

# Practical Responsibility

Katarzyna Paprzycka

## 1. Responsibility-Based Theories of Action

The recognition of the close relation between the concept of action and the concept of responsibility goes at least as far back as Aristotle. His account of voluntary action could be seen as being the source of two general strategies for understanding the concept of action.<sup>1</sup> One such approach is to determine when something is not an action first by studying a variety of interfering conditions.

- (e) The agent's  $\phi$ ing was a mere happening (non-action) iff external forces caused him to  $\phi$ .

But Aristotle described those cases as ones where the principle of action is not in the agent,<sup>2</sup> generating what might be thought of as a corresponding picture of what it means for a performance to be an action:

- (i) The agent's  $\phi$ ing was an action iff internal forces caused him to  $\phi$ .

In fact, however, (i) and (e) are independent of one another. From (e) it only follows that something is an action iff it is not caused by external forces. It says nothing about causation by internal forces. Indeed, from this point of view, the idea of internal forces causing performances is most naturally seen as a hypostatization of the absence of such causation by external forces.

---

<sup>1</sup> "What comes about by force or because of ignorance seems to be involuntary. What is forced has an external origin, the sort of origin in which the agent or victim contributes nothing — if, e.g. a wind or human beings who control him were to carry him off" (Aristotle, *Nicomachean Ethics*, trans. Terence Irwin [Indianapolis: Hackett, 1985], 1110a1-4).

<sup>2</sup> One must remember to avoid simple-minded interpretations here. The distinction is not (as suggested by the form of words Aristotle sometimes uses) between forces outside and inside the agent, for there can be the "wrong" kind of forces inside (e.g. spasms) and the "right" kind of forces outside the agent (e.g. persuasion). See Harry G. Frankfurt, "The Problem of Action," in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), pp. 69-79.

The latter approach has structured much of contemporary thinking about action. The former approach has been instantiated in some contemporary theories of agency. H.L.A. Hart's theory of action is one prominent example. Hart argued that action claims are properly thought of simply as ascriptions of responsibility. To say that John performed the action of dropping a glass is to ascribe responsibility to John for the event of the glass dropping (in the right conditions). One of the constraints on such a responsibility-based account is the unavailability of the concept of agency to assessments of responsibility. In fact, the approach might appear to be based on a conceptual mix-up. After all, in most instances, we base our judgments of responsibility on our judgments about actions. To claim that John is responsible for breaking the glass we must know that it is John who broke it, that he *did* it. This conceptual order seems to be also reflected in the very way we use the concept of responsibility: we are paradigmatically responsible for our actions. The fundamental problem of responsibility-based accounts of action is this: How can a person's action be understood in terms of whether it is appropriate for her to be held responsible if whether or not it is appropriate for her to be held responsible depends on whether or not she has acted?

One way out is inherent in the very idea of a responsibility-based approach to action. Such theories cannot rely on the concept of legal or moral responsibility lest they be satisfied with an account of legal or moral action. Rather they have to appeal to a broader concept of *practical* responsibility, which would be applicable even in cases that are legally or morally neutral. If so, then the problem is not as fundamental as it seems at first sight. There is *prima facie* nothing incoherent in thinking that our judgments about moral or legal responsibility are in part based on our judgments about actions and that our judgments about actions are based on our judgments about practical responsibility. The only problem then is to elucidate the concept of practical responsibility. This is the task of the present paper. I will develop an account of practical responsibility in terms of reasonable (in a sense to be specified) normative expectations (section 3). In the remainder of the paper, I will show that the account meets the criteria of adequacy set by Fischer and Ravizza's recent work.

## 2. Criteria of Adequacy for the Concept of Practical Responsibility

One might think that it would be rather hard to find criteria of adequacy for a concept of practical responsibility since the concept may appear to be a concoction tailored to serve the dubious (since unusual) responsibility-based theories of action. This impression is mistaken, however. Fischer and Ravizza's recent theory of moral responsibility implicitly points to a concept of practical responsibility. Their theory can therefore guide us in setting the criteria of adequacy for a concept of practical responsibility.

Despite the fact that Fischer and Ravizza aim to capture the notion of moral responsibility much of their work is directed at capturing the sense in which the agent can be said to be responsive to certain demands. They focus on the agent's responsiveness to reasons. But they are quite clear that there is nothing specifically moral about either the reasons to which the agent is said to be responsive or the concept they aspire to develop. Their concept of moral responsibility covers "morally neutral" actions: "one can be morally responsible for simply raising one's hand (where this is not a signal or in any way morally significant)."<sup>3</sup> They explain that on their (Strawsonian) view,

moral responsibility need not imply the actual application of a reactive attitude; it only requires that the agent be an *apt candidate* for such an application.<sup>4</sup>

It might be more appropriate to say, however, that they aspire to develop not quite a concept of moral responsibility but rather a broader concept of practical responsibility, where the latter is a necessary (but not sufficient) ingredient of the former.<sup>5</sup>

---

<sup>3</sup> John Martin Fischer and Mark Ravizza, *Responsibility and Control. A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998), p. 8.

<sup>4</sup> *Ibid.*, p. 8, emphasis added.

<sup>5</sup> A similar division into the two kinds of responsibility: moral responsibility and its prerequisite, practical responsibility, is strongly suggested by R. Jay Wallace's account (*Responsibility and the Moral Sentiments* [Cambridge, MA: Harvard University Press, 1994]). He argues that a person is morally responsible for *X* as long as it is fair to hold the person morally responsible for *X*. The concept of holding someone morally responsible in turn is articulated by appealing to moral normative expectations, i.e. to normative expectations that have specifically moral justification. This account is suggestive of a broader concept where the normative expectations to which it would be fair to hold a person are not restricted to

I shall therefore adopt the criteria of adequacy Fischer and Ravizza set for their account. The concept of practical responsibility must be able (1) to account for Frankfurt-style cases, and to correctly sort out our intuitions concerning ascriptions of responsibility for (2) omissions and (3) consequences. (4) In addition, as suggested already, insofar as the concept of practical responsibility is to ground theories of action, it must not appeal to the concept of agency.

I should point out that I am making no claims that the criteria of adequacy listed are exhaustive, but only that they are necessary conditions for a conception of practical responsibility.

### 3. Practical Responsibility

The concept of practical responsibility is not as rich as the concept of moral responsibility in at least two ways. First, the concept of practical responsibility must not appeal to specifically moral concerns. Second, it must be independent of the concept of action. One natural way of thinking about it is in terms of whether the agent is suitably responsive to be held practically responsible. Another way of putting it is in terms close to those of R.J. Wallace and Patricia Greenspan – whether our expectations of the agent would be reasonable. Indeed, I will argue that an agent  $\alpha$  is practically responsible for  $\phi$ ing (at  $t$ ) just in case it would be reasonable<sub>A</sub> (at  $t$ ) to expect of  $\alpha$  that  $\alpha \phi$  (at  $t$ ).<sup>6</sup> After a quick reminder of the distinction between normative and predictive expectations, I will explain the central notion of reasonableness<sub>A</sub>.

---

ones that have a specifically moral justification. Indeed, this is the skeleton of the concept of practical responsibility developed below.

<sup>6</sup> In all these contexts ‘ $\phi$ ’ occurs intensionally. An agent may be practically responsible for painting his house but not for offending his neighbor’s aesthetic sense, even though the two may describe the same event. From the account of practical responsibility there is a short step to the account of agency: Performance  $e$  is an action iff for some description ‘ $\phi$ ’ such that  $e$  prima facie fulfills the expectation of  $\alpha$  that  $\alpha \phi$ , it would have been reasonable<sub>A</sub> (at the time at which  $e$  occurs) to expect of  $\alpha$  that  $\alpha \phi$ . I argue for the adequacy of the account of action in (*Social Anatomy of Action: Toward a Responsibility-Based Account of Agency*, Ph.D. Dissertation: University of Pittsburgh, 1997).

