CHAPTER VII.

SELECTIONAL FORCE OF REASONS

In the preceding five chapters, I have formulated a responsibility-based account of action (understood as part of our conduct). I have argued that it correctly captures the distinction between actions and mere happenings. The object of the present chapter is to discharge my last commitment by showing how to conceive of the explanatory force of reasons.

Before going on, I should note that strictly speaking it is not part and parcel of the account developed in Chapters III-VI to include this discussion. The question how reasons relate to action is quite separate from the question what makes actions actions, with which I have been concerned so far.[1] In the preceding chapters, I have argued that in deciding whether a performance is an action or a mere happening, the agent need not be held to any actual normative expectations. In the present Chapter, I argue that the normative expectations to which the agent is actually held may help explain why the action was brought about. The question is worth addressing for at least three reasons.

---

[1] It is noteworthy that these questions tend to coincide on many accounts of action. For example, Davidson understands an action as a performance intentional under a description. At the same time, he takes it that a performance is intentional under a description just in case some reason that rationalizes the action under that description caused the action in the right way. This, in turn, means that a performance is intentional under a description just in case some reason explains the action under that description. The coincidence of the two questions depends on how one understands the notion of being intentional under a description. The questions are kept separate on Anscombe's account. This is because she proposes that a performance is intentional if a special "Why?" question applies to it (to which the appropriate answer is usually the reason for the action). In cases where the answer to the question involves citing the agent's reason for acting, the action will also be explained. However, there are cases where a special answer is given, viz. that there is no answer to the question, that there is no reason for which the agent acted, i.e. that the action cannot be explained in terms of reasons. It is those cases that show that the two issues are kept separate on Anscombe's account.

First, it is an issue that Davidson explicitly challenged the contextualists with.[2] Second, although the account developed so far does not immediately apply to the issues at hand, it does offer some vocabulary that is useful in handling the questions. Finally, the selectional account I propose will allow us to understand how it is possible for an agent to act on other people's wishes, and thus to further bring the view of explanatory nonindividualism out of the realm of the incoherent.

In section 1, I begin by explaining Davidson's challenge and briefly surveying some of the responses to it, thereby clarifying its nature. The opposition is crystallized between causal accounts, according to which the explanatory force of reasons must be conceived in causal terms, and various forms of teleological accounts of action explanation, which deny that this is the case. I will argue that rather than trying to understand the efficacy of reasons in causal terms we may try to understand it in terms of a selectional account. In section 2, I will describe some general features of selectional explanations and in particular Sober's useful distinction between selection for and selection of, which demonstrates that selectional explanations support the distinction between a selectional criterion being operative and it not being operative (but merely appearing as if it is) in the selection. In section 3, I develop the hypothesis that reasons can be conceived as selectional criteria by showing how one can account for the distinction between acting for and acting with reasons. To that extent, the causalist challenge is met. In section 5, I consider a way in which the causal theorist of action explanation might argue that despite the fact that one can account for the mentioned distinction without appealing to the idea that reasons are causes, there is still explanatory room left that can only be filled by that hypothesis. I argue that while there is still room for explanation, and while it can be filled by the hypothesis that reasons are causes, it can be filled by appeal to other explanations as well. In section 4, I show how the selectional account allows for the possibility of our acting on others' wishes.

It is worth emphasizing that the meaning of 'cause' is disputed. In the following considerations I will assume that the causal theorist of action explanation takes

---

[2] "Actions, Reasons, and Causes," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 3-19.

Davidson's physicalist understanding of the nature of causes.[3] Davidson's views are highly controversial on this score. In particular, it has been argued that by refraining to endorse Davidson's commitment to physicalism, one can avoid some problems that arise on his account.[4] It might be objected for this reason that by choosing an orthodox interpretation of the causal theorist's commitments, I am not being charitable enough toward the causal theory of action explanation. Such an objection would involve a misunderstanding of my goals in this chapter. I seek to understand a relation between reasons and actions, and I believe that the selectional account of that relation coincides with many of our intuitions. I have no qualms at all with a philosopher using the notion of "cause" as broadly as to encompass the selectional relation between reasons and causes I advocate. I do believe, however, that the selectional account illuminates that relation.

1. Davidson's Challenge

> Two ideas are built into the concept of acting on a reason…: the idea of cause and the idea of rationality. A reason is a rational cause.[5]

There is little dispute that the idea of acting on a reason involves the concept of rationality. To explain an action, it has been thought, is to rationalize it, to place it in the normative space of reasons the agent had when performing the action. In 1950s, the philosophical consensus was that the framework of reasons which is required to understand actions as actions rather than as motions of our bodies does not require us to understand reasons as causes. In fact, it was thought that the appeal to causality is misplaced, that there is no space for it in the hermetically closed framework of reasons. Since Davidson's breakthrough paper,[6] the consensus has changed to its exact opposite.

Davidson's argument for the causal theory of action explanation is simple. The question he asks is, What is the force of our ordinary action explanations? The force is

---

[3] Donald Davidson, "Causal Relations," in *Essays on Actions and Events, op. cit.*, pp. 149-162.
[4] John McDowell, "Functionalism and Anomalous Monism," in (eds.) Ernest LePore, Brian P. McLaughlin, *Actions and Events* (Oxford: Basil Blackwell, 1985), pp. 387-398.
[5] Donald Davidson, "Psychology as Philosophy," in *Essays on Actions and Events, op. cit.*, p. 233.
[6] "Actions, Reasons, and Causes," *op. cit.*

no doubt normative in that our ordinary rationalizing action explanations allow us to situate the agent's action in the space of the agent's reasons.[7] But this does not exhaust their force:

> A man driving an automobile raises his arm in order to signal. His intention, to signal, explains his action, raising his arm, by redescribing it as signalling. What is the pattern that explains the action? Is it the familiar pattern of an action done for a reason? Then it does indeed explain the action, but only because it assumes the relation of reason and action that we want to analyse. Or is the pattern rather this: the man is driving, he is approaching a turn; he knows he ought to signal; he knows how to signal, by raising his arm. And now, in this context, he raises his arm. Perhaps… if all this happens, he does signal. And the explanation would then be this; if, under these conditions, a man raises his arm, then he signals. The difficulty is, of course, that this explanation does not touch the question of why he raised his arm. He had a reason to raise his arm, but this has not been shown to be the reason why he did it. If the description 'signalling' explains his action by giving his reason, then signalling must be intentional; but, on the account just given, it may not be.[8]

Over and above telling us how an action was reasonable for the agent, Davidson argues, the explanation of the agent's action (as opposed to its mere rationalization) also points to the causes of the action, to what actually moved the agent. That this is so becomes apparent when we reflect on the fact that we intuitively allow the possibility of an agent having a reason, performing an action that is rationalized by that reason, and yet not performing the action because of that reason. This is a situation where the reason rationalizes but does not explain the agent's action. It appears that in order to account for the distinction we must appeal to some concepts beyond those available in the normative rationalizing framework. And Davidson believes that the concept of causality is the natural candidate. By construing the explaining reason as standing in not only a rational but also a causal relation to the action, the distinction between reasons that merely rationalize and those that in addition explain is captured.

Davidson's argument for the causal theory of action explanation can be summarized as follows:

---

[7] For ease and simplicity of writing, I will first of all tackle the question what it means for an agent to act on his own reasons. The account proposed will be general enough to encompass the case where the agent acts on another person's wishes, for instance. That it is I will demonstrate in section 4. Up until then, unless otherwise indicated, I will mention only the case where the individual acts for his own reasons.

[8] Ibid., pp. 10-11.

(1) Any theory of action explanation must account for the distinction between acting for reasons and acting while merely having reasons.

(2) No theory that appeals to concepts belonging just to the framework of reasons can account for that distinction.

(3) Only a causal theory of action explanation can account for the distinction.

Moreover, Davidson believes that the causal theory of action explanation implies that:

(4) Reasons are causes of actions.

Davidson's argument has been challenged in at least three ways. There have been attempts to deny (2), by showing that the distinction can be understood in terms of the rational force of reasons. It has been argued that in most cases we have no problem in identifying the reason for which the agent acts from the multiplicity of reasons the agent has, for reasons differ substantially from one another in the degree to which they are rational. For example, when a person abandons her family and friends, sells all her belongings and moves to Rangoon, we would reject the idea that her reason *for* doing so was the fact that she heard Rangoon is beautiful. This is not the sort of consideration for which she could have acted even though it might have been one of her reasons.[9] In other words, the thought is that the rational force of the reasons is sufficient to make the distinction between acting for and acting while merely having a reason. But the problem with such a response is that it is possible for agents to act *for* (not merely while having) bad reasons.

Some teleological theorists of action have denied (3). G.M. Wilson argued that teleological vocabulary is strong enough to support the distinction between acting for and acting with reasons. When we say that an agent acted *in order to* satisfy his desire, the statement does not leave it open for us to construe the action as being only rationalized by this desire. To say that an agent acted in order to satisfy his desire is to say that he acted because of it. Thus, contrary to Davidson's claim a teleological theory of action can meet

---

[9] This position is defended in Sergio Tenenbaum's *The Object of Reason: An Inquiry into the Possibility of Practical Reason.* Ph.D. Dissertation: University of Pittsburgh, 1996. The example is Tenenbaum's.

the challenge as well. The problem with this solution is that it does not seem very illuminating. One way of broaching the objection is to wonder whether any insight has been gained or whether this is not simply a way of restating the original challenge. Is not saying that the agent acted for a reason saying that the agent acted in order to satisfy the reason? Someone like Davidson would never doubt that the former entails the latter. But the question then is what is it to act in order to satisfy a reason. And it is here that Davidson offers an insight.[10]

One way of strengthening the causalist case (vis a vis the teleologist opponents) has been suggested by William Child. Child claims that Davidson's challenge-argument ought to be changed or at least supplemented. Not only ought one to require of any theory of action explanation that it give an account of the distinction in question, but also that it account for the fact that action explanations explain why the action occurred when it did.[11] He then argues that only a causal account of action explanation can meet the challenge. His argument is simple.

> Every event either has a cause or it does not. If it has a cause, then explaining why the event occurred must make reference to that cause. If it does not have a cause, then there is simply no explanation of why this particular event occurred when it did.[12]

---

[10] Wilson's counterargument is that Davidson's appeal to causality does not, contrary to appearances, explain what it is to act in order to satisfy a reason. In fact, a causal theorist of action reaches a dilemma. We can construe Davidson either as trying to explain what it means to act in order to satisfy a reason or as not trying to do that. If we take Davidson as undertaking the task of explaining what it means to act for a reason (contrary to the way Davidson seems to perceive his task at least since his "Freedom to Act," in *Essays on Actions and Events, op. cit.*, pp. 63-81), then Davidson does not give a very good theory of what it is to act in order to fulfill a reason because he needs to append the idea of causation by mental states with the unilluminating qualifier "in the right way." And if Davidson does not undertake the task of analyzing what it is to act in order to satisfy a reason then he should be the last to fault the teleological accounts for not analyzing it either. However, one might restate the challenge on behalf of Davidson. Although Davidson does not explain or aspire to explain what it means to act for a reason (to act in order to satisfy a reason), he gives and aspires to give an account of what underlies our disposition to describe some cases of acting in the context of a reason as acting for that reason (viz. when the reason causes the action) and others as acting while merely having the reason (viz. when the reason does not cause the action). It is at this point that Wilson seems to have to say that what underlies our disposition to describe some cases as actions for a reason and others as actions while merely having the reason is the *fact* that we recognize the former but not the latter as cases of actions in order to satisfy the reason. And this is hardly an account of the distinction.

[11] William Child, *Causality, Interpretation and the Mind* (Oxford: Clarendon Press, 1994), p. 92.

[12] Ibid., p. 92.

Since actions are events and as such have causes, an explanation why an action occurred when it did must "make reference" to the causes. Thus, if ordinary action explanations explain why an action occurred when it did, they must "make reference" to causes.[13]

Indeed, if we require that ordinary action explanations explain why an action occurred when it did then teleological accounts of action explanations will lose out. To say that the agent acted *in order to* further his desire is usually not to explain why the action occurred when it did. This point is admitted by von Wright who explicitly points out that his theory does *not* undertake the task of explaining why an action occurred when it did.[14] He construes ordinary action explanations as explaining the significance or the point of the action *given* that it occurred.

The problem with Child's rendition of "the basic argument" for the causal theory of action explanation is that it is not at all clear that our ordinary action explanations do indeed explain why the action occurred *when it did* — not in general, at least. Sometimes, they might. It might be that someone wagered to run around his house exactly when the town clock strikes twelve on a particular day. Then the explanation why he ran around the house at noon by appeal to his desire to win the wager does explain why the action occurred when it did rather than at some other time. But ordinarily this will not be the case. Peter may have plenty of reasons to finish his latest book (to get paid, to finally finish it as he is getting tired of writing it, etc.). But when he finally does it, none of the reasons are likely to illuminate why he has finished on Saturday, May 12 at 3pm.

---

[13] The reader will note that it does not follow from this argument that the reasons mentioned in the ordinary explanations of action are the causes to which the explanations must "make reference." It is also left very vague what exactly is required for an explanation to "make a reference" to a cause. In fact, Child abandons Davidson's thesis (4) that reasons are causes in favor of a weaker thesis that reasons explanations make reference to causes that are suitably related to reasons (e.g. the right kind of perceptual beliefs).

[14] "Von Wright's account gives no explanation of why the agent's behavior occurs or comes about, for the agent's intentions, beliefs, and desires are not here causes. An explanation of action, on [the] non-causal theory, gives the attitudinal conditions in terms of which to derive the understanding of the agent's behavior *as* the act that he performed — and that is sufficient to explain why the agent acted as he did — but it does not give the sufficient (causal) conditions of the occurrence of the behavior which is understood as action." (Frederick Stoutland, "The Causation of Behavior," in (ed.) Jaakko Hintikka, *Essays on Wittgenstein in Honor of G.H. von Wright* [Amsterdam: North-Holland, 1976], p. 302.)

Still one may feel that there is something to Child's suggestion. While it may not be necessary for a theory of action explanation to account for why an action occurred at a specific point in time, the theory ought to make this fact intelligible. In other words, rather than requiring that an action explanation explains why an action occurred exactly when it did, it ought to explain at least why it occurred within some reasonable limits of when it did.

