliar our two-factor account of reason-responsiveness may initially seem, it is worth emphasizing that our observations are broadly in line with received wisdom. We can recast the observations in more familiar terms by considering an effect we all take for granted in assuming that there is good point in holding someone to account for what they do. In making that assumption in relation to you, we assume that your accountability to certain reasons is a function of two factors: one, a disposition to prove accountable that is independent of our actually holding you to account; and two, your disposition to prove accountable as a result of our holding you to account. While our revision of the standard picture of reason-responsiveness introduces new language to describe the two factors at work in making you responsible or accountable, the distinction between those two factors is one that we are all disposed to acknowledge anyhow.

But doesn't the picture we describe jar with a familiar Aristotelian image? On that image, excellence or *aretai*a—in effect, a standing sensitivity to reasons—is the terminus that the good person may be expected to achieve

¹⁷ This notion of an ecological capacity, as we later discovered, resembles a notion defended by Manuel Vargas (2013); we borrow the term 'ecological' from his book.

and mere competence or *enkrasia*—a capacity to respond to reasons that is not fully based in sensitivity to reasons—is a second-best. Are we committed to rejecting that image in favour of a picture in which even the best of people are likely to be dependent on a sensitivity to others for the capacity to respond properly to reasons? No, we are not required to reject this image. For all that our view supposes, it is quite possible for someone to become so sensitive to reasons that, while they remain sensitive to audience, while it continues to matter to them that they act well in the eyes of the others they authorize, this concern does not play an essential role in making them responsive to reasons in any range of choice; their standing sensitivity to reasons—their excellence or *aretai* in that regard—is sufficient on its own to keep them on the path of virtue.

Does this then suggest that a sensitivity to others may play no actual role in the psychology of the virtuous: that while it may strengthen their disposition to respond to reasons, that strengthening is entirely redundant? Again we say, no. Think of the old political insight, already cited, that all power corrupts and that the only hope of keeping the powerful virtuous is eternal vigilance: that is, a sustained interrogation and examination of their performance (Pettit 1997: ch. 7). It may be that while you now have Aristotelian virtue, your political power is such that, if we ceased to be vigilant, if we enabled you to serve your own interests with invisibility and impunity, that would change your character and destroy your virtue. It may be, in other words, that your standing capacity to respond to reasons, however autonomous in its operation, is dependent for its survival on your sensitivity to others.

8.3 THE SOLUTION IN THE REVISED PICTURE

Our revision of the reason-responsive picture introduces a new element into the story of how we as agents develop and maintain our reason-responsiveness: we can become more responsive to reasons by being sensitized to the judgments of an authorized audience, present or foreseen. But the revision also introduces a new element into the story of how we are likely to perform, not as agents, but as audience. Given that the sensitization charted in the last section is salient to all, the revision means that when I interact with you ~~as audience with agent~~, then I am bound to recognize, implicitly or explicitly, that I play a sensitizing role and to credit you with a degree of responsiveness to reasons that presupposes my efficacy in playing that role.

This observation is the key to our way of handling the hard problem of responsibility. When I credit you with a specific capacity to respond to

reasons—or when I presuppose that capacity in holding that the reasons require you to do such and such—I do not speak as a detached observer who is liable to think of any failure of responsiveness on your part as just a fluke. I speak as someone invested in your proving to be responsive and as someone who thinks I can help to make you responsive. And that, so we shall argue, makes for a big break with the standard version of the reason-responsive approach.

The standard version of the approach, as we saw, fails in two salient ways. First, it does not make a place for the resentment with which I typically burn when I find that you have hurt me, not taking due account of my interests, or for the indignation I typically harbour when I learn of how you have jeopardized the happiness of another, neglecting considerations of common decency. And, second, it does not make sense of why I hold you up for censure or blame in the event of such a failure. It rightly registers that I credit you with a responsiveness to reasons when I cite the reasons why you should have acted otherwise; or, taking those reasons as understood, when I say that that you could have acted as they require. But it does nothing to suggest that my ascription of this responsive capacity should amount to anything more than an observation, perhaps inflected with some disappointment, that it was possible for you to have acted in accord with those reasons. It treats that ascription as a modal report akin to the report I might make on the racehorse in our earlier example: that notwithstanding the horse's failure to run at a champion pace under perfect conditions, still it was possible for it to have done so.