## A. Normative vs. Descriptive Expectations

The distinction between predictive and normative expectations is familiar. I expect of my mother-in-law that she treat me with respect and yet I expect that she will not. That no contradiction is involved is clear. Two different concepts of expectation are involved: the former expectation is normative, the latter predictive or descriptive.<sup>7</sup>

There are various ways of drawing the distinction between normative and descriptive expectations. Greenspan characterizes the distinction in terms of the notion of prediction, on the one hand, and the notion of holding the agent to a demand, on the other.<sup>8</sup> Wallace ties the notion of normative expectation with various reactive emotions we are inclined to feel when the expectation is frustrated (guilt, resentment).<sup>9</sup> The distinction can be sharpened by appealing to the metaphor of direction-of-fit introduced by G.E.M. Anscombe.<sup>10</sup> Predictive expectations (that  $p$ ), like beliefs, have a mind-to-world fit: if it is the case that not- $p$  the fault lies with the expectation. Normative expectations (that  $p$ ), like intentions, desires, etc., have a world-to-mind fit: if it is the case that not- $p$ , the fault is with the world which ought to be changed accordingly. More precisely, we can say that a person predictively expects that  $p$  when (among other things) he is disposed to dismiss the expectation as having been wrong if not- $p$ . A person  $\beta$

---

<sup>7</sup> The distinction has a long standing in sociology, where normative expectations are taken to define social roles (see e.g. Erving Goffman, *Stigma. Notes on the Management of Spoiled Identity* [New York: Simon & Schuster, 1963]). It has progressively come to occupy a more important place in philosophical literature. For example, Patricia Greenspan has used the notion of reasonable normative expectations to define freedom (“Behavior Control and Freedom of Action,” *Philosophical Review* 87, 1978, 225-240, and “Unfreedom and Responsibility,” in (ed.) Ferdinand Schoeman, *Responsibility, Character, and the Emotions* [Cambridge: Cambridge University Press, 1987], pp. 63-80). A similar distinction (though labeled regularity- and rule-engendered expectation) is at work in Steven Lee’s “Omissions,” *Southern Journal of Philosophy* 16 (1978), 339-354. R.J. Wallace appeals to the notion of normative expectations in giving a compatibilist theory of moral responsibility (*Responsibility and the Moral Sentiments, op. cit.*).

<sup>8</sup> “Unfreedom and Responsibility,” *op. cit.*

<sup>9</sup> This distinction is not crisp, because, as Wallace recognizes, predictive expectations are also often associated with various kinds of emotions. “For example, my expectation about the start of classes may be suffused with a feeling [of] anxiety that has its roots in my childhood experiences of school; the failure of my TV to go on as expected when I activate the remote control may provoke a fit of rage and frustration. But it is not in general the case that expectations of this sort — that is, beliefs about the future — are presumptively associated with any particular attitude” (R.J. Wallace, *Responsibility and the Moral Sentiments, op. cit.*, pp. 20-21).

<sup>10</sup> G.E.M. Anscombe, *Intention. 2nd edition* (Ithaca: Cornell University Press, 1963).

expects (in the normative sense) of another person  $\alpha$  that  $p$  when  $\beta$  is disposed to sanction  $\alpha$ 's failure to bring about  $p$ .

$\beta$  expects (in a normative sense) of  $\alpha$  that  $p$  when  $\beta$  is disposed to impose a negative sanction on  $\alpha$  if  $\alpha$  fails to bring it about that  $p$  and a positive sanction if  $\alpha$  does bring it about that  $p$ .

Correlatively, a person expects of himself that  $p$  when he is disposed to negatively sanction his failure to bring about  $p$  and positively sanction his success in bringing about that  $p$ .

Sanctions are to be understood very liberally. Negative sanctions in particular ought to include the reactive emotions Wallace speaks about. Being susceptible to feeling guilt, resentment or indignation are all forms of being disposed to sanction oneself or others in case of failure to fulfill the expectation.<sup>11</sup> But it includes sanctions of a lesser moral magnitude. Feeling dissatisfied or disappointed by oneself or by another, criticizing oneself or others, etc. are all forms of negative sanctions. But there are also positive sanctions. Various forms of reward or feelings of satisfaction or accomplishment are forms of positive sanction.

In order to understand normative attitudes in terms of sanctions, one must not attempt to reduce normative attitudes to people's (or communities') behavioral dispositions to sanction.<sup>12</sup> Rather, any understanding of normative attitudes must appeal to an already normative notion of sanction. Indeed, it must be the case not only that a person *does* or *tends* to sanction non-conforming behavior but that the person *ought to* sanction it.

---

<sup>11</sup> Wallace discusses cases of irrational guilt, where one feels guilty without believing that one has frustrated any expectations one accepts. In explaining how this is possible Wallace suggests that we must distinguish between the ends that one values and the ends one is motivated to pursue. In our terms, the distinction is one between expecting something of oneself and believing that such an expectation is reasonable. Usually these two attitudes go hand in hand, but it is possible for one to expect of oneself what one believes not to be reasonable, in which case one feels guilty (because one is disposed to sanction oneself) but irrationally or unreasonably because one believes that the expectation is unreasonable.

<sup>12</sup> For an extensive argument, see Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994), pp. 34-46.

The above characterization of what it means for one person to expect something of another does not attempt a reduction of the normative attitude of expectation to a mere disposition. When  $\beta$  expects of  $\alpha$  that  $p$ ,  $\beta$  is required to be disposed to negatively sanction  $\alpha$  in very specific circumstances, viz. when  $\alpha$  fails to bring it about that  $p$ .<sup>13</sup> In other words,  $\beta$  is required to be *correctly* disposed to negatively sanction  $\alpha$ . Likewise,  $\beta$  is required to be *appropriately* disposed to place a positive sanction on  $\alpha$ , when  $\alpha$  does bring it about that  $p$ .

It is usual to think that what is normatively expected are actions. We may expect of someone that he act decently, that he come on time, that he keep his promise, that he move the furniture, and so on. In all these cases, what is expected is an action (rather than a mere happening). This then is one more place at which one needs to be careful if one wants to offer an account of agency in terms of an account of practical responsibility that is grounded in the notion of normative expectations. I will distinguish two concepts of fulfillment (resp. frustration) of normative expectations. A normative expectation is *genuinely* fulfilled (resp. frustrated) only by actions, it is *prima facie* fulfilled (resp. frustrated) by performances in general, which includes actions and mere happenings alike. The expectation of John that he move the furniture is genuinely fulfilled only by John's action of moving the furniture. But it is *prima facie* fulfilled even if John moves the furniture as a result of his falling down, which accidentally pushes the piece of furniture in the required spot. For the sake of linguistic convenience, I will keep the idiomatic agentive complement of expectations. To avoid any circularity, however, I will appeal only to the concept of *prima facie* fulfillment (resp. frustration) of expectations. Thus, when I say that the expectation that John move the furniture is (*prima facie*) fulfilled, this will mean that the expectation is fulfilled by actions (John moving the furniture) and non-actions (John's falling on the furniture as a result of which it is moved) alike.

---

<sup>13</sup> The characterization would not be immune to the charge if the condition of negative sanction were "if  $\beta$  believes that  $\alpha$  fails to bring it about that  $p$ ."

## B. Reasonableness<sub>A</sub> of Normative Expectations

Any concept of responsibility in terms of normative expectations must subject them to a normative standard. Wallace elucidates the concept of moral responsibility in terms of when it is fair to hold someone to a moral expectation. Greenspan analyzes the concept of freedom in terms of when it would be reasonable to hold someone to a normative expectation. Both their concepts of fairness and reasonableness are fairly rich – too rich for a concept of practical responsibility. I shall suggest that in order to focus on the concept of specifically practical responsibility we need to focus on a single component of the concept of reasonableness.

Let us begin by observing that there are at least two ways in which a normative expectation may be unreasonable. One reason why an expectation of a person may be unreasonable is that, as we intuitively say, it is not “within her power” to do what she is expected to do. For instance, it would be unreasonable to expect of an athlete who broke a leg that she take part in a race, of a blind person that he drive a car, or of a newly arrived foreigner that he speak like a native. Another reason why an expectation may be unreasonable is of a different nature. It may be that the person has the power to do what we expect of him, but it may be nonetheless inappropriate for us to expect it of him. Let us suppose that you have a relatively ordinary relationship with your neighbors. You are polite to one another, occasionally help one another out in neighborly matters. But there are (many) expectations that it is simply inappropriate for you to hold them to, and not because it is not “within their power” to fulfill them. For instance, it would be inappropriate for you to expect them to regularly mow your lawn, to do your shopping, etc.