This concludes a very brief survey of the main currents pulling in various directions in the issue at hand. The causal theory of action explanations has a genuine appeal. But whatever other reasons for it, the main one remains the challenge of accounting for the force that an explanatory appeal to reasons has. The argument for the causal theory of action has the form of a challenge. It will be my aim below to try to argue that there is a way of meeting the challenge that has been overlooked. Much of the appeal of the idea that action explanations are causal comes from the blanket-uses of the term 'cause'. I will try to show that a very special (though still causal in *some* sense) way of understanding the teleological relation characteristic of action can meet the Davidsonian challenge. Rather than trying to give an account of the teleological relation in a causal-intentional way, we can try to understand it in a selectional way. Rather than understanding reasons as causes, I will suggest that we can understand them as selectional criteria. Such an account will meet Davidson's challenge (section 3). It will meet the Child-Stoutland challenge (section 5). And it will satisfy the criterion of adequacy we imposed early on: of allowing us to understand nonintentional explanations of action (section 4).

2. Selectional Explanations

The thought that certain processes in the world are directed toward, or pulled toward, ends seems inescapable. There are two paradigmatic areas where teleological thinking found its most immediate application. The first domain was the organic world, the object of the study of biology. The second was the domain of human action. In both cases, explanations that appeal to goals are integral to our understanding of the phenomena; without them it would be seriously incomplete.

The general problem of teleological explanation, explanation in terms of ends, is that ends typically reside in the future. To the extent that we are accustomed to giving explanations in terms of (efficient) causes (where the paradigmatic idea is that of a push by the past rather than a pull from the future), the idea of a teleological explanation seems problematic. It looks like an action-at-a-(temporal)-distance. The future end cannot (efficiently) cause an action.

One (causal) solution to this problem has been to find some efficient cause that is suitably related to the future end. An intention, it has been claimed, is just such a state. It is not the end itself, but it reflects, represents or embodies the end of the action. This (let us call it "causalist") interpretation of teleological relations has found quite a comfortable niche in the second of the domains of teleological relations, human action. But it has also been proposed in the other domain of biological phenomena. Lamarck's model of evolution explains why organisms are so perfectly adapted to their environments by appealing to *striving* on the part of the organisms to achieve better adaptation. By striving to be better adapted, the organisms achieve better adaptation. The achievement of the purpose is causally mediated by states of the organism that represent it.

Lamarck's solution was an adaptation of the causalist interpretation of teleological relations in the domain of biology. It has been replaced with a different model of teleological relations which also relies on causal relations but quite different ones. The selectional interpretation of teleological relations has been proposed by Darwin to account for biological adaptation. The thought is simple. Darwin thought that the model on which ends are realized by appealing to causal states that represent or reflect ends must be rejected. The way in which ends are achieved is mediated by a special configuration of causal processes, but none of the processes themselves could be seen as representing or embodying the end. That the purpose is achieved is, as it were, an emergent outcome of the operation of a variety of causal processes. So, in Darwin's case, the purpose of better adaptation is achieved because those organisms that are less well adapted tend not to survive, not to pass on their genes to future generations. The purpose exerts its influence *not* by being embodied in the causal states of the individuals, but rather by being embodied (or distributed) in the pressures to which the individuals are

subject. The selectional model provides an alternative way of avoiding the problem that teleological relations involve an appeal to action at a distance. Rather than thinking about the purpose as embodied in the causal states of the individuals, it presents it as embodied in the selective pressures.

Another (teleological) attempt at resolving the problem relies on questioning our scientistic inheritance, the custom of explaining phenomena in terms of efficient causes. Such an account questions the very impulse for trying to conceive of the end as in any way related to the efficient cause. A teleologist does not deny that the phenomena have causal explanations but asserts that there are two kinds of explanations one can give: teleological and causal; and there is no reason to think that the former must be reducible to or less fundamental than the latter. (The Stoutland-Child objection shows the limits of such a position for the theory of action.)

In the next section, I will try to make more concrete the proposal to exploit this analogy in the understanding of the way in which reasons relate to actions. At present, I would like to make two general points about selectional explanations. First, I will coin some simple terminology for discussing the nature of selectional explanations in general. This is important because the model of natural selection is but one kind of selectional explanation and we need to have some concepts to understand selectional explanations in general. Second, I will introduce the distinction between selection-for and selection-of [15] which will be the seed from which the distinction between acting for and acting with reasons will be cultivated.

The selectional account employs the idea of selection at a very abstract level. I will not claim that there is an analogue of natural selection in the domain of human action. It is thus important to begin by casting the conceptual net wide enough so as to comprise a variety of selectional phenomena. Let us begin by mentioning examples of four selectional phenomena, and then identifying some elements common to them.

(a) Before industrialization, there coexisted two subspecies of moths: black and white. In birch forests, white subspecies dominated slightly; otherwise, the two

---

[15] Elliott Sober, *The Nature of Selection. Evolutionary Theory in Philosophical Focus* (Cambridge, MA: The MIT Press, 1984).

subspecies lived in an equilibrium. This changed with industrialization. The black moths began to dominate the white ones, even in birch forests. This was because pollution darkened the bark of the trees, making white moths more visible to predators. Since the black moths could breed more easily without suffering comparable losses to the predators, they began to dominate the gene pool.

(b) The characteristic features of the cauliflower (highly developed flower bracts), brussels sprouts (highly developed multiple offshoots), cabbage (highly developed leaf growth), etc., are the result of long-term *artificial selection*, which started with a single wild cabbage plant. Specimens with the desired characteristics were interbred. Their offspring was carefully sorted for the desired characteristics, and then subjected to further breeding. As a result the desired characteristics achieved the developed state we know from grocery stores.
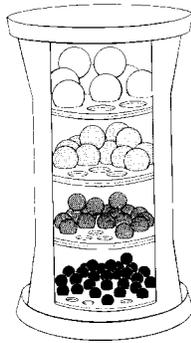
Figure 5. Sober's selection toy

(c) Sober describes a cylinder-shaped *selection toy* with four horizontal levels (Figure 5). Each level contains holes of the same size, but the holes on each level are larger than those on the level below. The toy is filled with balls of four sizes (equally sized balls are of the same color) such that the biggest (white) balls cannot pass through the biggest holes, the second biggest (yellow) balls can pass through the biggest holes but not through the second biggest holes, and so on. The shaking of the toy distributes the balls to their respective levels so that only the smallest (green) balls end up on the lowest level.

(d) All children in a room read at the third grade level. This is because individuals would not be admitted to the room unless they could read at the third grade level.[16]

All these examples involve some sort of selection. In cases (b) and (d) the selection is artificial, it is a matter of someone choosing or selecting certain objects. In cases (a) and (c), the selection is not a matter of someone choosing anything, and in this sense it is "natural." Cases (a) and (b) describe a *process* of selection (providing the breeders do not choose to stop in (b), in which case the organisms would be subject only to natural selection). After one cycle of selection, the organisms reproduce thus resupplying new organisms which in turn are subject to continued selection. This is not so in cases (c) and (d), where the selection is a one-time affair: once the individuals are selected, no new individuals are subject to the same selection.

Selectional explanations of phenomena rely on what one might figuratively call two focal points. One of the focal points is a "selecting mechanism" which selects the variety of individuals according to some selectional criterion. The second focal point is a "generating mechanism" which supplies individuals, objects, etc. on which the selection operates. Depending on whether the selectional phenomenon is a process or not, the mechanism continues to generate the individuals or objects. And so, (a) in the case of the process of natural selection, what generates the variety of organisms on which selection operates are mechanisms of genetic variation and reproduction. The selecting mechanism comprises the force of natural selection, which segregates organisms that are adapted to the environment (which live and reproduce) from those that are not adapted (which either die or do not reproduce). (b) In the case of artificial selection, what generates the variety of organisms on which selection operates are likewise genetic mechanisms and reproduction. The selecting mechanism, on the other hand, lies in the hands of the breeder. It is the breeder who chooses which organisms reproduce further. In cases (c) and (d), the individuals are not continuously generated. (c) In the case of the selection toy, all the individuals within the toy are subject to selection. The mechanism of selection (started by shaking the toy the right side up) consists of the levels of holes

---

[16] This is an example of Sober's.

letting the balls of appropriate size pass. (d) In the case of the selection of children, all the individuals who come are subject to selection. The selecting mechanism consists in a committee choosing the children with appropriate reading skills.

Another useful idea is the characterization of a selectional process as an adaptational process. One example of an adaptational selectional process is the process of natural selection. The element that is responsible for natural selection being an adaptational process is heritability of certain traits.[17] Simplifying, since the size of the neck in giraffe-ancestors was heritable, the fact that they were subject to selection for neck-size led to the increase of the proportion of longer-necked individuals in next generations.[18] It is this feature of the generating mechanism coming to coincide with the selecting mechanism that characterizes adaptational selection. In the limit, the generating mechanism produces mostly those individuals that the selecting mechanism would select. Natural selection with respect to the neck-size of giraffes has been adaptational in that throughout the incidence of giraffes with longer necks has increased in further generations. Nowadays, in fact, the adaptation is manifest in that the generating and the selecting mechanism coincide: only those giraffes with long necks are born, or in other words, only those giraffes that would be selected by the selecting mechanism (for neck-size) are supplied by the generating mechanism. Needless to say not all selection is adaptational, not even all natural selection. The property of not-having-lethal-mutations is selected for: organisms that have lethal mutations consistently die out. But the selection for not-having-lethal-mutations is not adaptational, since it is not heritable.

We can thus characterize the notion of a simple selectional system in functional terms as a system that comprises a number of individuals subject to a selectional mechanism (according to some selectional criterion). In addition, a simple selectional system has a generating mechanism, which enables the selection to be repeated on individuals generated.

---

[17] This is such an integral feature of natural selection that it is sometimes taken to be part of the very meaning of "selection." This is not the case, however, as the examples demonstrate.

[18] It is possible (in view of the possibility that some traits are recessive, e.g.) that at some point in the chain, the incidence of the selected trait in generation $k+1$ will be actually lower than in generation k, but in the

It is important to appreciate the fact that the characterization of something as a selectional system is functional. Here is an example where this becomes rather clear. Suppose that a child is very messy. He constantly throws toys on the floor. His parents spoil him continuously buying him new toys that the child throws on the floor. But the room is clean most of the time when the child is not currently throwing toys. This is because he has an aunt, who, obsessed with order, comes to his room and simply collects all the toys from the floor and throws them out. Here the generating system is the child throwing toys on the floor, and the selectional mechanism is his aunt's obsessively throwing away the toys from the floor.

One crucial fact about the process of selection has been emphasized by Elliot Sober and brought out in his distinction between *selection for* (properties) and *selection of* (objects).[19] The point of the distinction is that what is selected are objects, but they are selected according to a selectional criterion, i.e. insofar as they have certain properties. This is a crucial fact to appreciate about selection because it opens the door to the possibility of an object being selected according to one criterion while it appearing as if it could have been selected according to another. And it is this distinction that will allow us to understand the distinction between acting for a reason and acting with a reason in selectional terms.

Sober illustrates the distinction in terms of his selection toy. Recall that the toy is so constructed that all balls of the same size are also of the same color. The shaking of the toy results in the smallest sized green balls falling to the bottom of the toy. In such a case, it is true to say that the smallest balls are the objects that were selected, as it is equally true to say that the green balls are the objects that were selected. The concept of *selection of* objects is transparent. Not so for *selection for* properties. While it is true to say that smallness was the property selected for it is not equally true to say that greenness was the property selected for. Greenness was a "free-rider" as it were, the fact that green

---

long run, the incidence of the selected trait will increase. If the selection of no other traits competes with the given one, it will eventually dominate the whole population.

[19] Ibid., pp. 97-102.

balls were selected was coincidental, and was due to the distribution of properties among the objects that entered the process of selection.

This can be clearly seen by imagining an appropriate counterfactual situation. Suppose that the toy was filled with balls of varying sizes whose color was not so uniformly correlated with the size of the ball: e.g. if half the balls were green and half red, but each color had characterized different sizes of balls. In such a case, it would be true to say only that the smallest size balls were selected, not that green balls were selected (since among the green balls were balls of bigger size). In other words, the fact that in the actual case the smallest green balls were selected can be understood in terms of a "size-counterfactual": had the balls not been small, they would not have been selected. While the corresponding "color-counterfactual" is false: had the balls not been green they would not have been selected. Since, in the example, the former counterfactual rather than the latter is true, it was the size not the color criterion that was operative.[20]

Two points are crucial. First, selectional phenomena cover a much wider range than the process of natural selection, which is nowadays taken to be paradigmatic of such phenomena. Second, selectional explanations are capable of supporting the distinction between the operativeness of one selectional criterion and the operativeness of another. It is this feature of selectional explanations that will allow us to understand the distinction between the efficacy of one reason and the efficacy of another.

3. Reasons as Selectional Criteria

Let us now turn to the crucial question of understanding the distinction between acting for a reason and acting while merely having one. The question is this. Given what we know about the intuitive force we attach to action explanations (in particular, the fact

---

[20] This is a delicate point as there is no consensus on exactly how to understand the logic of counterfactuals. I use counterfactuals in a way that is supported by one side of the debate (see e.g. David Lewis, *Counterfactuals* [Oxford: Basil Blackwell, 1986]). In particular, I am assuming that the "size-counterfactual" and the "color-counterfactual" make the idea of selection-of and selection-for a little clearer. This does not mean that there might be theories of counterfactuals where this is not the case (see e.g. Michael J. Loux, ed., *The Possible and the Actual. Readings in the Metaphysics of Modality* [Ithaca, NY: Cornell University Press, 1979]). I should emphasize that I treat counterfactuals as a spring-off point. I will later dispense with them in favor of speaking of selection being caused by the agent's belief concerning what fulfills an expectation in question.

that we allow for the distinction between acting for and acting with reasons), how must we conceive of the relation between the agent, the reason, and the action in order for the conceived relation to be strong enough to capture the force of reason explanations? Davidson's answer to the question was to conceive of reasons as causally efficacious states of the agent. If such states do causally produce the action, they are the reasons for which (not merely with which) the agent acts. On Davidson's picture the reasons are conceived of as causally generating the actions that are rationalizable in view of such reasons.