The two failures of the standard reason-responsive approach may be put as follows: first, that it reduces the ascription of a specific reason-responsive capacity to a modal *report*; and, second, that it reduces it to a *modal* report. By representing the ascription as a *report*, the account fails to register the typically emotive character of communicating that you could have done otherwise. And by representing the ascription as purely *modal*, it fails to mark the difference between saying, as a matter of impersonal possibility, that you might have responded to the reasons and saying this in a way that justifies condemnation and blame. These two failures correspond, of course, to the two aspects of the hard problem of responsibility with which we began. That problem, as we saw, is: one, to explain our emotive, resentful engagement with the agent who fails to manifest a specific capacity to respond to reasons; and two, to explain why this engagement should take the form of blaming or condemning the agent for the failure. We now proceed to show how the revised version of the reason-responsive approach can handle these two challenges.

8.3.1 Ascribing a Capacity to Respond to Reasons Need Not be Merely *Reportive*

The first challenge for our account is to show why the ascription of a specific reason-responsive capacity is typically not merely reportive in character. Consider the case where you are facing a choice and I cite or gesture at various considerations relevant to how you make that choice. In doing this I convey a certain view of you: that you are responsive to those reasons and have the capacity to appreciate and act on what they require. For instance, suppose you are asked by a friend about your role in some episode of double-dealing and you have to choose between telling the friend a lie about your unflattering involvement and fronting up with the full truth about the part you played. I am aware of your position and I tell you that there are good reasons to be truthful, that they require you to be truthful, and by implication that you can be truthful: you have the specific capacity, unaffected by excusing factors, to respond to the reasons and tell the truth.

The standard story of reason-responsiveness does not require us to think that in doing this, I am doing anything more than recording the fact of your responsiveness to what the relevant, recognized reasons require and perhaps expressing my belief that you are likely to exercise it: this, on the ground that the responsiveness consists in a disposition to appreciate and act as the reasons require. But on the revised account, I also recognize that your responsiveness to reasons is partly a function of my presence and my presumptive view that the reasons require truth-telling. And on that story, I ascribe such responsiveness, at least in part, with a view to eliciting or reinforcing that very capacity: ultimately, with a view to getting you to tell the truth. Consciously or unconsciously, I seek to sensitize you to reasons in the very act of recording your sensitivity.

According to the revised account, no matter how sensitive I take you to be to the reasons relevant in the situation, I assume that you are also sensitive to me, as to an authorized audience: you are disposed in light of my manifest expectations as to how you will perform to become even more sensitive to the requirements of the relevant reasons. And in accordance with that story, I act on this assumption, adopting an interventionist role in nudging and guiding you to conform to what the reasons require and to tell the truth. In saying or implying 'You can tell the truth', then, I do more than record the responsiveness assigned; I do more than remind you of the robust possibility that you will tell the truth. I speak with a view to helping evoke or reinforce the exercise of that very capacity. I make the remark in conscious or unconscious pursuit of this desired effect, trying to move you to act appropriately.

Since the work of J. L. Austin (1962), we have been familiar with the notion of a performative utterance: a statement such that making it in suitable circumstances ensures that its truth conditions are fulfilled. Standard examples are ‘I order you to do such and such’ or ‘I baptize this ship so and so’. On one account of such performatives (Lewis 1983: ch. 12), to make an utterance of this sort is certainly to report that a certain state of affairs actually obtains—namely, the state of affairs (the order, the baptism) characterized by the utterance. But it is also to do more than offer a report, since the utterance actually realizes the state of affairs reported. It is a performative report, as we say, not just one of a purely descriptive character.

On the story about ascribing capacity that we are developing now, ‘You can tell the truth’ ~~does not aspire to be~~ just a descriptive report; it is not meant to communicate merely the possibility that you will tell the truth. And neither of course is it a report with performative significance: it doesn’t make it the case that it is possible that you will tell the truth. But it is what we may describe as an *evocative* report: a report on the possibility of your telling the truth, the purpose of which is, at least in part, to evoke or call into being the very possibility reported—namely, your telling the truth.

