“Agency” is a term of art. Its meaning and use might be discovered by reading and studying this volume. In contrast, “responsibility” is an everyday term, one that regularly appears in kitchen-table conversations and on the pages of newspapers. The present chapter is meant to relate them.

RESPONSIBILITY

Let us start by examining the relatively wide variety of everyday uses of “responsibility” or “responsible.”

In one, the word could be replaced, without loss of meaning, by “cause:” “The extreme heat was responsible for the engine failure” is equivalent to “the extreme heat caused the engine failure.”

In another, the word could be replaced, instead, by a word such as “duty,” “obligation,” or the more prosaic “job:” “Feeding the dog is my responsibility” means “Feeding the dog is my job” (or duty, or obligation). (In contrast, causing the failure was not the job of the extreme heat.)

Relatedly, the thought expressed by, e.g., “Maria is very responsible” could instead be expressed with, “Maria is conscientious; she takes her obligations seriously.” Someone who is irresponsible cannot be trusted with a job.

In a third and more complicated sort of use, “responsible” could be replaced with “at fault for” or “to be thanked for,” as in “Who is responsible for the delay?” or “Who is responsible for this lovely reception?” Here, it seems we mean something like, “Whose actions or omissions foreseeably caused this (bad or good) outcome?”

As is often the case with ordinary notions, these uses overlap and combine. Sometimes, in saying someone is responsible for something, we mean that they have an obligation, or duty, or job, because they are at fault for some bad outcome—as in (certain uses of), “Who is responsible for this mess?” and “The mess is John’s responsibility.” John made the mess, and now it is John’s to clean
up. (We do not use the word in this way in cases in which someone is to be thanked or rewarded, presumably because, in such case, it is we, not they, who have the obligation—of gratitude or of recompense.)

Sometimes, though, rather than impose upon John a duty or obligation of clean up, we impose upon John some fine, penalty, or punishment. In such cases, it seems to be because John’s actions or omissions foreseeably caused some bad outcome that some burden can now be aptly imposed upon John—he can be fined for the mess because he is responsible for it.

If we are in the business of imposing a penalty or burden on John because he is at fault for the mess, we might justify our doing so by saying, “he is a responsible adult.” In this case, we are not saying that John is conscientious or trustworthy (like Maria). Rather, we are saying that he is such that he can be aptly fined, penalized, or punished for his errors. He is such that he can be, as we say, held responsible. (In fact, it is sometimes suggested that to fail to burden John for his malpractice would be to fail to treat him as a responsible adult—to treat him with disrespect.¹)

Holding responsible is something we do to one another. One way in which we do it is by making clear that an action or omission creates some obligation—by making clear our expectations of clean up, e.g. Another, as just considered, is by imposing penalties, sanctions, or punishments.

However, there is an additional, more complex or subtle, way in which we hold one another responsible, one which has been the subject of voluminous philosophical reflection over the last six decades: We stand ready to respond to others with what Peter Strawson called “reactive attitudes.”²

Strawson introduced the term “reactive attitudes” in his landmark paper, “Freedom and Resentment.” These are attitudes or emotions, like resentment, gratitude, and indignation, which we form in response to what Strawson called the “quality of will” that we perceive in another. They contrast with an “objective attitude,” which we adopt in response to things that do not manifest a

---

¹ See, e.g., (Morris 1968).

² (Strawson 1962)
will of any quality. To illustrate: while you might be frustrated, even angry, if you find that a stray nail has punctured your tire and left you stranded, you will feel differently if you find that someone has slashed your tire, on purpose. You will resent the slashing. (If you resent the nail in the same way, you may recognize this is a mistake, a kind of animating of the physical world.) Likewise, if a snowstorm forces a closure that absolves you of a distressing and burdensome task, you will feel relief. But if instead someone notices your distress and takes action to relieve you of the burden, you will feel gratitude. (You would feel grateful for the snowstorm if you thought it orchestrated by a divine being with you in mind.) Resentment and gratitude are reactive attitudes—they are, as Strawson puts it, “natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions.” When contrasted with these, anger and relief are more objective attitudes.

When we respond to the quality of another person’s will with reactive attitudes, we are, in an important sense, holding that person responsible, treating them as a responsible adult. If instead we fail to react, or react with more objective attitudes, we are, in an important sense, exempting that person from responsibility.