In this section, I propose that we do not need to think about reasons as the generating causes of actions, not at any rate on the grounds given by Davidson. We shall see that if we think of the agent as a selectional system of sorts who selects her performances in accordance with her reasons (understood as selectional criteria), we can accommodate the distinction between acting for a reason and acting with a reason. This is the sole purpose of this section. Although we will see some considerations that would favor abandoning the thesis that all reasons are causes (section F), my goal is to suggest a way of applying the selectional metaphor to the case of agency and understanding the concept of acting for a reason accordingly. I will argue in section 5 that the account opens a way for the causal theorist of action explanation to still claim superiority for her account. But she will not be able to do so on the grounds that one cannot accommodate the distinction between acting for and with reasons otherwise.

I begin by suggesting in very broad strokes how one can conceive of the agent as a selectional system (section A). I then consider a preliminary example to consolidate some of our intuitions (section B). In section C, we will see more systematically how to make the distinction in two other examples. In sections D and E, I formulate the distinction between acting for reasons and acting while merely having reasons more systematically, listing and explaining certain constraints that are required. I end by suggesting some reasons for thinking that reasons might not be causes (section F).

A. An Agent as a Selectional System

The reason why the suggestion that an agent is some sort of a selectional system seems other-worldly is that we are by and large reliable in producing many bodily actions

that we intend to perform. We are by and large reliable in raising our arms, shaking our heads, walking, etc.[21] With respect to those types of actions, we are reliable in generating the performances that we would select as realizing our desires. It is also in cases where our reliability is disturbed that it becomes clearer how we could think of ourselves as selecting performances in accordance with our reasons. Particularly illuminating in this respect are (unintended) mistakes: pouring orange juice instead of water absent-mindedly, having something one did not want one's conversant to hear slip out, misreading a note, etc. In all these cases, the otherwise reliable agent generates an action she does not want to perform, an action that does not fit the selectional criterion.

I begin by clarifying the idea of what it means to say that an agent is reliable and distinguish two cases where she is not: she could be semi-reliable and anti-reliable. I then ask the question in which cases the concept of action finds application and conclude that it fails to apply only when the agent is anti-reliable. When the agent is semi-reliable, the idea of the agent as a selectional system becomes most clear. It will allow us to see the reliable agent as a special case of a selectional system.

There are performances with respect to which we are rather *reliable*. If I intend to raise my arm, I most likely will succeed in so doing since I am reliable in producing performances that realize such an intention. When I intend to raise my arm, I will raise an arm rather than a leg, I will raise an arm rather than sit motionless gaping at a screen. There are other performances with respect to which we are *semi-reliable*. Though by and large we would succeed in doing what we want to do, we would ordinarily not succeed on the first attempt. There is usually some slack: we produce two or three performances before the right kind of performance is produced. Occasionally, of course, we might produce the wanted performance right away, but not usually. Shooting baskets is something most of us are only semi-reliable at. Finally, there are types of actions with

---

[21] It might be noted that the fact that we are reliable with respect to many mundane bodily actions is relatively insignificant given the great number of types of actions with respect to which are not reliable (becoming rich, managing to be punctual, realizing political agendas, etc.). The reason why the fact that we are by and large reliable with respect to bodily actions is important for a causal theorist of Davidson's persuasion is that he treats all actions as identical to bodily movements. Davidson will then say that we are not generally reliable in seeing to it that our intentional bodily movements have the consequences that we intend them to have, but that we are nonetheless generally reliable in producing the bodily movements.

respect to which many of us are completely *anti-reliable*: we generally do not succeed in producing the wanted performance within a recognizable period of time. Juggling four balls is something most of us are anti-reliable at.

Most of us are reliable, semi-reliable and anti-reliable with respect to different performances. (Much of our social organization relies on that fact.) But it may pay to reflect on what would happen to the very concept of action if we were either exclusively anti-reliable or exclusively semi-reliable. It has been convincingly argued that our concept of action would not even get a grip if we were anti-reliable with respect to all kinds of performances. Nor indeed would the concept of having reasons. The very application of psychological vocabulary presupposes that our behaviors form certain relatively steady patterns. To suppose that an agent is anti-reliable with respect to all performance types, is to suppose that his behavior is not interpretable (in terms of reasons), and that no concept of action is applicable.[22]

But the situation is different if we suppose ourselves not to be anti-reliable but to be only semi-reliable. In such a case, the concept of action would still be applicable, the only difference is that our actions would stutter, as it were. The only addition that would have to be made is that there would have to be some element of recognition of the *right* performance, the performance that fits what we want, or in other words, the performance that is *selected* from among the other (unsuccessful) performances. For instance, the agent who produced a performance that did not fit what he wanted (the selectional criterion) could remark to someone or think to himself "This is not what I meant to do" or he could just try again, produce another performance until he reached the one that fit

---

[22] G.E.M. Anscombe, *Intention* (Ithaca: Cornell University Press, 1957); D. Davidson, *Essays on Actions and Events, op. cit.*; *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984); Daniel C. Dennett, *Brainstorms. Philosophical Essays on Mind and Psychology* (Cambridge, MA: Bradford Books, 1981); *The Intentional Stance* (Cambridge, MA.: Bradford Books, 1987).

what he intended[23] (the selectional criterion) in which case he could say or think "There" or simply stop producing further performances.[24]

The reader ought by now to have recognized that this is not just a wild philosophy-fiction story. There are many occasions where we are semi-reliable rather than reliable. Such occasions include certain kinds of activities. Sometimes we are prone to making mistakes of a certain sort (some people may be particularly prone to certain kinds of accidents: dropping glasses, stumbling over stones, forgetting appointments, etc.). Some sport disciplines where the best mark from a couple of trials counts seem to assume that we are semi-reliable in producing the performances. Many broadly conceived educational contexts allow for multiple trials. This includes activities such as paper writing, where the agent can write a number of versions of a paper and decides which one is "the paper he wrote" by submitting it. Before infants learn basic motor skills, they are at best semi-reliable in performing them.

A semi-reliable agent could be thought of as a selectional system. The agent generates performances of which he selects the performance that fits what he wants (the selectional criterion). A reliable agent likewise can be seen as a selectional system. But to say that he is a reliable agent is to say that he is disposed to generate performances that he is disposed to select. This means that a reliable agent produces only performances that he would select.

At this point, one might be inclined to wonder what causes the agent to generate the performances. But I will leave this question open for now and return to it later (section F). The point now is merely this: given the idea of an agent as a selectional system, where the notion of a reason functions as a selectional criterion for the action the agent is disposed to select from the performances he produces, we can account for the

---

[23] In Eastern-European languages, the non-Latin word for 'intention' is actually illuminating in this regard. The Polish 'zamiar' (intention) and 'zamierzyć' (to intend), for example, could be understood as composed of two component words: 'za' meaning 'for' and 'miara' meaning 'measure' ('mierzyć' meaning 'to measure'). In other words, if the lexicographic speculation is correct, 'to intend' would be understand as 'to take for a measure'.

[24] And it might be not only a matter of the agent's discretion to decide which of the performances fits the selectional criterion. For example, when a ballet teacher expects her pupils to perform a certain movement, she will in general be the authority on whether the movements the students perform fit the expectation. Of course, her authority is not absolute, she can be wrong. (I briefly discuss such cases in section 4.A.)

idea of a reason being operative in the action, abstracting from any mention of causes of actions. It may be true that actions do have causes, but they will not be invoked in laying out the distinction between acting for and acting with reasons. This means that even if reasons were causes, it suffices for our intuitive notion of acting for a reason that they be selectional criteria rather than causes. (In section 5, I consider further reasons why one might believe that the hypothesis that reasons are causes is necessary for our understanding of the explanatory force of reasons.)

In order to show that not too much is built into the idea of a selectional system already (I consider some further objections on this point later, see in particular section 6.A), it will pay to consider the following scenario. Suppose that an agent is reliable in responding to his intention to φ and in responding to his intention to ψ. Suppose that on an occasion he produces a performance that can be described as both his φing and his ψing. As a matter of fact, he also has reasons both to φ and to ψ. The fact that he produces the performance and so (since he is a reliable agent) selects it, does not tell us for what reason he acted. Likewise if he is a semi-reliable agent, and of three performances he finally selects one that is both a φing and ψing. We do not know for what reason he acted. This is because all we know immediately is that the agent selected a performance (in the sense of Sober's transparent "selection of"). What we do not know is for what reason he selected the performance (in the sense of "selection for").

This should make the idea that an agent can be thought of as a special selectional system a little clearer. It involves thinking of the agent as producing performances that are then selected according to whether they fit the reason or not. It does not explain what the causes of such performances are, nor how they are related to the selectional criteria.

B. A Preliminary Example

Let us switch gears a little bit and ask the question what would be required for an (ideal) interpreter to tell whether an agent, whom he conceives as this kind of selectional system, has acted for one reason rather than another while both rationalize the action.

Suppose that a gardener likes only white blooming plants, and that he also has a particular preference for miniature plants. In fact, the plants he cares for are all small — they do not exceed 2′. These preferences favor his garden containing only white-

blooming and only small plants. Suppose that the garden does in fact contain only small white-blooming plants and that the gardener is very diligent — whatever grows in the garden coincides with the gardener's preferences.[25] While this scenario suggests that the gardener constructed the garden to accord with both of his preferences, there remains the possibility that in fact only one of the preferences has been effective in the gardener's pruning and planting scheme, while the other preference is satisfied but coincidentally. It could be, for instance, that while the gardener likes white flowers and small plants, he shaped the garden exclusively to include small plants, and most of the ones that were available either do not flower at all or have white flowers, and the couple of small color-blooms he planted simply died. As a result, the garden coincidentally also fits his preference for white-blooming plants, though that preference was not efficacious in his constructing the garden.

But what exactly would it mean for the gardener to have followed one of the preferences rather than the other? What kind of knowledge would an ideal interpreter have to be equipped with in order to tell that the gardener followed his preference for small plants? One way of answering this question would be to consider what would happen in certain counterfactual situations.[26] As a matter of fact, the garden contains white flowers and small plants only. (Let us suppose for simplicity that only these two preferences are in play.) One sort of counterfactual situation (S–W) one would have to consider, is what the gardener would do if there were only small plants in his garden, but some of them had colorful blooms. The other sort of counterfactual situation (–SW) one would have to consider is what the gardener would do if there were only white blooming plants in his garden but some of them were rather tall.

The answer to the question which preference was efficacious depends on what the gardener would be disposed to do in those situations. If in situation (S–W) the gardener

---

[25] This corresponds to the assumption that the agent is reliable.

[26] The counterfactual situations correspond to those involved in the explanation of Sober's distinction between selection-of and selection-for (see p. 178, above). It is worthwhile noting that the counterfactuals are different from ones that would be involved if we were to suppose that the gardener's preferences were to be construed causally. For the counterfactual situations that would be relevant would consist in cases where the gardener has both preferences, has one preference but not the other, and has none of the

were disposed to prune all and only the plants that had colorful blooms, but would not be disposed to prune any plants in situation (–SW), given that only two preferences are in play, we should say that he follows the preference for white blooming plants.[27] If, on the other hand, in situation (S–W), he would leave the garden as is, but prune all tall plants in situation (–SW) then we ought to conclude that he follows the preference for small plants in constructing the garden.

What this example demonstrates is that we can answer the question what preference was efficacious in the gardener's construction of the garden as long as we know what the gardener would be disposed to do in relevant counterfactual situations.

C. Two Further Examples

One might worry that this simple suggestion works only for actions that could be intuitively construed as the agent involved in some kind of selective process — as the gardener is involved in positive selection of plants (by planting them) and negative selection of plants (by pruning them). Let us take an example similar to the one Davidson discussed in "Actions, Reasons, and Causes." (I should note that the case involves a simple bodily action, that of raising an arm. Since we are in general reliable in raising our arms, the agent's selecting of a performance is going to be identical with the agent's producing of the performance. I consider another example of an action we are semi-reliable at, later in the section.)

Let us suppose that a driver, as he approaches a turn where his friend happens to be passing by, raises his arm. He has two reasons to raise his arm. One of the reasons, *S,* is that he is approaching the turn, he wants to turn and *s*ignaling by raising his arm is a viable option to do so. The second reason, *G*, is that by raising his arm he will be *g*reeting his friend, which he wants to do also. Both reasons justify his own expectation

---

preferences. By contrast, all the counterfactual situations involved here presuppose that the agent has both preferences.

[27] Objection: The gardener could prune not only colorful plants but also some among the small ones if he considered them to be growing too profusely, for instance. Yet, this would be consistent with his following the preference for small plants. This is not a counterexample to the envisaged situation, however, since we are supposing that *only* two preferences are in play. This is a great simplification, of course. In reality, frequently many more reasons will be considered in play.

of himself to raise the arm. More precisely, the first reason justifies the expectation to signal a turn, the second reason justifies the expectation to greet a friend, but raising an arm (in the appropriate circumstances) constitutes a fulfillment of both expectations. On what expectation did the agent act, supposing that these were the only two reasons in play?

We need to consider appropriate counterfactual situations. The actual situation $[E^S E^G]$ is one where the driver's raising his arm fulfills both expectations justified by S and G. The three counterfactual situations will include: $[E^S\text{-}E^G]$ a situation where only the expectation to signal would be fulfilled by the agent's raising his arm (e.g., the agent nears the turn but his friend is not on the other side of the street), $[\text{-}E^S E^G]$ a situation where only the expectation to greet a friend would be fulfilled by the agent's raising his arm (e.g., the driver passes his friend long before the approach of the turn), $[\text{-}E^S\text{-}E^G]$ a situation where neither of expectations would be fulfilled by the agent's raising his arm (e.g. the driver does not either pass his friend or approach a turn).

We need to distinguish between the driver raising his arm (1) in order to signal the turn but not in order to greet a friend, (2) in order to greet a friend but not in order to signal the turn, (3) both in order to signal and in order to greet the friend, (4) just raising his arm for none of these reasons. Let us assume that the driver has and would have true beliefs in all these situations. We can also assume that the driver is reliable in responding to both expectations.