These two kinds of cases exemplify two different, though equally fundamental, concerns with the reasonableness of normative expectations. For want of better terminology, I shall speak of *reasonableness<sub>A</sub>* (agent reasonableness) to capture the first sort of case, and of *reasonableness<sub>N</sub>* (specifically normative reasonableness) in the second kind of case. It is possible for an expectation to be reasonable<sub>A</sub> but unreasonable<sub>N</sub>. Your expectation of your neighbors that they do your shopping would be reasonable<sub>A</sub> (because it is “within their power” to do so) but, under normal

circumstances, it would be highly unreasonable<sub>N</sub> for you to expect it of them. It is also possible for an expectation to be reasonable<sub>N</sub> but unreasonable<sub>A</sub>. A teacher may reasonably<sub>N</sub> expect of his student that she turn the assigned paper on time but the expectation may be unreasonable<sub>A</sub> in view of the fact that the student has been taken to the hospital. Only the former concept of reasonableness<sub>A</sub> matters to attributions of practical responsibility, which is to ground attributions of agency.<sup>14</sup> I will thus focus on it.

There are at least two ingredients in the metaphor of a performance being “within an agent’s power.” First, there is a sense in which the agent must be able to perform the action in question. If the agent could not succeed in performing the action, we would intuitively think that the action was not “within the agent’s power” at the time. An expectation of a two-year old child to win an Olympic swimming competition would surely be unreasonable<sub>A</sub>. Second, there is a sense in which the agent must be able to make a difference. If what the agent is about to do would happen whether or not she did anything, we would be inclined not to think that what happened was in her power. An expectation of someone that he ensure that the sun rises would be unreasonable<sub>A</sub>.

To see this more clearly, let us imagine that we want to test whether an agent can perform a certain type of action. To do so, we will give him a series of tasks, to which he will respond in the best possible way: we are assuming that he is cooperative, that there are no other designs, intentions or expectations in play, the agent is at ease, under no pressure, etc. The tasks are of two kinds, to  $\phi$  and not to  $\phi$ , and they are interspersed randomly in a series.

Four situations are of special interest. Suppose that an agent systematically frustrates the expectation to  $\phi$  (situations (iii) and (iv) in Table 1). When he is expected to  $\phi$ , he does not. In such a case, it would be unreasonable<sub>A</sub> to expect of the agent that he  $\phi$ . The (fully cooperative, capable, etc.) agent cannot succeed in fulfilling the expectation. Suppose that the agent regularly fulfills the expectation to  $\phi$  but frustrates

---

<sup>14</sup> Both concepts of reasonableness<sub>A</sub> and reasonableness<sub>N</sub> seem to be relevant to attributions of moral

the expectation not to  $\phi$  (ii). What this will mean is that the agent  $\phi$ s indiscriminately. In such a case, we would tend to think that the agent's  $\phi$ ing is not up to him, that the agent cannot make a difference, and hence that it would be unreasonable<sub>A</sub> to expect of him that he  $\phi$ . This configuration would obtain if we expected the agent to breathe, for example. Finally (i), when the agent fulfills all the expectations (when expected to  $\phi$ , the agent responds by  $\phi$ ing, when expected not to  $\phi$ , the agent responds by not  $\phi$ ing), we would tend to think that  $\phi$ ing and not  $\phi$ ing are "within the agent's power," that it is not unreasonable<sub>A</sub> to expect of the agent that he  $\phi$ .

	Expectation: $\phi$		Expectation: not- $\phi$	
(i)	fulfilled	( $\alpha \phi$ s)	fulfilled	( $\alpha$ does not $\phi$ )
(ii)	fulfilled	( $\alpha \phi$ s)	frustrated	( $\alpha \phi$ s)
(iii)	frustrated	( $\alpha$ does not $\phi$ )	fulfilled	( $\alpha$ does not $\phi$ )
(iv)	frustrated	( $\alpha$ does not $\phi$ )	frustrated	( $\alpha \phi$ s)

Table 1. Some result patterns of a simplified test sequence (in ideal conditions).

This simplified test scenario allows to make a little clearer some of our intuitions concerning especially the situations in which it would be unreasonable<sub>A</sub> to hold an agent to an expectation. The test scenario is, of course, unrealistic. It presupposes that the agent responds to the expectation in the very best conditions but such conditions are almost never present. It can be approximated, however, if most agents across a wide range of circumstances that approximate the ideal test conditions systematically fulfill or frustrate relevant expectations.<sup>15</sup>

We can dress the intuitions thus:

---

and legal responsibility.

<sup>15</sup> The notion of a systematic correlation is a theoretical placeholder. It is most natural to think about it as involving a counterfactual dependence paradigmatically (though not necessarily) of the causal variety.

(Success Condition):

It is prima facie unreasonable<sub>A</sub> to hold  $\alpha$  to an expectation to  $\phi$  if the expectation to  $\phi$  is systematically prima facie frustrated.<sup>16</sup>

(Difference Condition):

It is prima facie unreasonable<sub>A</sub> to hold  $\alpha$  to an expectation to  $\phi$  if (a) the expectation to  $\phi$  is systematically prima facie fulfilled while the expectation not to  $\phi$  is systematically prima facie frustrated.<sup>17</sup>

An expectation to win a fair lottery is prima facie unreasonable<sub>A</sub>. Such an expectation would be systematically prima facie frustrated. An expectation to speak all the known languages fluently would be systematically prima facie frustrated, so the expectation is prima facie unreasonable<sub>A</sub>. By contrast, an expectation to breathe would be systematically prima facie fulfilled while its contrary would be systematically prima facie frustrated; the expectation is thus prima facie unreasonable<sub>A</sub>. It would also be prima facie unreasonable<sub>A</sub> to expect of a person that she bring it about that the seasons change, for such an expectation would be systematically prima facie fulfilled, while its contrary would be systematically prima facie frustrated.

I will work under the hypothesis that no other conditions characterize prima facie unreasonableness<sub>A</sub>. An expectation is *prima facie reasonable*<sub>A</sub> just in case it is not prima facie unreasonable<sub>A</sub>. As agents, we are guilty until proven innocent. It is reasonable<sub>A</sub> to expect of us any performance unless there are special conditions that would make such an expectation unreasonable<sub>A</sub>.

---

<sup>16</sup> Note that the concept used here is that of prima facie frustration (resp. prima facie fulfillment) of expectations rather than agentive frustration (resp. agentive fulfillment). This is because the concept of reasonableness<sub>A</sub> is ultimately to serve in defining the concept of agency. All the concepts used must therefore be agentively neutral.

<sup>17</sup> The success and difference conditions roughly correspond to the positive and the negative condition in the *stit* theory (Nuel Belnap and Michael Perloff, "Seeing to It that: A Canonical Form for Agentives," in (eds.) H.E. Kyburg, Jr., R.P. Loui and G.N. Carlson, *Knowledge Representation and Defeasible Reasoning* (Dordrecht: Kluwer, 1990), pp. 175-199; Nuel Belnap, "Before Refraining Concepts for Agency," *Erkenntnis* 34 (1991), 137-169.). Unlike the positive condition, the success condition does not require that the success be guaranteed, however. The difference condition excludes the situations where the agent cannot make a difference but without committing us to incompatibilism, as we shall see.

The expectation that a student turn in homework on time is prima facie reasonable<sub>A</sub>. The expectation is not systematically frustrated, nor is it systematically fulfilled while its contrary is systematically frustrated. But when the student falls seriously ill, its reasonableness<sub>A</sub> is defeated. His falling ill is a defeating condition with respect to the expectation to turn in homework on time. There are two basic kinds of defeating conditions corresponding to the success and the difference condition: hindering and compelling conditions, respectively.<sup>18</sup>

- (1) An event of type *C* is a *defeating condition of the first kind (hindering condition)* with respect to an expectation to  $\phi$  iff the occurrence of an event of type *C* is systematically correlated with the prima facie frustration of the expectation to  $\phi$  and with the prima facie fulfillment of the expectation not to  $\phi$ .
- (2) An event of type *C* is a *defeating condition of the second kind (compelling or forcing condition)* with respect to an expectation to  $\phi$  iff the occurrence of an event of type *C* is systematically correlated with the prima facie fulfillment of the expectation to  $\phi$  and with the prima facie frustration of the expectation not to  $\phi$ .<sup>19</sup>

Breaking a leg is systematically correlated with the prima facie frustration to run a race and with the prima facie fulfillment of the expectation not to run a race. It is a defeating condition with respect to the expectation to run the race. Thus, while it may be prima facie reasonable<sub>A</sub> to expect of a person that she run the race, in view of the fact that she has broken a leg, it would be unreasonable<sub>A</sub> to hold her to the expectation. It is prima facie reasonable<sub>A</sub> to expect of a person that he walk. But this expectation ceases to be

---

<sup>18</sup> The terminology is based on von Wright's distinction between hindering (preventing) and compelling (forcing) acts (*Norm and Action* [London: Routledge & Kegan Paul, 1963], 54-55).