The evocative character of saying or implying that you can tell the truth in this example explains why that ascription is not a mere report and why, in making it, I may be emotively engaged in the exercise. The ascription is not primarily designed to communicate that it is possible for you to tell the truth, according to my beliefs. Why would I even bother to convey such a message? Under general circumstances, the aim in making it is to bolster and reinforce the very possibility it reports, persuading you to exercise the specific capacity ascribed. I proffer considerations that support telling the truth, relying on my ability in doing so to sensitize you to them, and not just relying on your standing sensitivity to reasons. I speak then with the aim of exhorting you to display the capacity and, being invested in your doing so, I speak in an engaged or emotive mode.

We have been arguing that my saying ‘You can tell the truth’ in the example given aims at evoking performance and that this explains why it does not count as a mere report: why it typically has an engaged character. If that claim is accepted, then it also helps to explain the engaged character of my observation, should you tell a lie, that you could have told the truth: that you could have done other than what you did. If the remark ‘You can tell the truth’ was appropriate prior to action, expressing an aspiration to help evoke the exercise of the capacity, then the remark ‘You could have told the truth’ will serve in the wake of action to reaffirm that aspiration or perhaps a more general form of the aspiration. In making the later remark, I certainly report the possibility that you might have told the truth. But in the context described, I do more besides. I express a view of you as someone

who was, and continues to be, susceptible and responsive to evocatively intended ascriptions of sensitivity: I don't give up on you, resigning myself to your lack of performance. And so it is unsurprising that I should be emotively invested in what I say when I declare that you could have done otherwise; I say this, not in the spirit of a detached observer, but in the spirit of someone exhorting you to do better (McGeer 2012; Pettit 2007; see too, Macnamara 2013).

These claims about second-person ascriptions of responsiveness to reasons, whether made before or after action, pair off with corresponding claims about first-person ascriptions. When I say to myself, 'I can tell the truth', I am not just reporting on a possibility I detect in myself, even a robust possibility. Playing audience to myself, I typically make the remark by way of exhorting myself to manifest the possibility or capacity ascribed.¹⁸ This marks the fact that I am capable of *sensitizing* myself to reasons so far as I care about living up to the standards that I take myself to endorse. Further, what holds of the post-hoc second-person utterance, 'You could have done otherwise' holds in parallel of the post-hoc first-person utterance, 'I could have done otherwise'. Suppose that despite having exhorted myself by the present-tense ascription 'I can tell the truth', I actually fail to tell my friend the truth. And suppose that I do not give up on myself as someone I can continue to exhort in the same manner, evoking performance in the very act of ascribing a capacity to perform appropriately. In that case the remark 'I could have told the truth' will have the same evocative force as its second-person counterpart. It will record a continuing possibility in the spirit of someone invested in having a certain effect on performance, not in the spirit of an impersonal reporter.

We began with the observation, problematic for the standard picture, that crediting an agent with responsiveness to reasons, whether to reasons cited or taken as understood, does not have the character of a detached report. But we have now seen that the picture in which responsiveness to reasons is a function of both a sensitivity to reasons and a sensitivity to audience allows us to make sense of the engaged, emotive character of ascribing capacity. The ascription of capacity in the second- and first-person can itself induce or reinforce that capacity: it can *capacitate* the addressee. And the natural way to take such ascriptions is as evocative reports in which speakers are engaged in supporting the performance sought.

What of the apparently disengaged case where ascriptions are used in the third-person, as when we say of some distant other that they could have

¹⁸ The view articulated here fits well with a more general account of self-knowledge that emphasizes the commissive aspects of self-reports. For a defence and elaboration of this view, see McGeer (1996, 2008a); Moran (2001).

done otherwise? We may surely express resentment or indignation in this case too. And such emotive engagement can make sense on our account. It is fully intelligible that while speaking of agents in the third person, we may still mean to speak evocatively. Assuming that we have a sense of the sorts of agents in whom we can elicit or reinforce a capacity to act on the reasons, we may speak of their failures as from the point of view of someone who might have made a difference. Indeed we may even signal our commitment to engaging in such evocative work, were the opportunity to arise.