(1) If he were to raise his arm only in situations where he were approaching the turn $[E^S E^G]$ and $[E^S\text{-}E^G]$, but not otherwise, then we can say that he raised his arm to signal a turn. More precisely, if the driver would raise his arm whenever he were approaching a turn (even if he were not passing the friend $[E^S\text{-}E^G]$), but would not have raised it if he were not nearing the turn ($[\text{-}E^S E^G]$, $[\text{-}E^S\text{-}E^G]$), we can say that he acted *in order to signal the turn.*

(2) If the driver would raise his arm whenever he were passing his friend (even if he were not nearing a turn $[\text{-}E^S E^G]$), but would not have raised it if he were not passing the friend ($[E^S\text{-}E^G]$, $[\text{-}E^S\text{-}E^G]$), he raised his arm *in order to greet the friend.*

(3) To say that both reasons are operative in an action is actually to say either of two things. It could be that the agent acts *in order to either signal a turn or to greet a*

friend. In this case, he would raise his arm were he either nearing a turn or passing a friend ([-$E^S E^G$], [E^S-E^G]), but would not have raised it if he were neither passing a friend nor nearing a turn [-E^S-E^G]. It could be that the agent acts *in order to both signal a turn and to greet a friend*. In this case, the driver would raise his arm only if he were both nearing a turn and passing a friend [$E^S E^G$], but would not have raised his arm either if he were approaching a turn but not passing a friend [E^S-E^G], or if he were not approaching a turn though passing a friend [-$E^S E^G$], or neither [-E^S-E^G].

(4) If neither of these situations arises, the driver acts *neither to signal nor to greet a friend*. If the driver would still raise his arm even if he were neither passing a friend nor nearing a turn [-E^S-E^G], then he acted for neither of the reasons.

We can summarize this in the following table:

	S	G	S or G	S and G	neither S nor G			
[$E^S E^G$]	+	+	+	+	+	+	+	+
[-$E^S E^G$]	−	+	+	−	−	−	+	+
[E^S-E^G]	+	−	+	−	−	+	−	+
[-E^S-E^G]	−	−	−	−	+	+	+	+

Table 2. Patterns of action. '+' represents the agent's disposition to raise his arm, '−' the disposition not to raise the arm. See the text for the explanation of row assignments. In columns, the agent raised his arm in order to: *S* — signal, *G* — greet a friend, *S or G* —either signal or greet a friend, *S and G* — both signal and greet a friend, neither *S* nor *G* — neither signal nor greet a friend.

The above example is relatively straightforward for it involves a case of an action we are by and large reliable in performing (the raising of an arm). We should, however, consider another example of an agent producing a performance that he is only semi-reliable at.

Let us imagine that an actor stands in front of a mirror, rehearsing his part in an upcoming play. This involves his trying out a variety of face expressions. To simplify, let us suppose that a particular scene could call for either an expression of disdain or of terror. The actor toys with three interpretations of the character, on one — he should be disdainful (*D*), on the other — he should be terrified (*T*), finally — he should be perfectly ambiguous (*DT*). These constitute reasons he has to play the scene emphasizing disdain

or terror or both/neither. However, he has trouble in producing the right kinds of expressions at will: he is only semi-reliable in producing them.[28] His expressions "stutter" (he continues to produce further ones) until he is satisfied. For clarity, let us assume there are (relatively) overt signs of selection: when the actor is not satisfied, he thinks to himself "oh, no" and tries again; when he is satisfied with the performance, he thinks to himself "yes" and goes on to the next scene. It may be useful to classify the performances he produces into four categories: *dt* — performances that are ambiguous between disdain and terror, *d-t* — performances that display more disdain than terror, *-dt* — performances that display more terror than disdain, *-d-t* — performances that fail to display either terror or disdain (which includes erratic facial expressions as well as expressions of different emotions, surprise say).

Let us now suppose that he performs the following sequence: *-d-t, -d-t, d-t, d-t, dt, -dt*, where only the last performance is the one that is selected (only then does he think to himself "yes" and continues with the scene). Since all four possible types of performances are exemplified in the sequence, and only one is accepted, it is pretty clear what interpretation he opts for. He wants to emphasize terror (*T*): the reason that is operative, the reason why he produces the performance *-dt* is to emphasize terror. In fact, this is also the reason why he produces the whole sequence of (as they happened to be) unsuccessful attempts at emphasizing terror.

But it is, of course, possible that the sequence he produces will not allow us to clearly identify the reason for which he acted the way he did. Consider the following sequence: *-d-t, dt*. In other words, he selects the performance that is ambiguous between displaying terror and disdain. This is compatible with his acting either for *D*, or for *T*, or for *DT*, but not with his acting for neither. We can identify the reason that is operative by considering appropriate counterfactual situations, where we consider what would happen if the last performance *dt*, that has been actually selected by the agent, were different: if instead of *dt* the agent produced either of the three other performance types. [$E^D E^T$] The actual last performance satisfies the expectation justified by *D* as well as the expectation

---

[28] It does not matter for our purposes whether the actor can learn to make the expressions at will. It will matter for the play at least that he can learn to realize one of the three interpretations.

justified by T. $[E^D\text{-}E^T]$ One counterfactual situation to consider is whether the agent would select the performance if it expressed disdain but not terror (*d-t*), i.e. if it fulfilled the expectation justified by D but did not fulfill the expectation justified by T. $[\text{-}E^DE^T]$ Another is to consider whether the agent would select the performance if it expressed terror but not disdain (*-dt*), i.e. if it fulfilled the expectation justified by T but frustrated the expectation justified by D. $[\text{-}E^D\text{-}E^T]$ The final question is what the agent would do if the performance frustrated both expectations (performance of type *-d-t*).[29]

We can now see how we can determine what reason the agent acted for. We would say that he brought about the performance *dt* to realize interpretation D, if he would select his performance were it to fulfill the expectation justified by D (i.e. *dt* or *d-t*) but he would not select the performance were it to frustrate this expectation (i.e. *-dt* or *-d-t*). In such a case, if he produced *-d-t or -dt*, he would continue producing further performances until he managed either *dt* or *d-t*. Similarly, he acted (*dt*) to realize interpretation T, if he would select his performance were it to fulfill the expectation justified by T (i.e. *dt* or *-dt*) but not otherwise (i.e. *d-t* or *-d-t*). Finally, he acted to realize the third interpretation, if he would select his performance were it to fulfill both expectations justified by T and D (i.e. *dt*) but not otherwise (i.e. *d-t, -dt* or *-d-t*).

D. Acting for a Reason

The core of the idea of an agent acting for a reason R can be captured rather simply: Agent α φs for reason R just in case α φs (where his φing fulfills a normative expectation justified by R, E^R) and α would have φed if his φing were to fulfill E^R but α would not have φed were his φing to frustrate E^R. A driver raises his arm to signal a turn just in case he raises his arm (thereby fulfilling an expectation to signal a turn) and he would raise his arm as long as his raising his arm would fulfill the expectation to signal a turn, but would not have raised it if the expectation to signal a turn would be frustrated by his raising the arm (for example, if he were not approaching the turn). The actor grimaces to convey an expression of terror just in case he would select his performance

---

[29] Note that in order to be interpretable, in case the agent were to produce a performance of type *-d-t*, he would have to produce another performance.

were it to fulfill the expectation to express terror (i.e. *dt* or *-dt*) but not otherwise (i.e. *d-t* or *-d-t*).

We need to be a little more systematic. Let us first observe that we may want to understand two concepts. First, we may want to understand what it means to say that an agent acts because she expects something of herself (in section 4, I explain what it means to say that the agent acts because someone else expects something of her). Second, we may want to understand what it means to say that an agent acts because of a reason. I treat both concepts as being related. To say that a person acts because of a reason is to say that she acts on an expectation that is justified by that reason. When the driver raises his arm in order to signal a turn (for that reason), he acts because he expects of himself that he signal the turn.[30] When the actor grimaces in order to express terror, he acts because he expects himself to produce a performance that realizes interpretation *T* of the character he plays.

What does it mean to say that an agent acts on an expectation of himself more generally? Let us first assume that (r) the agent is reliable in responding to the expectation by fulfilling it, and that (t) the agent has true beliefs regarding what performances fulfill or frustrate the expectation. We can say that an agent α ϕs because he expects of himself that he ψ just in case (a) α actually expects of himself that he ψ, (b) α ϕs and his ϕing fulfills the expectation to ψ, (c) α would have ϕed had his ϕing fulfilled his expectation to ψ but α would not have ϕed had his ϕing frustrated the expectation to ψ. Thus, assuming that (a) the driver expects of himself to signal a turn, and that (b) he raises his arm thereby fulfilling the expectation to signal a turn, and that (r) the driver is reliable in responding to the expectation to signal a turn by signaling a turn, and that (t) the driver knows when he signals a turn, we can say that he raised his arm because he expected of himself that he signal a turn just in case he would have raised

---

[30] The fact that the concept of acting for a reason is first cashed out in terms of the concept of acting on an expectation explains one peculiar fact about the concept of acting for a reason. The way in which we use the concept of 'having a reason' frequently obfuscates an issue that is important to the debate between individualism and nonindividualism. For given that we know that a teacher wants a student to write a paper, we will infer that *the student has a reason* to write a paper. We draw the same conclusion if we know that the student wants to write a paper. The student's having a reason does not yet settle *who* expects of the student that he write the paper, and *whose* desire justifies the expectation.

his arm had his raising his arm fulfilled the expectation to signal a turn but he would not have raised it had it frustrated the expectation to signal a turn.

Let us generalize further by removing the simplifying assumptions. If we do not assume (t) that the agent has true beliefs concerning what performances fulfill or frustrate the expectation in question, then we need to consider the agent's beliefs concerning what fulfills or frustrates the expectations at hand. Suppose that John expects himself to do something to wake himself up, and that he makes some coffee. What would it mean to say that John made the coffee because of the expectation? We can say that John acted on his expectation just in case he would have made the coffee if *he believed* that it would fulfill his expectation to help him stay awake, but he would not have made the coffee if he believed that it would frustrate the expectation (if he believed that yet another cup would make him drowsy).[31]

If we assume (r) that the agent is not reliable but only semi-reliable in fulfilling the expectation, then rather than considering what performance the agent would produce in appropriate counterfactual situations, we must consider which of the performances the agent might produce he would *select*. We said that the actor who produces a grimace that expresses terror and disdain at the same time does so in order to express terror just in case he would have selected a grimace just in case it expressed terror but he would not have selected one that did not express terror.

We thus arrive at the more general formulation:

(E) An agent α φs because of his expectation of himself that he ψ just in case (a) α actually expects of himself that he ψ, (b) α selects his performance *p* of φing, *p* fulfills the expectation to ψ, and he believes that *p* fulfills the expectation to ψ, (c) α would have selected *p* had he believed that it fulfilled his expectation to ψ but α would not have selected *p* had he believed that it would frustrate his expectation to ψ.

---

[31] For the purposes of this dissertation I am going to settle on the belief-talk since my primary dispute is with individualism understood as requiring that the agent act on her pro-attitudes. There is a potential, however, for developing the account along externalist lines suggested by Rowland Stout, *Things that Happen because They Should. A Teleological Approach to Action* (Oxford: Clarendon Press, 1996).

Note that what is required by clause (b) is not only that the agent φ but that he actually select his φing. The concept of acting on an expectation (and the concept of acting for a reason) applies only to cases of actions that are selected or recognized by the agent as being appropriate. The point is brought out when the agent is semi-reliable with respect to a performance type. Just as it would make no sense to ask why (for what reason) a stutterer uttered the first syllable in a stuttering sequence, so it does not make sense to ask why (for what reason) the semi-reliable agent produced a performance that he did not select.

I have already spent a lot of time justifying the specifically selectional clause (c). At this stage, let me only point out that it amounts to the thought that the agent selects the performance *because* he believes it fulfills the expectation to φ. In other words:

(E) An agent α φs on his expectation of himself that he ψ just in case (a) α actually expects of himself that he ψ, (b) α selects his performance *p* of φing, *p* fulfills the expectation to ψ, and he believes that *p* fulfills the expectation to ψ, (c) α selects *p* because he believes that it fulfills his expectation to ψ.

The force of the 'because' here can be supposed to be causal. I will return to this point later. At present, let me note that (i) the beliefs cause the selection of the action, not the action (see Figure 6), and that (ii) the content of the beliefs is rather peculiar (they are not the beliefs that are frequently cited in the rationalization of the action).

It is important to require that the agent *actually* hold himself to the expectation (clause (a)).[32] If the agent did not, we could not say that α acted because of his expectation of himself. Rather, we would have to interpret it as a case of α's acting because α *thought* that he expected it of himself.

Finally, let me note that there are three ways of construing the belief cited in clause (c). On one interpretation (enforced by assumption (a)), (EB), the belief presupposes that α actually holds himself to the expectation. On a weaker interpretation

---

[32] This point will become particularly clear when we consider the possibility of our acting on other's expectations of us.

(-EB), the belief in (c) would only presuppose that α believe that α holds himself to the expectation. On the weakest interpretation (-E-B), the truth of the belief in (c) would presuppose neither. It is implausible to suppose that the belief ought to be construed in the strongest way (EB). Indeed, this is why (a) is needed as a separate clause. It is also implausible to construe the belief in the weakest way (-E-B), as not even presupposing that α believes that he holds himself to the expectation. If the belief is shorn even of this presupposition then one might argue that for any person ξ, whenever α believes that a performance fulfills his own expectation of himself to ψ, α also believes that it fulfills ξ's expectation of him to ψ. The driver who believes that his raising his arm fulfills his own expectation of himself that he signal a turn also believes that it fulfills President Clinton's expectation of him that he signal a turn. One might perhaps distinguish two beliefs here. First, the driver may believe that the performance *fulfills* President Clinton's expectation of him to signal a turn. This belief presupposes that the driver believes that President Clinton expects of him that he signal a turn (-EB). Second, the driver may believe that the performance *would fulfill* President Clinton's expectation of him to signal a turn, were he to be held to the expectation by the President. This belief no longer presupposes that the driver believes that the President holds him to the expectation (-E-B).

This indicates that α's belief that his ψing would fulfill his expectation of himself, as the phrase is used in the above characterization, implies at least that α believes that α holds himself to the expectation. It is the former belief (-EB) that is intended in clause (c). We shall see that this will become an important point in our discussion below (section 4.C).

Given the above understanding of what it means to say that an agent acts on an expectation of himself, we can understand what it means to say that an agent acts for a reason.

(R) α φs for reason R (in order to satisfy R) just in case there is some expectation justified by R, E^R, and α φs because of E^R.

A given reason can potentially justify a number of expectations. The fact that one is nearing a turn and wants to signal it might justify a number of expectations: to signal the turn, to signal the turn by raising one's arm, to signal the turn by putting a blinker on, etc.

Our characterization requires only that there be an expectation that is justified by the reason and on which the agent acts.