<sup>19</sup> There are also defeating conditions of a third kind, which approximate situation (iv) in Table 1. Suppose that an agent's hands tremble erratically, severely impairing his job manufacturing electronic chips, say. However, despite the tremble his chances of succeeding are about 50%. In other words, given an intuitive grasp of 'systematic correlation', it would be wrong to say that the condition is systematically correlated either with the frustration or with the fulfillment of the expectation to connect the chip. Yet, neither the manufacturer of the chips (expecting the agent to connect the chips correctly) nor the rival manufacturer (expecting the agent to sabotage the chip production) can count on the agent. In such a case, our intuitive judgment that what is within the agent's power is limited can be manifested if we subjected the agent to a test. Suppose that the agent was to fulfill a series of expectations to connect the chips

reasonable<sub>A</sub> if the person is in fact physically forced to walk by another. The application of appropriate physical force is systematically correlated with the prima facie fulfillment of the expectation to walk and with the prima facie frustration of the expectation not to walk.<sup>20</sup>

We should note that given the characterization of defeating conditions of the first and the second kind, if *d* is systematically correlated with prima facie frustration of the expectation to  $\phi$  and with the prima facie fulfillment of the expectation not to  $\phi$  then *d* is systematically correlated with prima facie fulfillment of the expectation not to  $\phi$  and with the prima facie frustration of the expectation to  $\phi$ . This is to say that a defeating condition of the first kind is a defeating condition of the second kind for the contrary expectation.

It needs to be emphasized that the concept of a defeating condition is relativized to an expectation. An event-type that may be a defeating condition with respect to one expectation need not be a defeating condition with respect to another. Breaking a leg makes the expectation to run a race unreasonable<sub>A</sub> but it does not defeat the reasonableness<sub>A</sub> of the expectation to remember your friend's birthday.

Just as defeating conditions render prima facie reasonable<sub>A</sub> expectations unreasonable<sub>A</sub>, so *enabling conditions* render prima facie unreasonable<sub>A</sub> expectations reasonable<sub>A</sub>. A large portion of it is occupied by special abilities possibly due to special equipment. It is prima facie unreasonable<sub>A</sub> to expect of a person that she perform a pirouette. But it would be reasonable<sub>A</sub> to hold a skilled skater to the expectation.<sup>21</sup> Having a leg amputated will usually make the expectation to walk without support unreasonable<sub>A</sub>. It will count as a defeating condition of the first kind: it will lead to the

---

correctly and to connect the chips incorrectly, in a random order. In the long run, it would become evident that although the agent does occasionally fulfill the expectations, he systematically frustrates most of them.

<sup>20</sup> One could object that the prime candidates for defeating conditions (of the compelling kind) are desires. They appear to be systematically correlated the fulfillment of the expectation they justify and the frustration of its contrary. I address this worry below.

<sup>21</sup> The example, together with the observation, is borrowed from Annette Baier ("The Search for Basic Actions," *American Philosophical Quarterly* 8, 1971, 164).

systematic frustration of the expectation. However, when the agent is equipped with a prosthesis the expectation would no longer be systematically frustrated.

There are also *undefeating conditions*. Let us suppose that  $C$  is systematically correlated with the frustration of the expectation to  $\phi$ . Intuitively, when  $C$  occurs it is not “within the agent’s power” to  $\phi$ . A question that might be reasonably raised is: Is it “within the agent’s power” to see to it that  $C$  does not occur? If it is then the defeating power of  $C$  is itself defeated. Despite the occurrence of  $C$  it will be reasonable<sub>A</sub> to expect of the agent that he  $\phi$ . Drinking an immoderate amount of alcohol will reliably result in a loss of much control: it is systematically correlated with the frustration of a range of expectations (to drive safely, to behave responsibly, etc.). However, if it was reasonable<sub>A</sub> to expect of the agent not to ingest an immoderate amount of alcohol, then it is reasonable<sub>A</sub> to expect of the agent that she behave responsibly, that she drive safely, etc. Likewise, it may appear to be unreasonable<sub>A</sub> to expect of a committee member to be at the meeting of the committee if he is fast asleep. But as long as it was reasonable<sub>A</sub> to expect of him not to oversleep, it is also reasonable<sub>A</sub> to expect of him that he be at the meeting.

This allows one to see why desires, which are often systematically correlated with the fulfillment of the expectation they justify and the frustration of its contrary, do not render those expectations unreasonable<sub>A</sub>. If a desire to  $\phi$  is systematically correlated with the fulfillment of the expectation to  $\phi$ , it will render that expectation unreasonable<sub>A</sub> but only if it is unreasonable<sub>A</sub> to expect of the agent that she prevent her desire from affecting her action. It is arguable that some desires are like that. The compulsive-obsessive’s desire to wash his hands every five minutes is not something that he can control. But this is not so for most other desires. Even if someone’s desire for chocolate is very strong, so that whenever he has it he submits to it and gobbles up all chocolate in sight, it would (or at least might) be reasonable<sub>A</sub> to expect him not to eat the chocolate. It might be reasonable<sub>A</sub> to expect him to control the desire (for example, by making sure that there is no chocolate around). Contrary to appearances, then, not all desires render the expectations they justify unreasonable<sub>A</sub>.

We can thus offer the following characterization of reasonableness<sub>A</sub>:

It is reasonable<sub>A</sub> (at  $t$ ) to expect of  $\alpha$  that  $\alpha \phi$  (at  $t'$ ) if and only if either (a) no defeating condition (with respect to the expectation to  $\phi$  at  $t'$ ) is in force at  $t$ ,<sup>22</sup> or (b) such a defeating condition is in force at  $t$  but it has been countered by an appropriate enabling condition, or (c) such a defeating condition is in force at  $t$  but it is unreasonable<sub>A</sub> (at  $t$ ) to expect of  $\alpha$  that  $\alpha$  bring it about that it not be in force at  $t$ .

### C. Practical Responsibility

The concept of reasonableness<sub>A</sub> so construed is sensitive to temporal considerations. An expectation may be quite reasonable<sub>A</sub> at a certain point in time and only become unreasonable<sub>A</sub> when a defeating condition occurs later. Such a defeating condition may be defeated further at a still later time, and so on. In assessing whether an agent is practically responsible for  $\phi$ ing at  $t$ , we need to determine whether it would be reasonable<sub>A</sub> at  $t$  to expect of the agent that she  $\phi$  at  $t$ . The latter assessment will have to appeal to any defeating, enabling and undefeating conditions that may have occurred any time between the onset of the normative expectation and time  $t$ .

(P) An agent  $\alpha$  is practically responsible for  $\phi$ ing (at  $t$ ) just in case it would be reasonable<sub>A</sub> (at  $t$ ) to expect of  $\alpha$  that  $\alpha \phi$  (at  $t$ ).

This completes the very brief explanation of the sense of practical responsibility intended here.

The concept of practical responsibility just delineated is forward-looking. It is a practical correlate to what Kurt Baier calls task-responsibility.<sup>23</sup> Unlike backward-looking concepts of responsibility (in Baier's scheme: answerability, culpability, liability) it does not presuppose that an action has been performed. Furthermore no concepts are used that would presuppose that the distinction between what is agentive and what is not has been drawn. The concept of defeating conditions is formulated with the agentively

---

<sup>22</sup> Note that the characterization appears to miss cases where no defeating conditions occur but the expectation is prima facie unreasonable<sub>A</sub> (the expectation to make sure that  $2+2=3$ , e.g.). For simplicity, I will treat the case of prima facie reasonableness<sub>A</sub> and prima facie unreasonableness<sub>A</sub> as relative to a special tautologous defeating condition.