8.3.2 Ascribing a Capacity to Respond to Reasons Need Not be Merely *Modal*

We have seen that the standard version of the reason-responsive approach does not in itself explain why ascribing responsiveness to reasons should be engaged and emotive. We now turn to a second, even more important failure that the story exhibits. It fails to explain why the responsiveness that I ascribe or assume in dealing with you involves more than the *bare* modal possibility that you can act or could have acted as relevant reasons require. It does not save the phenomenology of ordinary interaction, in which the ascription of responsiveness is associated with condemnation of a failure to be responsive and indeed, though we shall not be speaking of this case, with commendation for success in doing so.

Condemnation or censure in the event of failure does not just consist in giving you a poor rating, even a poor rating that is engaged and emotive. It does not encode just the judgment, for example, that you would have been a better person had you been such as to be responsive to reasons in the case on hand. In that respect it differs from the non-censorious, albeit evaluative and perhaps emotive judgment, that you would have been a better person if, for example, you had received a proper education or been required to face greater difficulties in your youth. What censure communicates is the judgment that it was within your power to have exercised your responsiveness in the case where you fail: say, the case where you fail to tell your friend the truth about your part in the double-dealing example. You had a choice, so the idea goes, between telling the truth and telling an untruth; you had the capacity to tell the truth, being responsive to the considerations supporting that option; and you blew it: notwithstanding the fact that you could have done otherwise, you chose to tell a lie.

The standard story, as we saw, fails to explain the place of such censure. According to that story, being responsive to reasons means that, excuses being absent, it was possible as a matter of modal fact, indeed robustly possible, that you should have chosen as the reasons dictate: that is, it was

possible that you should have told the truth. The suggestion has to be, then, that the robust possibility of acting as the reasons dictate allows us to condemn you for not realizing that possibility in the actual world. Although it was robustly possible for you to tell the truth—although you are such that under many possible, even slight variations on that situation you would have told the truth—you did not actually do so. And so, according to this suggestion, we condemn you in order to mark the fact that you acted out of character; some unknown factor—perhaps sheer chance, perhaps a neural glitch—got in the way of your characteristic responsiveness to reasons. That you could have done otherwise means only that it was a fluke you did not do otherwise: it would have been much less surprising had you manifested responsiveness and spoken truly.

This account clearly fails to make sense of why we condemn your failure to respond to reasons in a case where there is a brute explanation for the failure; it fails to resolve the hard problem of responsibility, under its second aspect. To imply that it was just a fluke that you did not do otherwise communicates consolation—you are not so bad, after all—rather than condemnation; the unknown factor that explains your failure—the chance or the glitch—looks like an excuse that can let you off the hook. But as the revised version of the reason-responsive approach solves the hard problem under the first aspect, so we think that it can solve the problem under the second aspect. It can make sense of why my saying that you could have done otherwise in the wake of a failure to respond to reasons need not have a merely modal character.

Let us return once again to the perhaps special case where, prior to choice, I ascribe or assume a capacity in you to be responsive to the acknowledged reasons and to tell the truth to your friend. In explicitly or implicitly ascribing that capacity, I act in a consciously evocative manner, seeking to strengthen your standing sensitivity to the reasons I put before you or take to be before you. And I do so without leaving any uncertainty in place about the responsiveness I think I can evoke, as I might do if I said or suggested ‘You can probably do this’ rather than ‘You *can* do this’. I invest myself categorically in your proving responsive. Relying on your standing sensitivity to those reasons being enough for my sensitization to make it fully effective, I speak with a view to evoking full responsiveness to reasons. Assuming the role of sensitizer, I say in effect: ‘Go for it; act on those reasons!’

Insofar as I give my remark this evocative, injunctive significance, I do not use it to communicate a straightforward modal fact: *viz.*, that you are such that, in a range of possible scenarios where the relevant reasons remain in force, you more or less invariably tell the truth. I use it rather to encourage and enjoin you to pay attention to the reasons, confident that if you do so then that will be enough to get you to tell the truth. The message

conveyed is essentially evocative: ‘Pay attention to what I am supposing or pointing out—viz., that there are reasons to support telling the truth, and that you can and ought to do as the reasons dictate—and you will succeed in telling the truth!’¹⁹

If this is the message communicated in saying ‘You can tell the truth’—or more generally, ‘You can appreciate and act on the reasons’—what is the message communicated by saying in the case of failure: ‘You could have done otherwise?’ The message, as we have seen, cannot be that it was just a fluke that you did not appreciate and act on the reasons and did not tell the truth. That would not support condemnation or blame. But what then is it that I mean to communicate?