There is, however, one constraint that the expectation justified by R ought to meet:

(C) The class of all φings is not a subset of the class of fulfillments of the expectation E^R.

This means that the class of all φings is either a superset of the set of fulfillments of E^R or they partially overlap. Consider the driver who raises his arm to signal a turn. Constraint (C) amounts to requiring that it not be the case that all raisings of an arm are signalings of a turn. And this is certainly true. When one raises an arm in a restaurant one does not signal a turn.

The significance of (C) can be brought out by considering three claims:

(1) I raised my arm in order to signal a turn.

(2) I said "hello" in order to speak.

(3) I spoke in order to say "hello."

Of the three only (2) seems intuitively awkward. On hearing (2), we would be prepared to reinterpret what the agent means by saying "in order to speak." We might interpret (2) as the agent announcing that he said "hello" in order to say something in a crowd, or be noticed there. But it seems very awkward to think that the agent said something in order to just speak.

Constraint (C), coupled with our characterization of what it is to act for a reason, allows us to understand the awkwardness of (2). Let us think of all the above claims as having the form: "I φed in order to satisfy E^R," or "I φed because of R." In such a case, we can clearly see that constraint (C) is satisfied in (1). The class of arm raisings and the class of turn signalings partially overlap, so it is not the case that arm raisings are a subset of turn signalings. It is also satisfied in (3). It is not the case that acts of speaking form a subset of "hello" sayings, for the converse is true: "hello" sayings form a subset of acts of speaking. This also means that (C) is violated in (2).

Consider the way in which the violation of (C) in (2) but not (3) affects the application of our characterization. In (3), both counterfactual clauses of (R) are satisfied

non-emptily. In (2), they are not. In general, since φings are a subset of the fulfillment conditions of E^R, to say that α would not have φed if his φing were to frustrate E^R is trivially true, because his φing cannot frustrate E^R. In terms of (3), it is trivially true that α would not have said "hello" if his saying "hello" were to frustrate the expectation to speak because his saying "hello" cannot frustrate the expectation to speak.

The fact that we recognize claim (2) as awkward at least confirms that we intuitively accept a constraint like (C) on our idea of what it is to act in order to satisfy a reason.[33] Moreover, the direction in which (2) is likely to be reinterpreted is also telling in this respect. By interpreting what the agent means by 'speaking' in (2) as something to the effect of being noticed, we satisfy (C). For it is no longer guaranteed that when one says "hello" one will be noticed. So the idea that one can say "hello" in order to be noticed makes sense.

E. Acting for One Reason rather than Another

So far we have considered the idea of acting for a reason. We saw, however, that the contrast between the idea of acting for a reason and acting while merely having a reason is most vivid when an agent has two reasons but acts only on one. As before, the fundamental distinction is between acting because of one expectation and not because of another.

Let us assume that an agent selects his φing, and that his φing fulfills exactly[34] two expectations (to ψ and to ρ) to which he actually holds himself. We can determine which of the two expectations is operative. In all the cases, we are assuming that (a) α

---

[33] One might object here by appealing to a case on which J. Hornsby has put much emphasis, though her concern is very different (*Actions* [London: Routledge & Kegan Paul, 1980]). She imagines a case of an agent who tries to flex a particular muscle in his arm by clenching his fist. It seems intuitive to say in such a case that the agent clenches his fist in order to flex the muscle. Constraint (C) might appear to be violated in such a case since the class of successful muscle flexings is much wider than the class of successful fist clenchings. But a reflection shows that this is not the case in Hornsby's example. The agent clenches his fist in order to flex a very particular muscle. Not every fist clenching would result in his flexing this particular muscle. Hence (C) is satisfied.

[34] If the agent has more than two reasons, corresponding constraints would have to be added.

holds himself to the expectation to ψ ($E_\psi{}^{35}$) and to the expectation to ρ (E_ρ), that (b) α selects his performance p of ϕing, that p fulfills both expectations and that he believes that p fulfills both expectations. The following formulations give a causal rendition of clauses (c) (the counterfactual renditions of the clauses are listed in respective footnotes):

($E_\psi e_\rho$) α ϕs   because α expects of himself that he ψ rather than because α expects of himself that he ρ just in case (c) α selected p because α believed that it fulfills E_ψ and not because α believed it fulfills E_ρ.[36]

($e_\psi E_\rho$) α ϕs   because α expects of himself that he ρ rather than because α expects of himself that he ψ just in case (c) α selected p because α believed that it fulfills E_ρ and not because α believed it fulfills E_ψ.[37]

($E_\psi \vee E_\rho$) α ϕs because α expects of himself either that he ψ or that he ρ just in case (c) α selected p because α believed that it fulfills either E_ψ or E_ρ.[38]

($E_\psi \& E_\rho$) α ϕs because α expects of himself both that he ψ and that he ρ just in case (c) α selected p because he believed that it fulfills both E_ψ and E_ρ.[39]

---

[35] A minor notational point. I use a superscript following the shorthand for an expectation 'E' to indicate a that an expectation is justified by a reason, whose name appears in the superscript. By contrast, the content of the expectation appears in the subscript.

[36] α ϕs   because α expects of himself that he ψ rather than because α expects of himself that he ρ just in case (c) α would have selected p if α believed that it fulfills E_ψ (even if α believed that it frustrates E_ρ) but α would not have selected p were α to believe that it frustrates E_ψ (whether or not he believed that it fulfills E_ρ).

[37] α ϕs   because α expects of himself that he ρ rather than because α expects of himself that he ψ just in case (c) α would have selected p if α believed that it fulfills E_ρ (even if α believed that it frustrates E_ψ) but α would not have selected p were α to believe that it frustrates E_ρ (whether or not he believed that it fulfills E_ψ).

[38] α ϕs because α expects of himself either that he ϕ or that he ψ just in case (c) α would have selected p if α believed that it fulfills either E_ψ or E_ρ and he would not have selected p were α to believe that it frustrates both E_ψ and E_ρ.

[39] α ϕs because α expects of himself both that he ϕ and that he ψ just in case (c) α would have selected p if α believed that it fulfills both E_ψ and E_ρ and he would not have selected p were α to believe that it frustrates either E_ψ or E_ρ.

($e_\psi e_\rho$) α φs neither because he expects of himself that he ψ nor because he

expects of himself that he ρ just in case (c) α did not select p either

because he believed that it fulfills E_ψ or because he believed that it

fulfills E_ρ.[40]

We can use the recipe suggested in the previous section to obtain the corresponding

notions of acting for one reason rather than another, acting in order to satisfy either one

or the other reason, acting to satisfy both reasons and acting for none of the two reasons.

Let me illustrate on the example of acting for one reason rather than another.

($R_1 r_2$) α φs for reason R_1 while merely having R_2 just in case there is a

normative expectation justified by R_1, E^{R1}, and a normative

expectation justified by R_2, E^{R2}, and α φs because of E^{R1} not

because of E^{R2}.

F. Reasons as Selectional Criteria Rather than Generating Causes?

Reasons here (or more precisely, normative expectations that are supported or

justified by reasons) are conceived of not as generating causes[41] of the particular actions

but rather as criteria by which the actions are selected, as it were. The sense in which this

is a selectional model is extremely abstract. I am not postulating that there is anything

like an on-going *process* of selection as there is in the case of natural selection. A model

that is closer to what is meant is Sober's selection toy, except that what corresponds to

the balls in that case is here replaced by actual and possible actions. But it should be

born in mind that the idea of the agent's selecting performances does acquire some

substantiation when the agent is not completely reliable in generating the performances

he expects of himself, in generating the performances he selects.

It might be helpful at this point to contrast the proposal that reasons are

selectional criteria with the proposal that reasons are generating causes. In every case of

---

[40] α φs neither because he expects of himself that he ψ nor because he expects of himself that he ρ just in case (c) α would have selected p even if α believed that it frustrates both E_ψ and E_ρ.

[41] I use the phrase 'generating cause' to emphasize that I will leave room to claim that there is a sense of 'cause' that is compatible with the selectional account I am proposing.

an action done for a reason, there is going to be a normative expectation supported by that reason in order to fulfill which the agent acts, in the sense explained above of selecting the performance because the agent believes it fulfills an expectation. But, in every case of an action for a reason, there is going to be a generating cause of that action. In fact, there are likely to be many generating causes that come together to effect the action in question. Davidson's argument is that we *must* identify the generating cause of the action done for a reason with that reason. The proposal I have delineated diffuses the force of Davidson's argument. We *do not need* to identify the generating cause of the action with the reason for which the agent acts on the grounds that we could not otherwise account for the distinction between acting for and acting with a reason. I have proposed an account of that distinction without relying on any ideas concerning what causally generates the performance that is selected. Davidson's intuition is vindicated to the extent that some causal relation enters into the picture (viz. the belief causing the selection). But Davidson was wrong in supposing that the intelligibility of the distinction requires us to postulate that reasons are the generating causes of actions. (In section 5, I discuss another argument designed to show that reasons must be conceived as generating causes.)

One can assert the claim that reasons are selectional criteria with varying degrees of force. (i) One could claim that reasons are always selectional criteria and ought never to be identified as generating causes of actions. (ii) One could claim that reasons are always selectional criteria but could sometimes be identified as generating causes of actions. (iii) One could claim that reasons are always selectional criteria and always also generating causes of actions but that their explanatory power exhibited in ordinary action explanations relies on their being selectional criteria rather than on their being generating causes. Of the three positions, I take (ii) to be most plausible. (i) is too strong. It would be particularly implausible in view of visceral desires, like hunger and thirst, which are most naturally identified as generating causes of actions that lead to their satisfaction. But (iii) is also too strong.[42] Here are a couple of considerations that support the intuition

---

[42] (iii) is too strong if we think of causes as generating causes. At the end of the section, I allow that (iii) is an acceptable position for some uses of 'cause'.

that not all reasons are generating causes. In view of their intuitive plausibility, the defender of the thesis that all reasons are generating causes would have to produce an argument to the contrary.

Davidson's argument, which was designed to establish that all reasons are the generating causes of actions, fails. In section 5, I disarm another argument to that effect. One might insist on behalf of the causalist picture that the idea of reasons as generating causes is just natural and that the burden of proof lies with those who aim to deny that reasons are generating causes of actions. But this position is questionable in at least two ways. First, it is undeniable that the causalist picture is considered to be natural among philosophers nowadays. But this was not the case in the 1950s before Davidson's famous article. Second, it has been argued that our practices do not in fact support the picture that reasons are generating causes of actions. I want to briefly examine Child's rendition of this argument. I will show that while Child's position is too strong, the cases under consideration make it plausible to suppose that it is natural to identify some (not necessarily all) reasons with generating causes of actions.

It has been observed that frequently when we identify generating causes of actions, we do not identify them with reasons for which the actions were undertaken, but rather with more proximal events such as perceptual beliefs.[43]

> You ask me to pass the salt and I pass it, responding to your request, automatically as it were. This is an intentional act, though if 'intention' means anything like a state of mind, then I had no intention to pass the salt before I passed it; it went too quickly for that. Yet there was an intention embedded in that act (perhaps a belief too), the intention that the salt get to you in response to your request, an intention that could come before my mind only after I passed the salt, and which was not therefore a cause.[44]

Stoutland's point is that in such cases, we have a good folk-psychological understanding of the generating mental cause: the agent's hearing the request for salt. But this does not mean that we have equally good reasons for thinking that the agent's reasons for

---

[43] The point here is not that it would be impossible in such a case to identify the generating cause with the reason as well. The point is only that we as a matter of fact *do not* make the identification. The former claim would support a view like Child's that one could not understand our practices as being committed to the thought that reasons are generating causes. The second claim only supports the weak position that it is natural to suppose that we *do not* (not that we could not) identify all generating causes with reasons.

performing the action are to be identified with the generating cause of the action. Perhaps the reason why the agent passed the salt is that he is polite and appreciates politeness in general. But do we have good reasons for identifying the event that caused his action, which we already identified as his hearing the request, in addition to his wanting to be (or perhaps with his being) polite?

Too strong an answer to this question has been endorsed by William Child.[45] He claims that such events cannot plausibly be identified with the agent's reasons for acting. The reason why this may appear to be a conclusion that we are to draw from such cases is that it is implausible to identify the perceptual belief (as a type) with a reason (as a type). It need not be equally implausible, however, to identify the perceptual belief (as a token) with the reason (as a token) on a particular occasion. And it is only the latter that the causalist needs.

This demonstrates that the conclusion Child draws from such cases (viz. that not all reasons are causes) is too strong. It is sufficient for my purposes to draw a much weaker conclusion from such cases. They do not *demonstrate* that not all reasons are causes, for they do not demonstrate that the generating causes could not also (on a token by token basis) be identified with reasons. The cases merely register the natural ways for us to describe them. Accordingly, at most the cases *render it natural* (in absence of arguments to the contrary) to suppose that not all reasons are causes.

If so, however, then it is not clear that the causalist can simply resort to the thought that because the causal picture of reasons is so natural, the burden of proof lies with the challenger. For the picture according to which not all reasons are generating causes has also claims to being a natural picture. One reason in fact why one may think more of the burden lies with the causal picture lies in the fact that the claim defended by it (viz. that *all* reasons are generating causes) is so strong.

In sum, the fact that we frequently identify generating causes of actions with perceptual beliefs and not with reasons for action makes natural the view that reasons are not always generating causes of actions. It is important not to mistake the force of this

---

[44] F. Stoutland, "The Causation of Behavior," *op. cit.*, p. 319.
[45] W. Child, *Causality, Interpretation and the Mind, op. cit.*, p. 124.

claim for what it is not. I do not claim that this provides a conclusive reason to reject the thesis that reasons are always generating causes (this is Child's claim). My claim is much weaker, viz. that the fact that we do not frequently identify the generating causes of actions with reasons (but do identify them with other mental causes) constitutes a weak support for (in the sense of rendering it natural to hold) the view that not all generating causes of actions are reasons.

This leads to a worry. One might be concerned that I have focused all the attention on the selectional mechanism but left the generating mechanism completely in the dark. In fact, one could insist that we should have some intuitive conception about the generating mechanism, what causes the actions, if we then want to think about reasons as selectional criteria. But surely the most natural way of thinking about the generating mechanism is in terms of reasons causing us to produce actions. The above considerations actually show that this objection fails. We frequently think of perceptual beliefs as causing (i.e. generating) the actions. But we also have a more general, though blanket, way of thinking about the generating mechanism.