Strawson noticed three kinds of cases of exemption, three kinds of cases in which, when someone shows us malice or disregard, we respond in a less reactive, more objective, way. First, we might do so if we learn the person was in extreme or unusual circumstances. We might then simply ignore the ill will, treating it, so to speak, as the kind of thing that happens when people are under extreme stress. Second, we do not react in the usual way when interacting with people who are in some way incapacitated for ordinary interpersonal relationships. Strawson’s examples are immaturity and pathology. Finally, Strawson noted that we sometimes adopt an objective attitude for more pragmatic reasons: for therapeutic purposes, or out of scientific curiosity, or to avoid what he called “the strains of involvement.” The last should be familiar: Sometimes, in our ordinary relationships, we encounter an especially difficult or trying person. We sometimes then opt out, so to speak, of
ordinary interpersonal engagement: We treat that person as an "issue" or "challenge," someone to be simply handled or managed.

In recent philosophical discussion, responding to someone with negative reactive attitudes has been called "blaming," and being the appropriate target of such attitudes has been called being "blameworthy." These labels are unfortunate: the natural use of the English words "blame" and "blameworthy" do not pick out what Strawson had in mind, but rather what I called above being "culpable" or "at fault"—foreseeably causing some bad outcome for which you now rightly incur either obligations or vulnerability to some sanction or punishment.3

Perhaps because of the unfortunate label, philosophers have typically grouped reactive attitudes with sanctions and punishments, as though they were a burdensome treatment imposed upon the wrongdoer by the one reacting.4 This is a mistake. It overlooks a crucial difference in the agency at work.

It is tempting, even natural, to group the reactive attitudes with sanctions and penalties in part because they are (unfortunately, to my mind) often used as sanctions or penalties. Frequently enough, people wield their reactive attitudes, through their voluntary actions and communications, aiming to burden the wrongdoer—intending or hoping to make them bad for what they have done. Such "guilt-tripping" is not only possible but prevalent. Yet it does not show that the reactive attitudes, themselves, should be understood as sanctions or penalties: they should not.

To see this, note, first, that a sanction, penalty, or punishment is not simply a negative consequence that predictably follows from some action: a hangover is not a sanction. Neither is the pile of work, accumulated thorough procrastination, that now must be done by tomorrow.

---

3 Consider a lecture course going badly. The students will blame the instructor, thinking the lectures are boring, too difficult, or too flatly delivered. The instructor will blame the students, thinking they come to class inadequately prepared, they do not pay enough attention, or they are insufficiently curious. Each is thinking the other at fault for a bad outcome, and each is thinking it is the other's job to fix the situation. Quality of will, of the sort Strawson had in mind, is not centrally at issue; yet this seems a paradigm case of blame.

4 See, e.g., (Watson 1996) and (Wallace 1996).
Sanctions, penalties, and punishments are, instead, negative consequences that have been created and attached, by somebody (or some body), to certain violations. In response to such violations, sanctions, penalties, and punishments are imposed through voluntary action.

In contrast, reactive attitudes are not voluntary. Though, of course, one might communicate or express them via a voluntary action, the attitudes are not, themselves, adopted at will. The reactive attitudes are, as Strawson noted, “natural human reactions” to our perception of the quality of another’s will. Being the target of resentment is thus, in one way, more like having a hangover than being fined or sent to your room: it is not a burden voluntarily imposed in response to your misbehavior, but rather a natural consequence of your disregard, manifested in another.

However, while being the target of someone’s resentment is more like having a hangover than being fined, it is also importantly unlike having a hangover. These natural human reactions, though not voluntary actions, are also not simply involuntary reactions, like flinching or blinking. To treat the reactive attitudes simply like a hangover or a flinch would be to misunderstand the agency of the one who reacts—it would be to put the reactor in bad faith. We will return to this.

To summarize: The word “responsible” or “responsibility” is used in a variety of ways. Sometimes it simply notes causal relations. Sometimes it indicates obligations or duties, or conscientiousness with respect to one’s obligations or duties. Sometimes it indicates an obligation or duty that was created by some action or omission, as when one makes a mess and now must clean it up. Sometimes, to be “a responsible adult” is not just to be liable for cleaning up one’s messes, but also to be open to penalties, sanctions, and punishments for one’s actions or omissions. And, finally, we are, in an important sense, holding a person responsible when we respond to the quality of their will with what Strawson called the reactive attitudes. To be responsible, in this last sense, is to be

---

5 The ideas, and some of the text, of the last several paragraphs appear in both (Hieronymi 2019, in progress)
such that the quality of your will matters to others in the usual way. For lack of a better label, I will call it “responsibility as mattering.”