The message has to be consistent with my thinking, even after you failed, that what I said earlier in claiming ‘You can tell the truth’ remains true. If I thought that I was wrong to have made that earlier claim, then I would have to say now at the later time that it was not the case that you could have done otherwise. But I do say that you could have done otherwise and that means that what I said earlier remains true by my later lights. Thus, at the later time I have to think that, when you made your choice, paying attention to the considerations mobilized would have been sufficient to ensure success. And so it follows that you did not pay attention, thereby flouting my injunction. If I condemn you, as by hypothesis I do, then the obvious explanation is that I condemn you for not having heeded my implicit injunction: the instruction I conveyed to pay attention to the reasons.

Giving an instruction, as in saying ‘Do X’, always amounts to communicating something of the form: ‘Do X, or else’, where the else is a presumptively unwelcome consequence. That is what gives humour to the old (American) joke that when an unarmed UK police officer shouts ‘Stop’, she means ‘Stop, or I’ll shout “Stop” again’. In telling you to do X, whatever X amounts to, I always have to assume and communicate that, excuses being absent, your failure to act as instructed will put you on the hook, triggering an unwelcome consequence. Thus there will be no question of letting you off the hook on the grounds that some unknown, brute factor must have gotten in the way.

Why should the presumptive presence of a brute factor that explains your failure to do X not argue for letting you off the hook? The answer has to do with the logic of injunctions: there would never be any point in enjoining someone to do something if there were always some reason, *ex post facto*, to let them off the hook. My enjoining you to do X is only sensible if it is likely

¹⁹ In the words of Ian Hacking, I am not representing or reporting on how things are with you; I am intervening in how things are with you (Hacking 1983).

to get you to do X and it will not be likely to get you to do X if I am going to excuse any failure for which there is a presumptive explanation. Thus I must be prepared to maintain that the injunction was suitable, and that you are on the hook, in many cases where the failure is explicable.

This observation about injunctions or instructions enables us to explain the condemnatory force of my saying to you in the event of telling a lie to your friend that you could have done otherwise. In saying this, I am delivering the unwelcome consequence—in this case, an attitude or expression of censure—that was held out as a prospect in my earlier injunction to pay attention to the reasons. Now, on the occasion of your failure, I identify still with that injunction by rebuking or censuring you for not having done what I enjoined. The evocative character of the ascription of capacity implicit in my treating you at the moment of choice as responsive to reasons offers a straightforward explanation of why in the event of a failure to exercise that capacity I censure you. I censure or condemn you for not having followed the instruction that any such evocative ascription conveys.

Under this revised story we can still ask about what caused your failure. And it will remain the case, by the argument in the first section, that for many failures we can only invoke an unknown, brute factor in the role of explanation. But under ordinary criteria the presence of a factor that explains failure in this way need not undermine the appropriateness of my having enjoined you to pay attention to the reasons and avoid that failure. And I myself insist that that the injunction was appropriate insofar as I say in the wake of the failure that you could have done otherwise, thereby censuring you in a way that the injunction licenses. Had I come to think that the injunction was inappropriate—that the vitiating factor had the recognized status of an excuse—then presumably I would not reaffirm your specific capacity to have done otherwise and would let you off the hook.

This observation raises the question as to what leads us in any domain of injunction to treat some factors that explain failure as excuses and not to treat others in that way. Given the line taken here, we are directed to an independently appealing theory of excuses. On that theory, excuses are just those failure-explaining factors of which the following is true: according to assumptions encoded in our injunctive practice—these may vary, of course, across cultures—there is little hope of neutralizing their effect by holding people responsible in their presence.²⁰ And so, on that theory,

²⁰ As we noted in n. 8, non-exonerating excuses point to factors that implicate some past non-excused failure of reason-responsiveness on the agent's part. The assumptions encoded in our injunctive practice would presumably cast these as factors of a kind that can be neutralized by holding an agent responsible, not for the present action, but for the failures in the past that gave rise to them.

the features that explain failure without counting as excuses are just those factors—those glitches and chances—that are susceptible, according to our injunctive assumptions, to the regulatory effects of our holding one another responsible. There is more to be said on this topic but we cannot explore it further here.