We sometimes think of *the agent* as causing, generating or producing the actions. In fact, one could understand the appeal behind the agent-causality theories of action as lying precisely in the fact that they exploit this intuition. The idea of agent-causality has been employed in a questionable fashion to capture the distinction between actions and mere happenings.[46] But in light of our account, the idea of agent-causality is actually illuminating. It is not that there is some special sort of causation involved. Rather, the idea of the agent causing the actions conveniently covers the details of the generating mechanisms, thus leaving the force of the agentive function to the selecting mechanism. The idea of agent causality provides a blanket conception for the generating mechanism, allowing us to avert our attention from the causes of actions toward the ends in accordance with which the actions are selected.

---

[46] Roderick M. Chisholm, *Person and Object. A Metaphysical Study* (La Salle, IL: Open Court, 1976); Richard Taylor, *Metaphysics* (Englewood Cliffs, NJ: Prentice-Hall, 1983). For criticism, see Donald Davidson, "Agency," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 43-61.

In conclusion, the primary commitment of the selectional account lies in insisting that the identification of reasons with generating causes is not necessary to make the distinction between acting for and acting with reasons. This leaves it an open question whether we should nonetheless identify all reasons, some reasons or no reasons as the generating causes of actions. This is something that a selectional theorist need not be committed on at all. In this section, I have indicated some reasons to believe that it would be natural to think that at least some reasons are the generating causes of actions. I have not, however, presented any conclusive arguments against the suggestion that reasons are always generating causes. Neither, however, have we seen any conclusive arguments for the suggestion that reasons are always generating causes.

G. Summary

It is possible to take the arguments advanced in this section in too strong a fashion. It is not my intention to try to claim that once we have the idea of the agent as a selectional system, and of reasons as selectional criteria, the idea of causality vanishes from the picture altogether. But it depends on what idea of causality we have in mind.[47]

It is important to stress that the picture that arises employs or presupposes the idea of causal processes in at least two ways (Figure 6). First, I do not deny that the generated performances that the agent selects are caused. In fact, the selectional account is uncommitted as to the nature of the causes. I do not believe, however, that we have sufficient reasons to think that the performances are always generated by the reasons for which the agent acts. (I consider and challenge further arguments to this effect in section 5.) Second, I also do not deny that the very selection of a performance is caused, and I have suggested that it is most natural to think of it being caused by the agent's belief concerning what performances fulfill or frustrate the expectation justified by a given

---

[47] Davidson's physicalist notion of cause has been opposed by J. McDowell, "Functionalism and Anomalous Monism," *op. cit.* In a similar vein, Jennifer Hornsby has argued that the identification of 'cause' of an action at a personal level is not going to be illuminated in any way by the identification of causes of the action at the physical level ("Which Physical Events are Mental Events," *Proceedings of the Aristotelian Society* 81, 1980-1, 73-92; "Agency and Causal Explanation," in (eds.) John Heil, Alfred Mele, *Mental Causation* [Oxford: Clarendon Press, 1993], pp. 161-188). Yet another view in the vicinity has

reason. Still what is missing from the picture is the thought that the reason or the expectation justified by the reason causes the action.
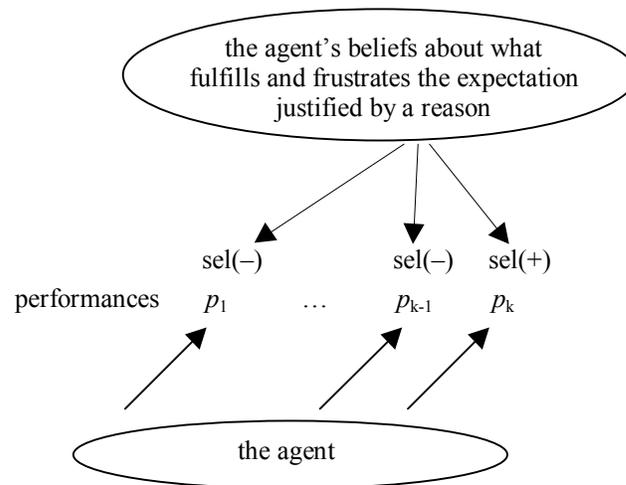

Figure 6. Causal relations involved in the agent's acting for a reason on the selectional model. 'Sel(+)' and 'Sel(−)' mark positive and negative selection, respectively.

I want to leave open the possibility that one could understand the selectional mechanism described as capturing a way in which reasons "cause" actions. One may want to say that what we mean when we say that the reason "causes" the action is precisely that the agent selects the performance because he believes that it fits the reason. If one understands the idea that reasons are "causes" in this sense, then there is no competition between the causal and the selectional account. But this will be because the selectional account illuminates the sort of "cause" that is at work.

4. Explanatory Nonindividualism Again

In Chapter I, I have argued that there are no conclusive reasons for holding explanatory individualism. I have indicated that the selectional account of the explanatory force of reasons will allow us to capture the idea that it is possible for an

been defended by Child (*Causality, Interpretation and the Mind, op. cit.*) who argues that while reason explanations are causal, reasons are not to be identified with causes.

agent to act on another person's wish, request, demand, without thereby acting on his own pro-attitude. In section C I consider the dispute between explanatory nonindividualism and explanatory individualism, rebutting an argument for the latter position. The argument relies on clarifying the notion of acting on another's expectation. I distinguish two concepts that could fall under that idea (section A). I focus on one of them showing how to extend the account to cover the concept of acting on another's expectation rather than one's own (section B). I conclude by considering our explanatory practices in light of the discussion (section D).

A. Acting on Another Person's Expectation of the Agent: Internalized and Non-Internalized Norms

Suppose that a ballet teacher, Mary, expects of Joe that he perform a particular configuration. Joe does perform the configuration just as expected. Is it something he has done *because* of Mary's expectation?

There are two ways to construe the idea of acting because of another's expectation. They depend on who selects the performance: the agent or the expector. Let us suppose that Joe internalized the norms involved in the expectation: he does know what it is to perform the configuration correctly. We might in fact suppose that he is at home where Mary does not even see him. It is most natural to think that Joe selects his performance in this case. Suppose, however, that Joe did not internalize the norms. Mary expects of Joe that he perform a particular configuration, but Joe is simply not competent in deciding whether a particular look-alike movement counts as being the movement she wants him to perform. In such a case, it would seem natural to think that it is Mary who selects Joe's performance. Joe produces successive attempts, and Mary selects them away until she finds one that is right. If she does not, she may shake her head upon which Joe continues.

These two different cases generate two different ways in which the selectional account will be applied.

> α φs because β expects of α that he ψ (internalized case) just in case,
>
> (a) β expects of α that he ψ, (b) α selects his performance p of φing,
>
> p fulfills β's expectation of α that α ψ, and α believes that p fulfills

β's expectation of α that α ψ, (c) α selects p because α believes that it fulfills β's expectation of α that he ψ.

We will say that Joe performs a particular ballet configuration (which he selects) because Mary expects it of him just in case (a) Mary does expect it of him, (b) Joe's performance does fulfill Mary's expectation and Joe believes that it does, (c) Joe selects the performance because he believes that it fulfills Mary's expectation of him.

α φs because β expects of α that he ψ (non-internalized case) just in case, (a) β expects of α that he ψ, (b) β selects α's performance p of φing, p fulfills β's expectation of α that α ψ, and β believes that p fulfills β's expectation of α that α ψ, (c) β selects p because β believes that it fulfills β's expectation of α that he ψ.

Joe performs a particular ballet configuration (which Mary selects) because Mary expects it of him just in case (a) Mary does expect it of him, (b) Mary selects Joe's performance, which as a matter of fact does fulfill her expectation and Mary believes that it does, (c) Mary selects the performance because she believes that it fulfills her expectation of Joe.

Unlike in the former case, the latter requires that the agent act under the supervision of the expector. The expector must in the very least be able to see what performances the agent generates. In the former case, on the other hand, the agent can be removed both spatially and temporally from the expector. The agent may still realize the expectations of his overpowering aunt, say, despite the fact that she is miles away.

In what follows, I focus on the cases where the agent selects his own performances. Such cases may appear to be more susceptible to the claim that one must give an individualist reconstruction of them. I should note, however, that the former cases are more interesting from the point of view of explanatory nonindividualism. They demonstrate that we are tied to one another in our agentive endeavors to a far greater extent than we might have thought.

B. Acting on Another Person's rather than One's Own Expectation

Suppose that β expects of α that α ψ, and that α expects of himself that α ρ. As a matter of fact α φs, fulfilling both expectations on this occasion. We can

straightforwardly understand what it means for α to act because of β's expectation rather than because of his own expectation.

> α φs because β expects of α that α ψ (rather than because α expects of himself to ρ) just in case (a) β expects of α that α ψ, and α expects of α that α ρ, and (b) α selects his performance *p* of φing, *p* fulfills both expectations on this occasion, and α believes that *p* fulfills both expectations, and (c) α selects *p* because α believes that it fulfills β's expectation of α that α ψ, not because α believes that it fulfills α's expectation of α that α ρ.[48]

It might be useful to emphasize again why clause (a) is needed. If only clauses (b) and (c) were required, we would not capture the idea of α φing because β expects of α that α ψ but at most say the idea of α φing because α believes that β expects of α that α ψ.

It will be useful to stress another point emphasized earlier (p. 192), for it will be important in our discussion in section C. I have indicated that the beliefs concerning what fulfills the expectation that figure in clause (c) ought to be construed at the very least as implying the beliefs that the relevant expectations are in force. In other words, α's belief that the performance fulfills β's expectation of α that α ψ implies at the very least that α believes that β expects of α that α ψ.[49]

Let me illustrate the characterization by considering the scenario of Milgram's experiments. An experimenter expects of an experimental subject *X* that *X* continue with the experiments (which involves administering shocks of ever growing intensity). The experimental subject, on the other hand, expects himself to abide by the norms he accepts, among others, not to hurt anyone unduly. At the beginning of the experiments,

---

[48] As indicated earlier, clause (c) can be understood in terms of appropriate counterfactuals: α would have selected the performance had α believed that it fulfills β's expectation of α (even if α believed that it frustrated α's expectation of himself), and had α believed that the performance would frustrate β's expectation of α, α would not have selected it (even if he believed that it would fulfill his own expectation of himself).

[49] As we shall see in section 4.C, if that were not the case then α selecting his performance because α believes that it fulfills β's expectation of α would imply that α selected his performance because α believed that it fulfills ξ's expectation of α that α ψ, for any ξ.

when the shocks are low, *X* administers a shock, thus fulfilling both expectations: he continues with the experiment (fulfills the experimenter's expectation of him) and since the shocks are low he does not hurt anyone unduly (fulfills his own expectation of himself). Which expectation is operative in his administering the shock? Why did he administer the shock? Milgram's experiments make it plausible to suppose that many people in fact act because of the experimenter's expectation of them. Suppose *X* was among them.[50] This would have meant that the following was true of *X*: *X* would have administered the shock as long as *X* believed that it fulfilled the experimenter's expectation of him that he continue with the experiment even if he believed that it frustrated his own expectation of himself not to hurt anyone unduly; but *X* would not have administered the shock if *X* believed that it would frustrate the experimenter's expectation (even if *X* believed that it would fulfill the expectation of himself). What Milgram's experiments have indeed shown is that more than half of his experimental subjects continued with the experiment (and so continued to fulfill the experimenter's expectation of them) even when their continuance meant that they would frustrate their own expectation of themselves. It is not implausible to take this as supporting the thought that the subjects act to realize the experimenter's not their expectations.[51]

Our characterization of the notion of acting for a reason appears to be inclusive enough to capture cases where a person acts because of another person's expectation of him. At the same time, there appears to be no need to reinterpret the case in terms of the subject having to have a pro-attitude of his own suitably directed toward the experimenter's wishes. The agent selects the actions that fit the other person's expectation of him and not his own expectation of himself.

Let us note, moreover, that the causal relationships involved are not mysterious at all. The agent causally produces (generates) the performances subject to selection. The

---

[50] I'm not trying to suggest that this is an adequate psychological profile of Milgram's subjects. In fact, it very probably is not. There might be many reasons and many expectations they support that form a complex web. My point is only that assuming this simplified map of expectations is correct, we can make sense of a subject acting because of the experimenter's expectation and using Milgram's evidence in support of that claim.

selection of one performance is effected by the agent's belief that the performance fits another person's expectation of him. When all this happens we say the agent acts because of the other person's expectation of him. As suggested at the end of section 3, we might say that the other person's expectation "causes" (in a selectional sense) the agent to act.

One might object, however, that the cases discussed make the nonindividualist's case easier, for the fulfillment and frustration conditions of the agent's own expectation of himself and another's expectation of him differ. What if they were the same? What if the agent performed an action of φing while desiring to φ and while another person desired the agent to φ? Could we apply the apparatus to this sort of case as well? I consider such an example in the next section.

C. Explanatory Individualism vs. Explanatory Nonindividualism

Nothing so far justifies the individualist reconstruction of cases of acting on another person's expectation. We will remember that the individualist reconstruction takes the following form: 'α φed because β wanted α to φ' must mean 'α φed because α wanted to satisfy β's desire that α φ'. I have argued that there are no conclusive arguments for adopting it already in Chapter I, though at an abstract level. Let us consider what such a reconstruction amounts to on our account. If it turned out that from the selectional interpretation of 'α φed because β wanted α to φ' it would follow that 'α φed because α wanted to satisfy β's desire that α φ', this would constitute a strong support for explanatory individualism. I will argue that no such implication holds. However, I will also show why it would be easy to think otherwise.

Let us consider the relevant reconstructions. Take acting on one's own reason first. Suppose that Jane wants to go to medical school. Her desire justifies her expectation of herself that she go to medical school. She does in fact enter medical school because she wants to. On our account, this means that she selects her entrance as

---

[51] It should be noted, however, that the case cannot be construed as demonstrating the falsehood of explanatory individualism. For the experiment does not show (and has not been designed to show) that in realizing the experimenter's expectations, the subjects were not realizing some expectation of their own.

something that fits the operative reason (she does not withdraw, as she might if by mistake she was accepted by law school, etc.). Moreover, it means that she would have selected her entrance as long as she believed that it fulfilled her expectation of herself to go to medical school, not otherwise.