<sup>23</sup> Kurt Baier, "Responsibility and Action," in (eds.) Michael Bradie and Myles Brand, *Action and Responsibility* (Bowling Green, OH: Bowling Green University Press, 1980), pp. 100-116.

neutral concept of prima facie fulfillment (frustration, respectively) of expectations. As a result, the concept of practical responsibility can be used to ground the notion of agency, thus satisfying one of criteria of adequacy set out in section 2. We will now see that it satisfies the others.

#### 4. Practical Responsibility and Frankfurt-Style Cases

Frankfurt-style cases have been designed to show that one can be morally responsible for an action despite the fact that one could not have done otherwise. Suppose that Jones decides to kill the mayor of the town. He carries out his plan to the letter, shoots the mayor who dies as a result. Unbeknownst to Jones, evil scientists have implanted a device into Jones' brain which, were Jones to decide not to kill the mayor (or waver after his decision), would have swayed Jones to kill the mayor anyway. The intuitions about cases of this sort have been almost uniform. Jones is responsible for killing the mayor. At the same time, it has been claimed, Jones could not have done otherwise: he could not have not killed the mayor.

We need to consider the question whether the agent in Frankfurt-style cases is practically responsible (according to the present account) for performing the action he actually performs. The account will be supported if it can answer the question positively. While Frankfurt-style cases have not been formulated to test our intuitions concerning practical responsibility, they presuppose that the agent performs an *action* in the actual sequence of events. On the account sketched, this implies that he must be practically responsible for performing it. There are, however, reasons to think that this is not the verdict dictated by the account.

Consider the example once again. Is Jones practically responsible for killing the mayor? In other words, is it reasonable<sub>A</sub> to expect of Jones that he kill the mayor (in the actual sequence of events). One might think that it is not. After all, given the presence of the counterfactual intervener it is settled that the mayor will die at Jones' hands. It would thus seem that the presence of the counterfactual intervener is systematically correlated with the prima facie fulfillment of the expectation that Jones kill the mayor and the prima

facie frustration of its contrary. It should therefore count as a defeating condition of the compelling kind.

Let us first note that the presence of the counterfactual intervener is a rather peculiar kind of condition, especially if it is compared with the intervener's actual intervention. The structure of the cases can be captured thus. There are two possible paths that lead to the death of the mayor at Jones' hands. The first is the "normal" path where Jones decides to kill the mayor ( $D_K$ ) and follows his decision through. The second would be the result of the counterfactual intervener's interference with Jones's attempt not to kill the mayor ( $C$ ). The counterfactual intervener would take over the control of Jones's body and lead him to the same result. There is no doubt at all that if the latter were to happen Jones would not be responsible for the mayor's death. In fact, Jones would not have killed the mayor at all — at most the mayor would have died at Jones' hands, which were but an instrument of the counterfactual intervener. Indeed, the intervention is a classic defeating condition. It is systematically correlated with the fulfillment of the expectation that Jones kill the mayor and the frustration of the contrary expectation. Hence it would be unreasonable<sub>A</sub> to expect of Jones that he kill the mayor. While  $C$  is a defeating condition,  $D_K$  is not.<sup>24</sup> What is special about the Frankfurt-style cases is that the very presence of the counterfactual intervener, i.e.  $D_K$ -or- $C$ , ensures that the mayor dies at Jones' hands. So  $D_K$ -or- $C$  is a defeating condition of the second kind: it compels Jones to kill the mayor.<sup>25</sup> It would then appear that Jones is not practically responsible for killing the mayor and so not morally responsible either.

---

<sup>24</sup> One might think that to the extent that a decision to  $\phi$  is systematically correlated with the fulfillment of the expectation to  $\phi$  and the frustration of the expectation not to  $\phi$ , it should count as a defeating condition of the compelling kind. However, the defeating power of decisions will usually be defeated. Just as desires do not render the expectations they justify unreasonable<sub>A</sub> (as long as it is reasonable<sub>A</sub> to expect of the agent that he not act on them), so as long as it is reasonable<sub>A</sub> to expect that the agent not act on a decision, decisions likewise do not render correlated expectations unreasonable<sub>A</sub>.

<sup>25</sup> One may think that some theoretical leverage could be gained from paying attention to the complement of the expectation (the expectation that Jones kill the mayor vs. the expectation that the mayor dies at Jones' hands). This is in fact the basis for a *stit*-theoretic response to Frankfurt-style cases. (See N. Belnap, M. Perloff, "Seeing to It that," *op. cit.*; N. Belnap, "Before Refraining Concepts for Agency," *op. cit.*, and Chapter 5 of *Facing the Future: Actual Agents, Real Choices*, by Nuel Belnap, Michael Perloff and Ming Xu, with contributions by Paul Bartha, Mitchell Green and John Horty, forthcoming, available at

This would be the end of story were it not for the fact that, as I remarked, our judgments of practical responsibility are, as they should be, sensitive to temporal considerations. Let us consider two kinds of Frankfurt-style cases. The simpler case is one where Jones makes his decision (at  $t_0$ ) and will not change his mind. In other words, the counterfactual intervener will intervene only immediately after time  $t_0$ . The more complex case is one where although Jones does decide to kill the mayor at  $t_0$ , he might change his mind later. So, the counterfactual intervener could intervene not only immediately after  $t_0$  but also any time before the time that the mayor actually dies ( $t_K$ ).

Take the simpler case first. If Jones chooses at  $t_0$  to kill the mayor (and as we are simplifying the case, the counterfactual intervener is no longer a factor), it is reasonable<sub>A</sub> to expect of Jones that he kill the mayor. This is because the condition  $D_K\text{-or-}C$  is no longer at work.<sup>26</sup> Its power lasts only as long as it is unresolved which of the pathways will be taken. Analogously, if Jones chooses at  $t_0$  not to kill the mayor, it is reasonable<sub>A</sub> to expect of him not to kill the mayor up until the time of the counterfactual intervener's interference. It is only thereafter that it becomes unreasonable<sub>A</sub> to expect of the agent that he kill the mayor in view of the actual intervention. In other words, the defeating power of such a defeating condition is only apparent. Once the time comes when it is resolved whether  $D_K$  or  $C$  will occur, the defeating power of the condition will disappear.

Consider the more complex case. We are now supposing that the counterfactual intervener may intervene at any later time  $t_i$  (prior to time  $t_K$  of the mayor's death) were Jones to waver at that time and decide not to kill the mayor. In other words, we must suppose that though the condition  $D_{Kt_0}\text{-or-}C_{t_0}$  is no longer at work after  $t_0$ , similar conditions  $D_{Kt_i}\text{-or-}C_{t_i}$  are at work, all of which jointly ensure that the mayor will die at

---

<http://www.pitt.edu/~belnap/ff/>. I discuss it in some detail in "Flickers of Freedom and Frankfurt-Style Cases in the Light of the New Incompatibilism of the *Stit* Theory," forthcoming in *Journal of Philosophical Research*.) This response is unavailable to us, however, since the concept of practical responsibility must not presuppose the concept of agency.

<sup>26</sup> Note that the disjunctive defeating condition is at work only as long as it is unsettled which of the disjoined conditions is at work. Although the disjunction " $D_K$  or  $C$  is present" will be still true when it is settled that  $D_K$  but not  $C$ , the disjunctive defeating condition stops operating – the only condition that is at work is  $D_K$ .

Jones' hands one way or another. It might therefore seem that it would be after all unreasonable<sub>A</sub> at  $t_K$ , when Jones actually kills the mayor of his own accord, to expect of him that he do so, because up until  $t_K$  the counterfactual intervention might occur. This would lead us to conclude that Jones was not responsible for the killing after all.