The line we have been developing may explain why my remarking ‘You could have told the truth’ in the event of failure carries censure with it, when it is preceded at the time of choice by the corresponding remark, ‘You can tell the truth’. But what of the more general case where I was not aware of your choice in advance and was not in a position to ascribe a capacity in an evocative way? What of the case where I become aware of what you did and comment that you could have done otherwise? How can this ex post ascription of capacity have the condemnatory force that we attributed to it in the case just discussed?

We believe that it can carry that condemnatory force just insofar as I assume that you, like any agent in our community, are subject to the general injunction to pay attention to the relevant reasons and I personally identify with that injunction, as I might identify with an earlier injunction of my own. Assuming the absence of excuses, I do not need to have addressed you explicitly in advance. I can take you to have been fully aware of the general injunction to pay attention to the reasons in such a case, as supported by the community at large. And identifying with the injunction to which I take you to have been subject, I blame you for not having paid attention, insisting that you could have done otherwise. I impose on you the penalty that promised, by shared assumption, to attend any failure to abide by the injunction.

What we have just argued in the second person case holds also in the first person. When you explicitly or implicitly think to yourself ‘I can respond to the reasons and tell the truth’, you ascribe that capacity in a sensitizing sense and you endorse the self-injunction ‘Pay attention to the reasons and you will act as they require’. And so you rebuke and reprimand yourself when in the wake of failure you admit ‘I could have told the truth’. You blame yourself for not having paid attention properly to those reasons—for not having obeyed that self-injunction—and, in consequence, for not having actually told the truth.

Does the interpretation just given to the claim that you could have done otherwise extend beyond the first and second person to uses of that sort of sentence in contexts where we speak of someone in the third person? Not necessarily, since in making that utterance we may simply wish to mark the fact that other paths were not blocked and that it was possible that the agent should have done otherwise. But we may preserve the charged tone, of course, or an echo of that tone, if we wish to indicate condemnation. In

such a case we identify with the general injunction to which we take the person to have been manifestly subject, and we communicate that we blame them for not having heeded it.²¹

CONCLUSION

When I ascribe a specific capacity to you to respond to relevant reasons in a certain choice, or when I assume such a capacity by telling you how the reasons require you to act, then on the standard picture with which we started I do nothing more than indicate that things are a certain way with you. I communicate that, absent excuses, you are the sort of person who would register and then act on those reasons in a range of contextually salient scenarios that resemble the actual situation. But what happens in any case where you do not respond appropriately to the relevant reasons? The standard picture must ultimately invoke an unknown, brute factor—a chance or a glitch—to explain the failure. And this appears to let you off the hook in the manner of an excuse; it suggests consolation rather than condemnation, registering that your failure was entirely out of character.

The hard problem of responsibility is to show why, within the terms of any naturalistic approach, an emotively engaged, normative condemnation remains appropriate in the presence of such a brute explanation. On the standard approach this problem persists. And the persistence of the problem may well explain the temptation to think that such responsibility—the kind of responsibility compatible with blame—presupposes a non-naturalistic, contra-causal conception of free will. This line would deny that there is any naturalistic explanation available—any explanation in terms of chance or glitch—for why a culpable agent failed to act appropriately. It would hold that the only explanation can be a failure on the agent's part to have exercised their free will appropriately: that no other sort of failure would merit resentful or indignant condemnation.