Take acting on another person's reason, as I have proposed to understand it above. Suppose that Jane's father, a doctor, also wants Jane to go to medical school. His desire justifies his expectation of her that she go to medical school. What would it mean to say that she enters medical school because he wants her to? On our account, this means that she selects her entrance as something that fits the operative reason, viz. his expectation of her justified by his desire. Moreover, she would have selected her entrance as long as she believed that it fulfilled her father's expectation of her that she go to medical school, but not otherwise.

Finally, consider the individualist rendition of acting on another person's reason. Once again, Jane's father's desire for her to be a doctor justifies his expectation of her that she go to the medical school. What would it mean to say that Jane went to the medical school because she wanted to satisfy her father's desire that she become a doctor? On our account, this means that she selects her entrance as something that fits the operative reason, viz. her desire to satisfy his desire. Moreover, she would have selected her entrance as long as she believed that it fulfilled her expectation of herself to fulfill her father's expectation that she go to medical school, but not otherwise.

In summary, all the cases presuppose (a) that Jane expects of herself that she go to medical school and that Jane's father expects of her that she go to medical school. Furthermore, in all these cases, (b) Jane selects her entrance, and the selected performance fulfills both expectations, and Jane believes that the performance fulfills both expectations. We can then distinguish the three cases by appeal to clauses (c). Jane goes to medical school because of her own expectation of herself in case she would have selected her performance because she believed that it fulfilled her expectation of herself. Jane goes to medical school because of her father's expectation of her in case she would have selected her performance because she believed that it fulfilled her father's expectation of her. Finally, Jane goes to medical school because she expected of herself that she fulfill her father's expectation of her in case she would have selected her

performance because she believed that it fulfilled her expectation of herself that she fulfill her father's expectation of her that she go to medical school.

What then is the relation between Jane's going to medical school because her father desires her to and her going to medical school because she desires to realize her father's desire of her? I said that the case for nonreductive explanatory individualism would be rather secure if our reconstruction licensed the inference from 'Jane went to medical school because her father wanted her to go' to 'Jane went to medical school because she wanted to satisfy her father's desire that she go'. The reconstruction of the cases is similar except for the belief that is relevant to the selection. In the one case, Jane believes that the performance (she selects) realizes her father's expectation of her that she go to medical school. In the other, she believes that the performance realizes her own expectation of herself that she fulfill her father's expectation of her that she go to medical school.

The question that must be answered then is whether Jane's selecting her entrance because she believes that it fulfills her father's expectation of her implies that she selects it because she believes that it fulfills her own expectation to fulfill her father's expectation of her. The question is in other words whether Jane's selecting a performance because of one belief implies her selecting the performance because of another belief. Prima facie it is implausible that it be so. Since I have construed the 'because' in causal terms, it is prima facie implausible that when something is caused by one event it must be caused by another event.

What if, one might wonder, Jane's belief that her performance realizes her father's expectation implies her belief that her performance realizes her own expectation to fulfill her father's expectation? Indeed, one might note that the truth conditions of the beliefs are identical. This does not yet mean that one belief implies the other. However, the identity of the truth conditions may be what contributes to making the dispute so intractable. How then ought we to settle the question? One way is to see what the beliefs imply. I have noted that Jane's belief that the performance realizes her father's expectation of her implies at least that Jane believes that her father holds her to the

expectation.[52] Likewise, her belief that the performance realizes her own expectation of her implies at least that she believes that she holds herself to the expectation. Indeed, if it did not, then nothing would stand in the way of saying that Jane's belief that the performance realizes her father's expectation of her implies her belief that the performance realizes President Clinton's expectation of her. And if so, then by the above reasoning, her selecting her performance because of her belief that the performance realizes her father's expectation of her would imply that she selected her performance because of her belief that the performance realizes President Clinton's expectation of her. I take this to be absurd enough to suggest that indeed the relevant beliefs are construed in such a way that they imply that the agent also believes that the expectation she believes to be fulfilled is also in force.

If that is so, then the claim that Jane's belief that the performance fulfills her father's expectation implies that she believes that the performance fulfills her own expectation of herself is implausible. For the belief that her father holds her to the expectation does not imply the belief that she holds herself to the expectation to fulfill her father's belief. At least, it is not clear why an explanatory nonindividualist ought to be convinced otherwise. It is perhaps more clear why an explanatory individualist might think this implication to be plausible. But then the individualist argument against the nonindividualist would be question-begging.[53]

I conclude then that on our reconstruction of the relevant cases, it is plausible to think that it is not the case that whenever an agent acts because another person expects her to so act, she acts because she expects of herself to fulfill that person's expectation of her. This is not to say that it is never the case that an agent acts for both reasons. Quite to the contrary, this may often be the case. But to echo von Wright, it would be sheer

---

[52] See p. 192, above.

[53] An individualist might reply that her thought is not that one belief implies the other (in the abstract, as it were). Rather the thought is that the agent's selecting the performance because of the belief that it fulfills her father's expectation implies that she expects of herself that she please her father. But we must then ask: What reason is there for supposing the latter implication to hold? I have shown that someone who accepts the selectional account of acting for a reason is not committed to thinking that any such implication holds. This suffices to show that someone who accepts the selectional account can coherently adopt the position of explanatory nonindividualism.

prejudice to suppose that the agent can never act on another person's expectation without at the same time acting on the agent's own expectation.[54]

This puts the ball in the individualist court. I have not argued that there could be no arguments that the individualist can resort to. But the most natural argument is not available to the individualist. I conclude then that unless further arguments are forthcoming explanatory nonindividualism is a natural position for someone who accepts the selectional account.

D. Our Explanatory Practices

The account suggested allows us to obtain a clearer picture of our practices of explaining one another's actions. We sometimes explain our actions by appealing to the *reasons* that justify the expectations the action fulfills. "I took the umbrella because it would rain" cites a fact (a reason) that justifies the expectation of myself to protect myself from the rain. The explanation can appeal to my *expectation* of myself to so protect myself, as in "I took the umbrella because I intended not to get wet." Finally, the explanation can emphasize the justification of the expectation by pro-attitudes and beliefs "I took the umbrella because I believed it would rain and wanted to avoid getting wet."[55]

Aside from the explanations that appeal to the agent's intentions and reasons, we also explain actions in terms of other people's wishes and expectations of the agent. We can appeal to the *expectations* of the agent (as in explanations in terms of what other people requested, demanded, etc.) or we can appeal to the *reasons* others had for such expectations (as in explanations that appeal to what other people wanted of the agent, or wished the agent would do).

Other explanations can also be understood in these terms. For instance, explanations that appeal to the agent's social role or position implicitly invoke normative expectations associated with such a role. Likewise, many explanations in terms of the

---

[54] Georg Henrik von Wright, "Explanation and Understanding of Action," in *Practical Reason* (Ithaca: Cornell University Press, 1983), p. 55, cited on p. 18, above.

[55] Annette Baier ("Rhyme and Reason: Reflections on Davidson's Version of Having Reasons," in (eds.) Ernest LePore, Brian P. McLaughlin, *Actions and Events, op. cit.*, pp. 116-129) argues that normally we cite facts as reasons. It is only when the agent's having the reason might be called into question that we would revert to the citation of beliefs or desires.

agent's character will invoke the normative expectations that the agent (as a person with such a character) will or should have of himself.

E. Final Remarks on Explanatory Individualism

I have demonstrated in sections A and B that there is no special problem in extending the account of acting for reasons developed in section 3 to cover the case of an agent acting because of another person's reason. To that extent, the account developed offers a reason for holding a nonindividualist position. In section C, I have considered the question whether there is any reason for someone who accepts the account to suppose that when an agent acts because of another person's expectations of her, she also thereby acts because of her own expectations of herself. The availability of such a reason would justify the position of explanatory individualism and thwart the prospects of explanatory nonindividualism. I have concluded that nothing in the account dictates the position of explanatory individualism.

One may object that the account still does not explain how the other person's expectation affects the agent. There is no psychological mechanism that has been presented. But one may reasonably inquire what kind of psychological mechanism was expected. Most probably, the expectation concerned the identification of some pro-attitudes on the part of the agent that would move him to action. In other words, what one expected to find is a rationalization of the agent's action, an account that would make it reasonable on the agent's part to act as he did. As I noted in Chapter I, this expectation is licensed by our adherence to normative individualism, to the belief that every action can be rationalized by the agent's reasons. Our concern, however, has been with showing the possibility of explanatory nonindividualism, which is quite compatible with normative individualism. The point is only that it is possible for an action to be explained in terms of another person's expectation and not always in terms of the agent's own expectations of himself.

I would like to close this section by noting that although I do not believe that explanatory individualism is the correct view about action explanation, there are nonetheless grains of truth in it which are preserved in the nonindividualist account I have proposed. Much of the motivation behind the individualist tendency was the

thought that the nonindividualist picture involves a kind of action at a distance. It seemed prima facie incoherent that another person's expectation could affect the agent's action without mediation through some intentional attitudes of the agent. The nonindividualist view advanced is far from supposing that such an action at a distance occurs. A crucial role in the model is played by the agent's selecting performances that fit and do not fit the reason which is operative in the action. I have argued that it is most plausible to construe the agent's belief concerning what performances fulfill the expectations justified by the reason as being causally involved in the selection. I have, however, resisted the specifically individualist claim that the mediation has to go through the agent's pro-attitudes.

5. Two Further Problems

The suggestion that we ought to conceive of the agent as selecting actions to fit normative expectations supported by the reasons for which the agent acts leaves the following issue completely in the dark. How is it that the agent actually performs the action? Given that the performance of an action is a causal process, it seems absolutely mysterious to suppose that the reason for which the agent acts, the goal that he intends (or is expected) to realize, is not somehow causally involved in the generation of the performance.

One way of putting the point is that while the selectional account simply has to assume that the agent is disposed to generate certain performances, the causal theory of action explanation can actually *explain* why the agent is so disposed. The agent is so disposed because the reason, which is his causally efficacious state, exerts causal pressure thus disposing the agent to the performance of the right sorts of actions in the right sorts of circumstances. This is essentially an objection that has been launched against G.H. von Wright's account by his otherwise sympathetic reader F. Stoutland:

> I raise my arm and my arm rises. Why does my arm rise just then? Can it be merely a brute (but fortunate) fact that when I intend to reach for a book and believe I must raise my arm to do so, that my arm rises so that by that behavior I can intend to get a book? … Von Wright writes that 'it is an empirical fact that a

> man *can do* various things when he decides, intends, wants to do them'.[56] The problem … is that [his account] appears to render this fact unintelligible, not only by making it unclear why it obtains but making it difficult to understand how it could obtain. If the behavior by which I intend a result has a Humean cause as sufficient condition, then it is a mystery why behavior occurs *when* I act.[57]

Two problems are usefully distinguished. One worry concerns an account of the very generation of action. If reasons are not causes then what causes the agent to act in accord with them? This is the first problem, the problem of spontaneity. The second problem, the problem of congruence, can be stated in the following way: We saw that the selectional account of acting for reasons presupposes that the agent is either reliable or semi-reliable in responding to his reasons.[58] We have seen (section 3.A) that there are conceptual reasons to exclude the possibility of our being in general anti-reliable.[59] But we have also seen that while our concept of action would be slightly different if we were in general semi-reliable, for it would have to take into account our agentive stutter, it would be still recognizable as a concept of action. Indeed, the distinction between acting for and acting with a reason would still find application if we were semi-reliable. The second problem of congruence is this. It is by and large true that, in most circumstances, for many types of bodily actions (especially simple actions like raising an arm, flipping a coin, putting on eye glasses) we are in most circumstances reliable. Let us call this fact (F).

(F) For most agents, in most circumstances, the agent is disposed to produce bodily actions that she is disposed to select.

The question is *why* (F) holds.

---

[56] Georg Henrik von Wright, *Explanation and Understanding* (Ithaca: Cornell University Press, 1971), p. 81.

[57] Frederick Stoutland, "Von Wright's Theory of Action," in (eds.) P.A. Schilpp, L.E. Hahn, *The Philosophy of Georg Henrik von Wright* (La Salle, IL: Open Court, 1989), p. 323.

[58] Let us recall that we have characterized an agent as being reliable insofar as he is disposed to produce performances that he is disposed to select. By contrast, an agent is semi-reliable if he is not completely reliable and it takes several unsuccessful attempts before he produces a performance that he would be disposed to select. An agent is anti-reliable if there is no regular or reasonable coincidence between his generating and selecting mechanism, if he does not produce the performance he is disposed to select within a reasonable amount of trials.

[59] See footnote 22.

Teleological theorists of action have usually responded by producing conceptual arguments to the effect that our concept of action would find no application if we were anti-reliable.[60] The problem is that the necessity of our not being anti-reliable (if the concept of action is to find any application at all) does not yet show that we must be reliable. It shows that we could be reliable but we could also be semi-reliable.

It is thus that a causal theorist might claim superiority by being able to account for both problems. The hypothesis that reasons are causes leaves no mystery with respect to the cause of the action, and it also explains why we are by and large reliable rather than semi-reliable. We are reliable because reasons as causes dispose the agent to produce just the right performances, just the performances that would be selected by the agent as according with her reasons. If the hypothesis that reasons are causes was the only way to account for both issues, it would constitute an argument for the causal theory of action explanation. But it is not.

A. The Problem of Spontaneity: Why do We Act at All?

It is worth beginning by inoculating oneself against one way in which this worry arises. One may be tempted to think that conceiving of an agent as being moved by reasons implies conceiving of the agent as having to be put into motion by a reason. The Aristotelian thought is that the agent would do nothing unless he were to be moved by a reason. On such a picture, it is extremely natural to associate the motivational power of reasons with their literally pushing the agent into motion.

A convincing way of getting rid of this picture has been suggested by John Dewey. Dewey reminds us of an analogy with physics. One of the questions that dominated Aristotelian physics was the question how motion is possible at all. The presupposition of this question is that the "initial" state of an object is to be at rest, and so that what calls for explanation is the fact that the object moves. It is this presupposition that has been questioned in physics. Dewey suggests that it should likewise be questioned in psychology:

---

[60] "That [we are in general reliable] is a contingency. But it is nothing to be surprised at. For it is a condition which the world must satisfy if we are to entertain our present notions of action and agency."