But the reasoning leading to the conclusion is faulty. The defeating conditions do not actually operate up until  $t_K$ . There is a time lapse between the time  $t_K$  at which the mayor dies (is killed by Jones) and the time  $t_{K-\epsilon}$  at which Jones can no longer change his mind about killing the mayor (when the revolver has just been fired, or when the right neural impulses have been sent to the muscles of the fingers). At this time ( $t_{K-\epsilon}$ ), the series of  $D_{K_{t_i}}$  - or -  $C_{t_i}$  conditions ceases to work. For it is no longer possible for Jones to waver. Thus,  $D_{K_{t_i}}$  - or -  $C_{t_i}$  type conditions do not manage to render the relevant expectations unreasonable<sub>A</sub> at the time that counts, viz.  $t_K$ , for they cease to operate at  $t_{K-\epsilon}$ .<sup>27</sup> And the same reasoning applies. Once, at  $t_{K-\epsilon}$ , it is settled that  $D_K$ ,  $C$  is not longer a factor and neither is  $D_K$ -or- $C$ .

Another way of putting the last point is this. The defeating power of the  $D_K$ -or- $C$  conditions stems from peculiar disjoining of a genuine defeating condition (the intervention) with what is not a defeating condition (the agent's decision). As long as it is not clear which of the paths (the path where  $D_K$  is operative or the one where  $C$  is operative) is the actual one,  $D_K$ -or- $C$  functions as a defeating condition. However, part of the very description of a Frankfurt-style case makes it abundantly clear that the agent *performs an action* achieving a certain result *without* the counterfactual intervener's *intervention*. It is thus clarified in the very structure of the cases that  $C$  is not operative

---

<sup>27</sup> One might imagine another version of the case, where the counterfactual intervention occurs not when the agent has a change of heart, but rather when something were to go wrong. For example, were the wind to displace the bullet in such a way that the mayor would not be shot, the counterfactual intervener would ensure (by means of another device implanted in the mayor's brain, say) that the mayor moves just the right amount so that he is shot after all. But there again has to be a time lapse (imposed by the laws of physics not psychology this time) by which the intervener must actually intervene in order for the changes to have the right effect. It is at this point that the spell of the defeating condition is lost.

but only that  $D_K$  is. Since  $D_K$  is not a defeating condition, it is reasonable<sub>A</sub> to expect of the agent that he perform the action.

Jones can be held responsible for following through his decision to kill the mayor at  $t_K$  since it is reasonable<sub>A</sub> to expect (at the time of the action  $t_K$ ) of Jones that he kill the mayor at  $t_K$ . If the counterfactual intervener did actually intervene leading Jones to the killing, the intervention would render the expectation unreasonable<sub>A</sub>.

## 5. Practical Responsibility, Actions and Omissions

What is central to Frankfurt-style cases is the thought that an agent can be morally (and practically) responsible for doing something even though she could not have prevented it. As has been first noted by van Inwagen,<sup>28</sup> there appear to be cases of omissions, where despite having appreciated the Frankfurt-style lesson we are nevertheless inclined to judge the agent as not morally responsible for the omission precisely on the grounds that she could not have prevented the result. I will call them van Inwagen-Fischer-style (or IF-style, for short) cases. This has led many (among them van Inwagen and Fischer, though Fischer revises his earlier stance in *Responsibility and Control*) to suppose that there is an asymmetry between our attributions of responsibility for omissions and actions.<sup>29</sup> But this turns out not to be the case and the situation is even more puzzling because there are cases of omissions for which we can be held morally responsible even though we could not have prevented them.

---

<sup>28</sup> Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983).

<sup>29</sup> The original thesis appears in Peter van Inwagen's *An Essay on Free Will* and John Martin Fischer's "Responsibility and Failure," *Proceedings of the Aristotelian Society* 86 (1985/86), 251-270. Fischer has elaborated it in a paper with M. Ravizza ("Responsibility and Inevitability," *Ethics* 101, 1991, 258-278), and upheld in *The Metaphysics of Free Will. An Essay on Control* (Oxford: Basil Blackwell, 1994). In the meantime, the thesis has received much critical attention, see Randolph Clarke, "Ability and Responsibility for Omissions," *Philosophical Studies* 73 (1994), 195-208; Harry G. Frankfurt, "What We Are Morally Responsible for," in *The Importance of What We Care About, op. cit.*, pp. 95-103 and "An Alleged Asymmetry between Actions and Omissions," *Ethics* 104 (1994), 620-623; Ishtiyaque Haji, "A Riddle Regarding Omissions," *Canadian Journal of Philosophy* 22 (1992), 485-502; Alison McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency," *Philosophical Review* 103 (1994), 453-488; David Zimmerman, "Acts, Omissions and 'Semi-Compatibilism'," *Philosophical Studies* 73 (1994), 209-223. In their most recent work (*Responsibility and Control, op. cit.*), Fischer and Ravizza elaborate their account so that the thesis turns out to be in fact false.

Here is a case of an IF-style omission for which the agent is not responsible. Jones does not have any fancy mechanism in his brain. He is strolling along the beach when he sees a child struggling in the water. Though he believes he can rescue the child with little effort, he decides not to go to the trouble. The child drowns. Unbeknownst to Jones, had he jumped into the water, the sharks swimming nearby would have attacked him. So Jones could not have saved the child after all. It appears that Jones is not responsible for failing to rescue the child precisely because he could not have rescued her. This is not deny that Jones is responsible for something. He is responsible for his failure to *try* to save the child but he is not responsible for his failure to save the child.

In the IF-style cases, the defeating condition (here: the presence of the sharks is a defeating condition with respect to the expectation to save the child) does not counterfactually depend on the agent's decision. It is a condition that operates up front, as it were. In view of the presence of the sharks, it would be unreasonable<sub>A</sub> to expect of Jones that he save the child. This is because the presence of the sharks is assumed to be systematically correlated with the prima facie frustration of the expectation to save the child. As long as it is unreasonable<sub>A</sub> to expect of Jones that he prevent the sharks from attacking, it is unreasonable<sub>A</sub> to expect of Jones that he rescue the child.<sup>30</sup> The presence of the sharks is thus a standard defeating condition with respect to the expectation to save the child and its presence alone accounts for our judgement that Jones is not responsible for saving the child.

Here is an example of an omission that exactly parallels the structure of Frankfurt-style actions.<sup>31</sup> Brown has an implant similar to Jones'. She is walking along the beach and sees a child struggling in the water. Though Brown cannot swim, she can throw a life jacket but decides not to. The child drowns. Unbeknownst to Brown, had she shown any inclination to try to save the child, the implant would have been activated

---

<sup>30</sup> This echoes Frankfurt's response to the case: "The real reason [why Jones bears no moral responsibility] is that what he does has no bearing at all upon whether the child is saved. The sharks operate both in the actual and in the alternative sequences, and they see to it that the child drowns no matter what John does" ("An Alleged Asymmetry between Actions and Omissions," *op. cit.*, 623).

as a result of which Brown could not attempt to rescue the child after all. As it happens, the implant did not need to be activated. In this case, the intuition seems to be that Brown is morally responsible for failing to throw the jacket to the child even though she could not have done otherwise.

The case is exactly parallel to Frankfurt-style cases of actions. It might appear that it is unreasonable<sub>A</sub> to expect of Brown that she not throw the life jacket, for given the arrangement of the case, the expectation not to throw the life jacket will be systematically *prima facie* fulfilled while its contrary – *prima facie* frustrated. The defeating condition is a composite of the decision not to throw the life jacket *T* and the scientist's possible intervention *C*. The occurrence of *T-or-C* is systematically correlated with the *prima facie* fulfillment of expectation not to throw the life jacket and the *prima facie* frustration of the contrary expectation. Once again the case is constructed in such a way that it is clear that only one of the conditions (*viz. T* but not *C*) has operated (the counterfactual intervener does not in fact intervene). We can therefore conclude that no defeating condition operated at the relevant time, and so that Brown was practically responsible for throwing the life jacket.

The account sketched can thus sort the intuitions correctly without presupposing that there is a deep asymmetry between actions and omissions. The difference concerns rather the structure of the cases. In Frankfurt-style cases of both actions and of omissions, the defeating condition ceases to operate with the agent's decision (in the simplified case) or when the agent can no longer waver. But in the IF-style cases of omissions, the defeating conditions is present throughout. If I am right then it ought to be possible to construct an IF-style case of action, where the agent should likewise not be held responsible.