AGENCY

Turning, now, to the term of art: “agency.” We can already see a variety of different ways in which an investigation of human agency might interact with questions about responsibility and a variety of different things one might have in mind if one wanted to focus, narrowly, on “responsible agency.”

Starting with the first use of “responsible:” An investigation into human agency might have something to say about which of the indefinitely many events causally downstream of my action count either as part of my action or else as salient causal consequences of my action, such that I am responsible for them in the causal sense. To take a well worn example, if I am moving my arm in order to operate the pump, in order to replenish the water supply, in order to poison the inhabitants of the house, in hopes of hastening the end of the war, then it seems that, in moving my arm, I am therein pumping and therein both replenishing and poisoning the water supply. If the inhabitants are poisoned, I will have poisoned them. However, poisoning the inhabitants seems merely a consequence—an intended consequence—of my action of poisoning the water supply. I.e., it does not seem that moving my arm just is poisoning the inhabitants, in the way it just is pumping the water and poisoning the supply (even less that it just is hastening of the end of the war). An investigation into the metaphysics of actions and omissions may help us to draw lines between action, intended consequences, hoped for outcomes, and other downstream causal consequences. (Of course, it is possible that conclusions about responsibility will instead help to determine the boundaries of human action.)

---

6 Scanlon has recently introduced the term “moral reaction responsibility.” (Scanlon 2015) He explicitly means the term to parallel burdens or obligations. Because I do not think reactive attitudes should be thought of in parallel with burdens and obligations, I have chosen another term.

7 The example belongs to (Anscombe 1957).
Moving to the second idea, that of being “at fault,” we might focus investigations of agency on the ability to foreseeably cause outcomes one should understand to be good or bad. This will require some ability both to foresee and to evaluate that which one will cause. Certain sorts of outcomes (such as the lovely reception) will, in addition, require an ability to form and execute complex plans.\(^8\) Such agency goes beyond what might be thought of as the most basic sort, e.g., that displayed in simple organisms pursuing biological needs.\(^9\)

We saw, next, that sometimes those at fault for a mess incur an obligation to clean it up, and sometimes they are (perhaps in addition) liable to sanction, penalty, or punishment. An investigation of human agency might yield insights relevant to the ethical question of whether a person at fault for a bad outcome can aptly or justly be burdened with an obligation, penalty, or punishment.

To illustrate, the mess may be the child’s fault—they foreseeably caused it—and yet the resulting mess may be more than the child can be asked to clean: they lack the required capacities. In this case, facts about the child’s capacities as an agent serve as input into ethical reasoning about where obligations lie. For a second illustration, it seems unfair to sanction or penalize a person for a violation if they lacked an adequate opportunity to avoid it. Thus, again, investigation into our capacities for agency—for guiding ourselves in such a way as to avoid violations—might serve as input into ethical reasoning about when sanctions, penalties, or punishments are fair.

One version of the traditional “problem of free will and moral responsibility” can be seen as a particularly fraught result of this last sort of investigation: On a natural line of thought, a moral sanction or punishment is fairly imposed or truly deserved only if the wrongdoer “could have done otherwise.” However, it seems to some that investigation into our agency reveals that we never have

\(^8\) For planning agency, see (Bratman 1987).

\(^9\) See, e.g., (Burge 2009). One might identify an even more basic form of agency in, e.g., chemical agents.
the ability to have done otherwise. Such thinkers conclude that we are never rightly deserving of moral sanction or punishment, and thus never morally responsible.  

While such skeptics believe that facts about human agency undermine moral responsibility, others believe that features of morality will guard against the skeptical conclusion. Many of these non-skeptics agree with the skeptic that facts about agency and facts about moral requirement must fit together so as to ensure we are capable of avoiding moral wrongdoing—in a slogan, they agree that “ought implies can.” However, these non-skeptics then part company with the skeptic in at least two different ways. Some reason that, since we evidently are morally responsible, we must have the relevant ability (using “ought implies can,” they draw conclusions about agency from facts about moral requirement). Others believe that morality will stretch or shrink, so to speak, to accommodate the facts about agency, whatever they turn out to be.

While I am sympathetic to the thought that morality can accommodate the facts, whatever they turn out to be, I doubt that agency and moral requirement must fit together in the way suggested by “ought implies can.” Humans are born into the world in need of moral development and, tragically, such development often goes awry. People thus often arrive at adulthood selfish, petty, insensitive, ruthless, etc.—people arrive at adulthood without the capacities required to satisfy moral demands. Yet those demands do not, for that reason, yield or shrink to fit. It seems common that moral demands rightly apply to people who lack the capacities required for their satisfaction. We are, we might say, subject to original sin.