The revision of the reason-responsive approach adopted here enables us to resist this temptation to renounce naturalism. On the line taken, when I ascribe or assume a capacity in you to respond to relevant reasons in a certain choice then I certainly presuppose that you satisfy the conditions

²¹ We assume that such emphasis and colouring will equally explain why we can embed ascriptions of capacity in the antecedents of conditionals, without treating them as straightforward modal reports. We say that if someone could have done otherwise (signalling that we take the agent to have breached a salient injunction with which we identify), then they ought to be censured. In effect, we say that if someone breached such an injunction, then they are subject to blame.

identified in the standard picture. But the revision of the standard picture allows us to see that I do something else as well. I invest emotively in your exercising the capacity to respond to reasons. And in explicitly or implicitly ascribing that capacity I enjoin you to pay attention to the reasons relevant to the choice, blaming you if you fail to obey it. More generally, if I take you to have had that capacity in a choice where I was not present, and if I identify with a presumptive injunction requiring you to respond to the relevant reasons, then equally I blame you for having breached that injunction.²²


On the general viewpoint from which we approach the problem addressed in this paper, the practice of holding one another responsible is of immense importance in human life, providing for a sort of mutual scaffolding or capacitation and enabling us to lift our performance to a level we might not otherwise have attained (McGeer 2008b; Pettit 2007). The hard problem addressed is that this practice, however useful it may seem, appears to falter in the presence of failures that, under naturalistic assumptions, we explain by reference to brute factors. Our aim has been to show how it is possible to retain a naturalistic approach—specifically, the reason-responsive version of such an approach—and yet make sense of why the practice of holding one another responsible can apply in the presence of such failures. In particular we hope to have shown that it is precisely the scaffolding or capacitating aspect of that practice that makes sense of how it can survive the hard problem of responsibility that those failures raise.²³

²² The revised, reason-responsive approach not only gets us over the hard problem of responsibility. It is worth noting that in doing this, it also blocks a familiar regress argument against the possibility of naturalistic responsibility that we mentioned in n. 9. This holds that if at any level there is a brute explanation for the presence of a trait like laziness or weakness of will that itself explains a failure to respond to reasons in a given action, we cannot blame the agent for the failure explained. While agreeing that at some level there must be a brute explanation for the presence of such a trait, we can resist the conclusion drawn. Since we can blame an agent for a failure that has a brute explanation, by the argument of the paper, we must surely be able to blame the agent for a failure that is explained by a trait like laziness or weakness of will for whose presence the agent is responsible. And equally we must be able to blame the agent for the presence of that trait itself, even when there is a brute explanation for its presence. We can do so insofar as the agent has or had the capacity to respond to reasons and overcome the laziness or weakness of will, at least in light of the fact that we and others hold them to account for doing so and thereby help to evoke the capacity within them. Whatever the level at which we take an agent to have the specific capacity to respond to reasons, at that level we can hold the agent responsible for any failure to exercise that capacity.

²³ We are very grateful for the many useful comments we received when an earlier version of this paper was presented at the NOWAR conference in New Orleans in November 2013 and at a conference at the University of Melbourne in March 2014. We are also indebted to comments from the editor, David Shoemaker, from two anonymous referees, and from Michael Smith.