It is in general easier to describe an IF-style omission than an IF-style action but perhaps the following example will bring the point home. It does not involve a counterfactual intervener. Smith wants to switch on the light. He presses the switch. The

---

<sup>31</sup> The case is borrowed from I. Haji, "A Riddle Regarding Omissions," *op. cit* A similar case is constructed by H. Frankfurt, "An Alleged Asymmetry between Actions and Omissions," *op. cit*

light comes on. It might appear that Smith switched on the light. However, it so happens that unbeknownst to him, the light would have come on at exactly the moment that he actually pressed the switch but independently of his intervention (the light is regulated by a computer). It seems intuitively plausible to describe the case as that of Smith having nothing to do with the light going on. (Indeed he could not have done otherwise: he could not have not switched on the light.) It would be inappropriate to hold Smith responsible for switching on the light. Smith might still be held responsible for flipping the switch but not for actually switching the light on. This is indeed borne out if we ask whether it was reasonable<sub>A</sub> to expect of Smith that he switch on the light. Given that the light will come on at  $t$ , it is unreasonable<sub>A</sub> to expect of Smith that he (not) switch on the light at  $t$ . The fact that the light will come on at  $t$  is systematically correlated with the prima facie frustration of the expectation that Smith not switch on the light at  $t$  and with the prima facie fulfillment of the expectation that Smith switch on the light. Hence, it was unreasonable<sub>A</sub> to expect of Smith that he (not) switch on the light at  $t$ .

On the account presented, the asymmetry thesis is false, which coincides with many intuitions on the matter. The account also makes clear what structure is responsible for our intuitions aligning as they do. The IF-style cases can be easily reconstructed as cases where the responsibility judgement is defeated by a defeating condition. Frankfurt-style cases, on the other hand, are ones where the power of the defeating condition is ultimately dissolved before the action takes place.

## **6. Practical Responsibility and Consequences of Actions**

Exactly the same form of puzzle as in the case of omissions arises with respect to responsibility for consequences. There are three cases in particular around which Fischer and Ravizza build their account. Two of them (“Missile 1” and “Missile 2”) are cases where the agent is morally responsible for the consequence even though she could not prevent it from happening. The remaining one (“Missile 3”) is a case where the agent is not morally responsible for the consequence and where we are inclined to think that this is because she could not prevent it from happening. Fischer and Ravizza develop a two-pronged account of responsibility for consequences – the two prongs correspond to the

two causal paths: the path leading to the action, and the path leading from the action to the consequence. First, they require that the mechanism which leads to the action be moderately reasons-responsive. Second, they require that the path from the action to the consequence be sensitive to the action. I will not argue in the abstract that both of these requirements are subsumed under the requirement of reasonableness<sub>A</sub>. Instead, I will consider how the account applies to the particular cases they focus on.

It may be useful to be reminded, however, that from our point of view one should not expect for there to be anything analogous to the two prongs. One should be able to ascertain in principle whether the agent is practically responsible for  $\phi$ ing at any point. What matters for the assessment of practical responsibility is whether it is reasonable<sub>A</sub> to expect of the agent that he  $\phi$  up until the very time when he is expected to  $\phi$ . In other words, the judgment of practical responsibility will be sensitive to everything that happens up until the time of  $\phi$ ing.<sup>32</sup> Two points need to be kept clearly in mind, however.

First, we need to consider how the occurrence of the action affects the agent's practical responsibility for its consequences. To do so, we simply have to ask whether it would be reasonable<sub>A</sub> to expect of the agent that the consequence occur given that he has performed such and such action. Suppose that the agent performed an action of  $\phi$ -ing (the agent was practically responsible for  $\phi$ -ing) at  $t$ , and that as a result of his  $\phi$ -ing a consequence  $C$  occurred at  $t'$ . If there is a systematic correlation between the agent's  $\phi$ -ing and the occurrence of  $C$ -type consequences then it might appear that the agent's  $\phi$ -ing is a defeating condition with respect to the expectation that the agent bring it about that  $C$ . However, in such cases it will have also been reasonable<sub>A</sub> to expect of the agent that

---

<sup>32</sup> This commits the account to a certain resolution of issues concerning the individuation and in particular the time-indexing of actions. On a currently popular view, though an action can be described in terms of its consequences, the timing of the consequences does not affect the timing of the action (Donald Davidson, "Agency," in *Essays on Actions and Events* [Oxford: Clarendon Press, 1980], pp. 43-61). Suppose Susan fires a gun at  $t$ , wounds a bird (at  $t'$ ) and the bird dies (at  $t''$ ). On Davidson's view, Susan performs one action at  $t$ , and that action can be described in terms of consequences that occur later. Thus somewhat awkwardly, Susan kills the bird at  $t$  (even though the bird dies only at  $t''$ ). By contrast, on the framework developed here, one can speak of three actions occurring (or being completed) at the time that the various consequences occur.

he (not)  $\phi$ , thereby defeating the defeating power of the condition. To give a concrete example, suppose that Keith's seasoning a dish ( $\phi$ , at  $t$ ) made it into a masterful entrée ( $C$ , at  $t'$ ). Since Keith is a master-chef, fully reliable in seasoning dishes to perfection, Keith's seasoning the dish is a defeating condition with respect to the expectation that he create a masterful entrée. The reason why Keith is nonetheless practically responsible for creating a masterful entrée even after he seasoned the dish, is that in this case it was reasonable<sub>A</sub> to expect of Keith that he (not) season the dish.

Second, it should be born in mind that on the account of practical responsibility developed here, we are always practically responsible for events under certain descriptions. By contrast, Fischer and Ravizza develop separate accounts of responsibility for actions (event-particulars) and consequence particulars and of responsibility for consequences under certain descriptions (consequence universals). They do not argue for the need to introduce the former and without further argument I see no need for it. I will therefore rest content with the account of practical responsibility as it has been laid out.

Let us now turn to the cases.

*Missile 1*: An evil agent, Elizabeth, decides to launch a missile toward Washington, D.C. An evil scientist has implanted a device into Elizabeth's brain. If Elizabeth were to waver in her decision, he would employ the device taking over control of her body and steering her to launching the missile any way. Once the missile is launched, Elizabeth cannot prevent it from hitting the city. As it happens, Elizabeth launches the missile toward Washington, D.C., on her own (there is no need for the counterfactual intervention). Elizabeth is morally responsible for launching the missile (even though she could not have prevented it) as well as for hitting Washington, D.C., (even though she could not have prevented it).

The first case "Missile 1" can be understood along the lines I suggested for understanding Frankfurt-style cases in general. Elizabeth is practically responsible for launching the missile as well as for bombing Washington, D.C., somewhere or other (i.e.

for the consequence that Washington is bombed somewhere or other).<sup>33</sup> Consider these in turn.

At  $t_0$ , Elizabeth makes her choice to launch the missile ( $D_L$ ) and, as a matter of fact, launches it. Prior to  $t_0$ , it would be unreasonable<sub>A</sub> to expect of Elizabeth to launch or not to launch the missile because, as the case is set up, the disjunctive condition  $D_L$ -or- $C$  operates: if Elizabeth does not decide to launch the missile, she will be led to do so via the intervention of the counterfactual intervener ( $C$ ). However, at  $t_0$ , Elizabeth's resolution in effect suspends the operation of the counterfactual intervener. The only condition that operates is  $D_L$ . Since  $D_L$  is not a defeating condition and no other defeating conditions occur, it is reasonable<sub>A</sub> at the time of the launch to expect of Elizabeth that she launch the missile. So, Elizabeth is practically responsible for launching the missile.

It is likewise reasonable<sub>A</sub> to expect of Elizabeth that Washington, D.C., is bombed (somewhere or other). We may think of the bombing of the city as being a consequence of her launching the missile. There are a number of possible defeating conditions that might make it unreasonable<sub>A</sub> to expect of Elizabeth that she bomb Washington, D.C. For example, the missile may be "blank" or the apparatus may be such that the missile would steer off course, etc. However, none of such defeating conditions are present. Another defeating condition is Elizabeth's action of launching the missile provided that Elizabeth is quite expert in launching missiles on chosen targets because then her launching a missile toward any target of choice will be systematically correlated with the expectation that she bomb the target of choice (*ceteris paribus*). Despite the presence of this correlation, however, it will be still reasonable<sub>A</sub> to expect of Elizabeth that she bomb Washington, D.C., in this case, because it was reasonable<sub>A</sub> to expect of Elizabeth that she (not) launch the missile. In the absence of any further defeating conditions, Elizabeth is practically responsible for bombing Washington, D.C.