---

10 See, e.g., (Pereboom 2001) and (Strawson 1994).

11 See, e.g., (Chisholm 1964) (Kant 1997).

12 The strategy is pursued in very different ways by (Wolf 1980) (Wallace 1996) and (Strawson 1962) I provide a sympathetic reading of Strawson’s strategy in (Hieronymi forthcoming).

13 I am appropriating the idea of original sin. It is, properly, the Christian doctrine according to which each person is born into the world tarnished with the original sin of Adam and Eve, therefore unable to achieve righteousness on their own, and thus in need of God's grace. I am appropriating it as a culturally available opposition to the popular idea that “ought implies can” (an idea that may also find origins in Christian doctrine).
If we are, as I suspect, subject to original sin—if we are not always able to satisfy moral demands—this should inform our understanding of sanction and punishment. If, as the skeptic maintains, our ideas about sanction and punishment somehow presume that we can always avoid moral wrongdoing, then we may have to reinterpret them, to avoid that false presumption.\(^{14}\) It seems to me that this could be done—the fact of original sin need not show that sanctions or penalties are always unfair.\(^{15}\)

To my mind, the final sort of responsibility, “responsibility as mattering,” provides the most interesting material for investigations of agency. Reflection upon it prompts us to ask, not, what kind of agency is required to incur obligations or be fairly penalized?, but rather, to what kind of agency do we aptly respond with reactive attitudes? And, what kind of agency (if any) is involved in those attitudes, themselves? And, finally, why do we not respond with these attitudes to children or those subject to certain kinds of pathology or those in extreme or unusual circumstances—is this because of a limitation in their agency, or does the explanation lie elsewhere?

Here is my own answer to the first and second of these questions, and a hunch about the third:

The reactive attitudes are reactions to, as Strawson put it, the quality of another’s will. That “will,” as I would understand it, is comprised of those states of mind that manifest or embody one’s take on the world—one’s take on what is true, worthwhile, to be done, threatening, delightful, awe-inspiring, etc. A person’s will prominently includes their beliefs and intentions, but also includes their trusts and distrusts, admirations and contempts, cares and concerns, etc.\(^{16}\) The reactive attitudes react to one particular quality of a will: the way in which other people figure into a person’s take on the world—whether, e.g., the interests of others are worth heeding, whether their needs are

\(^{14}\) On the interpretation of punishment, see (Nietzsche, Clark, and Swensen 1998, second treatise, aphorisms 12–14, Clark and Dudrick 2012). I am particularly indebted to conversation with Mark C. Johnson. See (Johnson 2018)

\(^{15}\) For a justification of criminal punishment compatible with original sin, see (Scanlon 1998, Chapter 6).

\(^{16}\) Cf. Harry Frankfurt, “if we consider a person’s will is that by which he moves himself, then what he cares about is far more germane to the character of his will than the decisions or choices he makes.” (Frankfurt 1988, 84)
worth protecting, whether they are treated as of equal importance, etc. Thus, when one person is held responsible by another person, in the responsibility-as-mattering sense, one set of attitudes, one person’s take on the world and their place in it, generates a reaction in a second set of attitudes (which often prompts a further reactive change in the first, etc.).

We next ask about the agency we enjoy with respect to these reactive attitudes—and the answer is not straightforward. On the one hand, we are not simply passive with respect to them. Resenting or admiring is not like having a headache, seeing spots, or catching a cold. In manifesting or embodying our take on things, these attitudes are, in some way, up to us. On the other hand, this range of attitudes is not voluntary—we cannot simply adopt them at will, in the way we can (if able-bodied and well resourced) raise our right hand, dance a jig, or cook a meal at will.

Importantly, non-voluntariness is essential to these attitudes: any state of mind that manifests or embodies your take on what is true, important, wonderful, admirable, offensive, contemptible, etc., cannot be voluntary.\(^\text{17}\) If an activity is voluntary, if it is done at will, it reveals your take on what is all-things-considered worth doing.\(^\text{18}\) But, for this reason, nothing voluntary could manifest your take on the distinct questions of what is true, important, wonderful, admirable, offensive, contemptible, etc.\(^\text{19}\)

Notice, too, that these attitudes are aspects of ourselves for which we can be asked to provide a justification, rather than merely an explanation. We can be asked to defend them with reasons that

\(^{17}\) One might put the point in a way that sounds paradoxical: Non-voluntariness is an essential feature of the will. I think this is true. While ordinary action is voluntary, willing (intending) is not. See, again, (Frankfurt 1988) and (Hieronymi 2006, in progress).