References

- Austin, J. L. (1962). *How to Do Things with Words* (Oxford: Clarendon Press).
- Brennan, G. and Pettit, P. (2004). *The Economy of Esteem: An Essay on Civil and Political Society* (Oxford: Oxford University Press).
- Clarke, R. (2009). "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism." *Mind* 118(470): 323–51.
- Cross, T. (2012). "Recent Work on Dispositions." *Analysis* 72(1): 115–24.
- Elster, J. (1983). *Sour Grapes* (Cambridge: Cambridge University Press).
- Fischer, J. M. and Ravizza, M. (1998). *Responsibility And Control. A Theory of Moral Responsibility* (Cambridge: Cambridge University Press).
- Frankfurt, H. G. (1969). "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66(23): 829–39.
- Gardner, J. (2007). *Offences and Defences: Selected Essays in the Philosophy of Criminal Law* (Oxford: Oxford University Press).
- Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (Cambridge: Cambridge University Press).
- Hieronymi, P. (2007). "Rational Capacity as a Condition on Blame." *Philosophical Books* 48(2): 109–23.
- Lewis, D. (1969). *Convention* (Cambridge, MA: Harvard University Press).
- Lewis, D. (1983). *Philosophical Papers*, vol. 1 (Oxford: Oxford University Press).
- Macnamara, C. (2013). "'Screw You!', & 'Thank You'." *Philosophical Studies* 165(3): 893–914.
- Manley, D., and Wasserman, R. (2008). "On Linking Dispositions and Conditionals." *Mind* 117(465): 59–84.
- McGeer, V. (1996). "Is 'Self-knowledge' an Empirical Problem? Renegotiating the Space of Philosophical Explanation." *Journal of Philosophy* 93: 483–515.
- McGeer, V. (2008a). "The Moral Development of First-Person Authority." *European Journal of Philosophy* 16(1): 81–108.
- McGeer, V. (2008b). "Trust, Hope and Empowerment." *Australasian Journal of Philosophy* 86(2): 237–254.
- McGeer, V. (2012). "Co-reactive Attitudes and the Making of Moral Community." In *Emotions, Imagination and Moral Reasoning*, ed. R. Langdon and C. Mackenzie, 299–326 (New York, NY: Psychology Press).
- McGeer, V. (2014). "P. F. Strawson's Consequentialism." In *Oxford Studies in Agency and Responsibility*, vol. 2, ed. D. Shoemaker and N. Tognazzini, 64–92 (Oxford: Oxford University Press).
- Moran, R. (2001). *Authority and Estrangement: An Essay On Self-Knowledge* (Princeton, NJ: Princeton University Press).
- Pettit, P. (1997). *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press).
- Pettit, P. (2001). "The Capacity to Have Done Otherwise: An Agent-Centred View." In *Relating to Responsibility: Essays for Tony Honoré on his Eightieth Birthday*, ed. P. Cane and J. Gardner, 21–35 (Oxford: Hart Publishing).

- Pettit, P. (2007). "Joining the Dots." In *Common Minds: Themes from the Philosophy of Philip Pettit*, ed. G. Brennan, R. E. Goodin, F. C. Jackson, and M. Smith, 215–344 (Oxford: Oxford University Press).
- Pettit, P. (2015). *What We Do For Each Other: Ethics with Attachment, Virtue and Respect* (Oxford: Oxford University Press).
- Pettit, P., and Smith, M. (1996). "Freedom in Belief and Desire." *Journal of Philosophy* 93: 429–49.
- Scanlon, T. M. (1998). *What We Owe To Each Other* (Cambridge, MA: Harvard University Press).
- Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, MA: Belknap Press of Harvard University Press).
-  Schlick, M. (1939). *The Problem of Ethics* (New York: Prentice-Hall).
- Seligman, M. E. P., Railton, P., Baumeister, R. F., and Sripada, C. (2013). "Navigating Into the Future or Driven by the Past." *Perspectives on Psychological Science* 8(2): 119–41.
- Smart, J. J. C. (1961). "Free-Will, Praise and Blame." *Mind* 70(279): 291–306.
- Smith, A. (1982). *The Theory of the Moral Sentiments* (Indianapolis, IN: Liberty Classics).
- Smith, A. (2007). "On Being Responsible and Holding Responsible." *The Journal of Ethics* 11(4): 465–84.
- Smith, A. (2008). "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138(3): 367–92.
- Smith, M. (2003). "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion." In *Weakness of Will and Practical Irrationality*, ed. S. Stroud and C. Tappolet, 17–38 (Oxford: Clarendon Press).
- Strawson, G. (1994). "The Impossibility of Moral Responsibility." *Philosophical Studies* 75(1): 5–24.
- Strawson, P. F. (1962). "Freedom and Resentment." *Proceedings of the British Academy* 48: 187–211.
- Vargas, M. (2013). *Building Better Beings: A Theory of Moral Responsibility* (Oxford: Oxford University Press).
- Vihvelin, K. (2004). "Free Will Demystified: A Dispositional Account." *Philosophical Topics* 32(1/2): 427–50.
- Wallace, R. J. (1994). *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press).
- Watson, G. (2004). *Agency and Answerability: Selected Essays* (Oxford: Clarendon Press).
- Wolf, S. (1987). "Sanity and the Metaphysics of Responsibility." In *Responsibility, Character and the Emotions*, ed. F. Schoeman, 64–2 (Cambridge: Cambridge University Press).
- Wolf, S. (1990). *Freedom Within Reason* (New York: Oxford University Press).