---

<sup>33</sup> Recall that in the conceptual order of things, the distinction between agentive and nonagentive complements does not affect the workings of the concept of practical responsibility. Thus being practically responsible for the action under the description "bombing Washington, D.C." and for the consequence that Washington, D.C., is bombed amount to the same.

In the above case, the counterfactual intervention occurs prior to the agent's action. Fischer and Ravizza consider another case where the counterfactual intervention affects the path between the action and the consequence.

*Missile 2:* As before, Elizabeth launches the missile. This time, however, there is no intervener who can steer Elizabeth to launch the missile, rather there is another woman, Carla, who would launch the missile, if Elizabeth decided not to. Once again, Elizabeth is morally responsible for the launch even though she could not have prevented it (Elizabeth cannot prevent Carla from launching the missile).

The second case, "Missile 2", is handled analogously. Prior to Elizabeth's launching the missile (*L*), the defeating condition *L-or-R* operates: if Elizabeth does not launch the missile, Carla will launch it instead (*R*). *L-or-R* is systematically correlated with the fulfillment of the expectation that Washington, D.C., is bombed (somewhere or other). Since Elizabeth launches the missile (*L*), Carla's possible intervention is no longer a factor: *L-or-R* ceases to operate. The only defeating condition in operation is *L*, i.e. Elizabeth's launching the missile, but since it was reasonable<sub>A</sub> to expect of Elizabeth that she launch the missile, its defeating power is defeated.

It is important to note at the start that one should not expect that there would be much difference between the way in which the first two cases are handled by the account of practical responsibility given. As stressed, practical responsibility is an intensional concept: what we are practically responsible for (under the above construal) is always sensitive to the way that the event is described. In Fischer and Ravizza's parlance, what we are practically responsible for are universals. By contrast, Fischer and Ravizza claim to develop an account of responsibility for both particulars and universals.

Finally, consider the third case where it seems that the agent is morally responsible for the consequence precisely because she could not have prevented it.

*Missile 3:* Joan, who knows that Elizabeth has launched the missile toward Washington, D.C., has in it in her power to deflect the missile in such a way that a less populous area of the city will be hit. She cannot, however, prevent the missile from hitting the city altogether. When Joan does use her weapon, she can be morally responsible for the fact that one section of the city (rather than another) is hit, but she is not morally responsible for the fact that Washington, D.C., is bombed

(somewhere or another), precisely because she could not have prevented it.

In this case, Joan is not practically responsible for preventing Washington, D.C., from being hit altogether, even though she is practically responsible for redirecting the missile and so for preventing a more populous area from being hit or, in other words, ensuring that a less populous area is hit.

The reason why it is unreasonable<sub>A</sub> to expect of Joan that she prevent Washington, D.C., from being bombed altogether is due to the fact that a defeating condition is present: Elizabeth has launched the missile. Since Elizabeth is quite expert in launching missiles, her launching the missile is systematically correlated with the destruction of the target of her choice. At the same time, this defeating condition is not defeated since (as the case is described) it is unreasonable<sub>A</sub> to expect of Joan that she prevent Elizabeth from launching the missile. The situation changes, however, when we consider whether Joan is practically responsible for redirecting the missile. For here Elizabeth's launching the missile is not systematically correlated with the frustration of the expectation that Joan redirect the missile or with the fulfillment of that expectation. Likewise, Joan is practically responsible for ensuring that a less populous area of Washington, D.C., is hit. While Elizabeth's launching the missile is systematically correlated with the frustration of the expectation that Joan prevent Washington, D.C., from being hit altogether, it is not systematically correlated with the frustration of the expectation that Joan prevent the more populous areas of Washington, D.C., from being hit.

The application of the account of practical responsibility yields the same results as that of Fischer and Ravizza. It has the advantage of not requiring a separate treatment of the two prongs: the pathway leading to action and the pathway leading from action to consequence. It offers a unified approach in this respect. It treats the Frankfurt-style cases of consequences (Missile 1 and Missile 2) in the same way that it treats all Frankfurt-style cases – by demonstrating that the apparent power of a defeating condition dissipates by the time the action takes place, i.e. by the time that matters for the assessment of practical responsibility. Although in such a case, it is impossible for the agent to prevent

the result from happening, the agent is practically responsible for it because – appearances to the contrary – no defeating condition is in operation at the time of the performance. By contrast, in the last case (Missile 3) the agent is not practically responsible for the consequences with respect to which defeating conditions operate.

The consideration of all the cases shows that the connection between inevitability of an outcome and practical responsibility for it is not straightforward. The fact that an outcome is inevitable does not yet show that one is not practically responsible for it. Minimally, one has to trace the source of the “inevitability.” One needs to ask whether the condition that makes an outcome inevitable has been defeated by another condition. (The mere fact that one is asleep and so unable to attend a meeting does not yet demonstrate that one is not responsible for failing to attend it. In fact as long as it was reasonable<sub>A</sub> to expect of one that one not be asleep, one is practically responsible for attending it.) And one needs to ask whether the condition that makes an outcome inevitable is still in operation at the time of the performance. (In Frankfurt-style cases, the disjunctive condition that makes the outcome inevitable does not in fact operate at the time of the action.)

## **7. Compatibilism or Incompatibilism?**

The conception of practical responsibility just sketched is not committed to either the position of compatibilism or incompatibilism. This carries with it the advantage that it can appeal to both.

It may appear, however, that the account is committed to incompatibilism. After all, defeating conditions may appear to function much like “determining conditions.” And although the characterization of defeating conditions in terms of systematic correlations is admittedly vague, it is not as vague as to leave in any doubt that a cause is systematically correlated with its effect. But if so, then the truth of determinism would indicate that for every action there is a sufficient cause that is systematically correlated with it. In other words, for every action it will be unreasonable<sub>A</sub> to expect of the agent that he perform it. Suppose the agent  $\phi$ s. The expectation to  $\phi$  will be unreasonable<sub>A</sub> because there would be an event  $c$  (that is the sufficient cause of the agent’s  $\phi$ ing) which

is systematically correlated with the agent's  $\phi$ ing. Thus, insofar as we are speaking of responsibility, we must be committed to incompatibilism.

I do not want to reject incompatibilism in favor of compatibilism. But I do want to dispel the impression that incompatibilism is forced on this framework. The above argument does not appreciate that the relation of being a defeating condition with respect to an expectation is intensional. One event  $c$  may be a cause of another event  $e$ , and that may be sufficient for there to be a systematic correlation between the two events but not necessarily under the descriptions that matter to the assessment of reasonableness<sub>A</sub> of expectations. Thus, the conception is not committed to incompatibilism. At the same time, it is not committed to compatibilism either in that nothing in it precludes the thought that the systematic correlations hold between such events under the descriptions that matter to the assessment of reasonableness<sub>A</sub>.

## 8. Conclusion

I have put forward a concept of practical responsibility. Its adequacy and usefulness will only be revealed in how well it is suited to the development of a conception of agency. For now, I have only shown that it is minimally suited for this purpose. It meets the criteria of adequacy for a concept of responsibility as they are laid out by Fischer and Ravizza and it does not presuppose the concept of agency. The latter characteristic suffices to demonstrate that H.L.A. Hart's responsibility-based theory of action should be freed from the layers and layers of dust that have covered it after seemingly devastating critiques.<sup>34</sup> This alone is a worthwhile achievement.

---

<sup>34</sup> Although many more points have been in question in Hart's theory, the "mix-up" of the conceptual order of the concept of responsibility and action was at the heart of most of the critiques. See George Pitcher, "Hart on Action and Responsibility," *The Philosophical Review* 69 (1960), 226-235; Joel Feinberg, "Action and Responsibility," in (ed.) Alan R. White, *The Philosophy of Action* (Oxford: Oxford University Press, 1968), pp. 95-119; Christopher Cherry, "The Limits of Defeasibility," *Analysis* 34 (1974), 101-107.