\(^{18}\) You may, of course, do something that you do not believe to be all-things-considered worth doing. But what you believe to be worth doing and what you take to be worth doing need not march in step. If you do something intentionally, then you have, therein, taken it to be, all-things-considered, worth doing, even if you believe otherwise. See (Hieronymi 2009).

\(^{19}\) One might think one could, e.g., be offended at will. Note, though, there are reasons that you might take to show being offended is worth doing—e.g., it would improve your bargaining position—that you do not take to show that anything was offensive. If being offended could be done at will, one could adopt it for such reasons. But a state of mind that could be adopted for such reasons would not be one that manifests or embodies offense. (None of this is to deny the possibility of motivated offense—anymore than denying that belief is voluntary rules out wishful thinking or self-deception. See (Hieronymi 2006).)
show the content of our beliefs to be true, the object of our cares to be important, that which we prefer to be desirable, etc. They thus contrast with our headaches, the spots we see, or the colds we catch. We can explain these, but not justify them—we do not adopt or maintain a headache or a cold for reasons of our own.  

I have elsewhere suggested that we understand the agency we enjoy with respect to these attitudes as the agency we enjoy when we answer questions for ourselves—that our take on the world and our place in it can be well understood as our answers to questions about whether this is true, or whether that was offensive or malicious, or whether this is important, or all things considered worth doing, etc. 

Notice, we have just brought into view another kind of status, one that might be a form of responsibility thus far overlooked: the status of being answerable, in the sense of being subject to a request for one’s own reasons.

Being answerable is, indeed, an important status, and one might call it a form of responsibility (though it is not, I think, a natural use of the everyday word). Importantly, though, answerability is more fundamental than, and should not be confused with, responsibility as mattering. A creature, even a person, can be answerable without being responsible in the responsibility-as-mattering way—as shown by Strawson’s categories of exemption. Not every instance of the relevant form of agency carries the relevant sort of significance. It seems to me likely that a wide range of creatures, actual and merely possible, are capable of answering questions, in the relevant sense—they plausibly have a take on the world of a sort for which they have their own reasons and so they are, at least in

———

20 G. E. M. Anscombe brought to our attention the fact that, if you are intentionally Φ-ing, a certain why-question is, as she put it, “given application:” you can be rightly asked “Why you are Φ-ing?” The same is true of entire range of attitudes that manifest or embody your take on the world. (Anscombe 1957)


22 The nearby “accountable,” in the sense of “obliged to provide an account,” may provide a bridge.
principle, answerable. But not all of these will be capable of giving offense or of aptly prompting resentment, indignation, trust, gratitude, etc.

I suspect that what takes us from answerability to responsibility-as-mattering is not any special enhancement of agency, *per se*, but rather our ability to stand in complex forms of interpersonal relationships. As adults, we typically are able to recognize each other’s expectations and demands, to form expectations of our own, and to stand in relations in which we not only grant legitimacy to each others’ expectations but also know, collectively, that we each do so. In other words, I suspect that what takes us from answerability to responsibility is not further facts about our agency, *per se* (not further facts about our ability to do things, affect things, or control things) but rather important facts about our sociability—about our capacity to take one another into account, to matter to one another in distinctive ways.23

These last thoughts, though, are idiosyncratic. In the earlier ones, I hope to have provided a useful taxonomy of, first, the variety of things that people might have in mind when thinking about responsibility, and, second, the corresponding ways in which an investigation of human agency may either illuminate or be illuminated by such thought.24

**Recommended Reading**


—The above two anthologies compile recent work on “blame” and moral responsibility.


—This short paper challenges the idea that responsible human agency is ultimately grounded in our capacity for choice or decision.

23 (Hieronymi 2014, in progress).

24 Thanks to Mark C. Johnson for helpful comments and conversation.

—A standard anthology of work on free will and moral responsibility.


—This volume includes a short, opinionated primer on the problem of free will and moral responsibility, a reprinting of P. F. Strawson’s seminal “Freedom and Resentment,” and a close, line-by-line reading of that difficult article.
REFERENCES


Hieronymi, Pamela. in progress. *Minds that Matter*.