> The idea of a thing intrinsically wholly inert in the sense of absolutely passive is expelled from physics and has taken refuge in the psychology of current economics. In truth man acts anyway, he can't help acting. In every fundamental sense it is false that a man requires a motive to make him do something. … Anyone who observes children knows that while periods of rest are natural, laziness is an acquired vice — or virtue.[61]

Moreover, not only are we naturally active, so that we do not need stimulation by reasons to remove us from a passive state, but we find ourselves in complex webs of reasonable expectations. As such, even our staying motionless may count as something we do, for I may be acting by *omitting* to do something. Indeed, as we saw in Chapters V-VI, in only very special circumstances do we actually get off the agentive hook. Most of the time we are doing something.

Bearing this thought in mind ought to relieve the impression that the very notion of agency becomes inert unless we conceive of reasons as causes. But it is not unreasonable to ask what causes actions. And the most plausible answer is that there is no unified account to be given. I have already pointed out in section 3.F that while some causes of actions are plausibly identified with desires (visceral desires, e.g.), in other cases, the causes are more plausibly construed as perceptions, noticings, etc.

But actions may also have non-psychological causes. Consider the following example. I want to jump into the water as I am trying to swim my few laps for the day. I stand over the water, almost prepared to jump. But, for whatever (if any) reason I do not jump in yet. Perhaps because I am daydreaming, or something has caught my attention. My mother, who stands behind me, gives me a gentle motherly push, which causes me to jump in. It is important to emphasize that the push is gentle: if it were not, it might count as a defeating condition. But this push is just "a reminder." As it turns out, however, it causes me to jump in. (Were I not pushed, I would not have jumped in at this moment.) This is a case where it still plausible to think that I jumped in to swim my few laps, but where my desire to jump into the water was not a cause (not an immediate cause) of my action.

(G.H. von Wright, *Explanation and Understanding, op. cit.*, p. 132.)
[61] John Dewey, *Human Nature and Conduct* (New York: The Modern Library, 1957), p. 112.

I conclude that while the hypothesis that reasons are causes constitutes a simple way of accounting for the etiology of action, it is not the only explanation there is.

B. The Problem of Congruence

Why do we by and large generate the bodily performances we are disposed to select? If we supposed that reasons are also the causes of actions, it might appear that the mystery would be resolved.[62] But there are other ways of accounting for (F). In fact, I would like to suggest that there is not, nor need there be, a unified account of why (F) holds. Instead, we can invoke various considerations that support (F) and are compatible with the selectional (and more generally teleological) construal of reasons.

Functional Explanation in Interpersonal Contexts. In interpersonal contexts, where the reasons in question justify expectations that one person has of another, a different kind of explanation why agents by and large fulfill the expectations placed on them may be in order. The explanation in question is familiar from the adaptational interpretation of Marxist explanations.[63] Rulers tend to maximize their power not because of their more or less hidden desires but rather because those rulers who did not exhibit appropriate tendencies lost in the competition with those rulers who did. The bourgeoisie adopts self-serving beliefs which justify their oppressive activities not because of any deeply rooted

---

[62] This is not necessarily the case. The mystery is regenerated on Davidson's anomalous monism. Insofar as Davidson insists on only a token-identity between mental states and physical states and insofar as he insists that mental causation proceeds in virtue of mental causes being physical causes, it becomes quite mysterious how one type of mental state (desire to φ) can reliably cause another type of mental event (the action of φing) without relying on any finite number of types of physical causal relations. (The worry was first formulated by Frederick Stoutland. See his "The Causation of Behavior," *op. cit.*; "Oblique Causation and Reasons for Action," *Synthese* 43 (1980), 351-367. For further discussion, see J. Heil, A. Mele, *Mental Causation, op. cit.*). A particularly helpful response is due to Peter Smith who argues that Davidson must respond by appealing to general functionalist principles. The reason why there is no mystery is that unless a desire to φ (and consequently whatever physical states it happens to be identical to) by and large caused φing, it could not count as a desire to φ.

[63] Leszek Nowak, "Theory of Socio-Economic Formations as an Adaptive Theory," *Revolutionary World* 14 (1975), 85-102. Leszek Nowak, ed., *Dimensions of the Historical Process. Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 13* (Amsterdam: Rodopi, 1989). G.A. Cohen, *Karl Marx's Theory of History. A Defence* (Princeton: Princeton University Press, 1978).

desires but because it increases their chances of surviving in competition with rulers who do not hold such beliefs.[64]

Typically, the explanation relies on pointing out that individuals in a certain social position are subject to selectional pressures. Applied to our question, as long as it is true that the agent who is held to an expectation is subject to a selective pressure that would eliminate him from a social position he occupies, it will be true that agents in that social position are disposed to produce performances that they are disposed to select. This, together with some further assumptions regarding the stability of habits, for instance, allows us to understand how an individual (in a certain social position) is reliable in generating the performances that are appropriate in such a situation.

This does not establish the truth of (F) in all contexts. But it does show it to be plausible in social contexts where the agent's position is at stake in case she is not disposed to fulfill the expectations.

Skill and Learning. This brings us to the contingent fact about us that we are capable of developing skills and learning how to respond to many stimuli, some of which might be reasons. Why we are capable of learning is beyond the scope of a philosophical theory of action explanation. But there is a worry here that ought to be addressed.

A causalist sympathizer may still argue that even if one agreed that intentions do teleologically and selectionally guide the acquisition of skills, this would still leave room for the causalist interpretation. For when the agent acquires the skill he must respond to some internal state of his that is a representation of some state of the world. Otherwise, he could not be responsive to the situation. When he acquires the skill, this presumably means that he becomes responsive to a certain internal representational state. This state is none other than a reason. In other words, the causalist may appeal to a functionalist understanding of mental attitudes. Reasons are simply those states that (among other things) cause the appropriate actions. This appears like a position that is very hard to

---

[64] Interestingly, Denise Meyerson has recently argued that the functional interpretation of false consciousness is incoherent unless it is supplemented with an individualistic explanation that appeals to the rulers' desires (*False Consciousness* [Oxford: Clarendon Press, 1991]). I rebut her contention and show that her arguments for individualism exhibit an individualist bias ("Must False Consciousness Be Rationally Caused?", forthcoming in *Philosophy of Social Sciences*).

undermine. If it is prerequisite for a state to count as being a reason that it cause the action, then indeed, it is rather hard to see how one could possibly deny that the reason does cause the action.

But we can undercut this argument by denying that one must identify whatever internal state causes the action as the reason. One would have to so identify the state if the only intelligible account of action explanation were the causal account. I have been trying to argue that the selectional account is a viable alternative.

Some Causes May be Reasons. While I do not believe that we have sufficient reason for thinking of reasons as causes in general, still this is no reason to deny that in some cases, our reliability in response to reasons might be accounted for by the fact that they are also causes. I have already suggested that many phenomenologically prominent and biologically grounded desires are most naturally construed as causing actions. Other causes of actions, as we saw above (section 3.F), are more plausibly construed as perceptions, noticings, etc.

It is important to point out that even if some causes of actions are understood as reasons (the desire for water, for instance) this does not eliminate their proper function as selectional criteria. Someone in a state of extreme thirst may be caused to chaotically grab for anything in sight in search for water. The desire for water, we might say, causes these chaotic movements. But it also functions as a selectional criterion in accordance with which the agent then chooses the bowl that contains water rather than the one that contains vinegar, say.

Difficulty of task. Paradigmatically, we are reliable rather than semi-reliable in performing simple bodily actions: raising one's arm, walking, jumping, turning around, and so on. What is true about such actions in general is the fact that they have rather wide fulfillment conditions. A lot of arm movements count as successful arm raisings. A lot of leg movements count as a successful walking motions. And so on. The reason why this is a relevant consideration is that the wider the class of fulfillments the more likely the agent is to succeed at fulfilling the expectation. Correlatively, the wider the class of fulfillments the better the chance that the agent will be reliable in responding to a reason to raise his arm rather than merely semi-reliable in such a response.

We might contrast the wide class of mundane bodily motions with not so mundane bodily motions: certain motions in ballet or Chinese opera, where almost every detail of an arm raising is subject to evaluation, where what one expects is fulfilled in a much narrower class of cases. Most of us, though reliable in raising our arms, could be safely taken to be semi-reliable at best in cases of such more sophisticated actions.

In sum, I have argued that other explanations are available to account for the problem of spontaneity and the problem of congruence than the causalist explanation relying on the thought that reasons are causes.

6. Objections

A. Reasons Cause the Agent to Select the Action

One might reasonably raise a question concerning the nature of the selection that the agent is supposed to effect. Suppose that we deal with a case where the agent's action stutters: the agent produces several performances and selects one that fits what he wants to do. Let us say that an actor sits in front of a mirror producing smiles; he intends to express a complex of emotions with the smile in an upcoming play. He is not satisfied with the smiles he generates, so he keeps on going until one fits what he wants. One might try to argue here that we must think that the agent's selecting the last smile (which accords with what he had reason to do) is itself caused by that want. So even if we do not construe the reasons as causing the performances, they must be construed as causing the agent to select the right performances.

Another way of putting the objection is that the agent's selecting an action must be conceived as itself an action that is done for a reason. This would be unintelligible on pain of infinite regress. If the agent's selecting an action is itself an action done for a reason, then we would have to refer to the idea of the agent selecting his selecting of an action. And if his selecting his selecting of an action were again conceived as an action we would have to refer to the idea of the agent selecting his selecting his selecting of an action. And so on ad infinitum.

Fortunately, neither do we need to think of the agent's selecting the performance as the agent's action done for a reason.[65] The agent who selects the performance merely *recognizes* it as fitting his reasons. The only resources required for an agent to select a performance as according with a reason are for him to have a conception of a performance fulfilling an expectation supported by that reason. To the extent that the agent has a conception of a performance fitting a reason (i.e. is capable of correctly sorting performances that fit the reason and those that do not), he is able to select a given performance as fitting the reason. We do not need to appeal to the idea of the reason causing such a selection. For the agent to select the performance is simply to apply his conception of the reason.

B. Reasons Always Cause Actions Done for a Reason

If we understand the idea of a reason "causing" an action in a selectional sense then it is true that reasons always cause actions done for a reason. What if one wanted to insist the reason causes the action in the sense of generating it (subcomponent of the above picture)? I have not offered conclusive reasons for rejecting the suggestion. Rather I have only related reasons various philosophers have proposed for being skeptical of it. The primary commitment of the selectional model lies in insisting that most of the reason's "work" (including the part responsible for giving an account of acting for rather than acting with the reason) lies not in the generational component of the model but rather in the selectional component. It would thus not be impossible for some selectional theorist to agree that reasons causally generate the actions.

In view of the considerations for being skeptical that reasons always causally generate the actions they explain (see section 3.F), one might ask *why* one would want to insist that reasons do in fact *always* causally generate the actions they explain. I see two types of arguments for the claim. First, one might argue that this is the natural picture of

---

[65] The agent's selecting a performance is still an action of his as long as it would have been reasonable$_A$ to expect of the agent that he select the performance. If the agent were steered to select the performance by someone else (through an implant in his brain, for instance), his selecting of the performance would not be something he did.

the relation between an action and those reasons that explain it. Second, one might argue that the claim that reasons cause actions is of theoretical significance.

The first consideration will have differing import depending on the value that one attaches to the natural picture of things. But in any case, there is nothing in the selectional account to offend the natural intuitions, provided that they do not claim to be more sophisticated than they are. I have been emphasizing that the account can be construed as precisely explaining what we mean when we say that reasons "cause" actions. To say that our natural intuitions exclude the selectional meaning of "cause" proposed here would be, however, to overtheorize them.

The second consideration carries more substantial weight. And so Donald Davidson has argued that we must understand the relation between the action and the reasons that explain it in causal terms or else not be able to make the distinction between acting for and acting with reasons. As long as we take the term 'cause' to encompass the selectional sense of "cause" advocated here, then nothing I have said contradicts Davidson's conclusion. But then it is also true that reasons need not causally generate actions. Davidson's argument fails otherwise. For I have precisely offered an account of the distinction between acting for and acting with a reason that does not rely on the idea that the reason causally generates the action.

C. The Account Is Secretly Individualist

Despite my claims that the account extends to cover actions done because of what someone else expects of the agent but not because of what the agent expects of himself, it might be objected that the account is really individualist at heart. Consider for instance the fact that it relies crucially on the agent's belief that the action fulfills somebody else's expectation. Granted that the expectation involved is somebody else's but the belief is still the agent's. Moreover, the belief is *about* somebody else's expectation. And this is just the thought the individualist insisted upon: that others' attitudes are relevant only insofar as they are mediated by the agent's attitudes suitably directed toward them.

The response to this objection is straightforward. It is true that the account appeals to the agent's belief about another person's expectation.[66] And it is true that the shape of this thought has been at the heart of the individualist account all along. But this does not show the selectional account to be individualist rather than nonindividualist. The explanatory nonindividualist insists that it is possible for an agent to act on another person's pro-attitude without at the same time acting on the agent's own pro-attitude (not belief!) suitably directed toward the other's pro-attitude; the explanatory individualist takes this to be impossible. I have argued that the selectional account allows us to understand the nonindividualist thesis and the fact that it appeals to mediation by the agent's belief in no way undermines it. At the very best, it shows that the individualist had some good intuitions but that she was mistaken about their exact form. The nonindividualist should have no problems in admitting that there is a grain of truth in the individualist thought. Indeed, all such grains allow us to better understand why the view has been so captivating.

One might push the objection further, however. One might observe that in fact the only attitudes that are involved are the agent's attitudes. It is the agent's belief about another expectation that is causally involved in selecting the performance. Some of the agent's attitudes may be involved in generating the performance. But the other person's expectation is not involved at all. In fact, one might object that the agent's belief is quite sufficient for the account and hence that the other person's expectation is not necessary at all.

Two points need to be emphasized. First, the objection confuses the levels of relevance. The selectional account is intended to make clearer what it means to say that another person's expectation is operative in the agent's action, just as it is intended to explain what it means to say that the agent's expectation is operative in his action. Neither in the case where the agent acts on his own expectations nor in the case where he acts on others' expectations is it the case that the expectation is involved in the action on

---

[66] Note that the objection targets specifically actions done on another person's expectation where the agent internalizes the norms involved. It would not hold for cases where the norms are not internalized, where it is the other person who selects the agent's performances (see section 4.A). I have decided, however, to make the primary case for nonindividualism based on the former kind of cases.

the same level as the belief about whether a performance fulfills the expectation or whatever (if any) attitude generates the performance. Rather, the "operative" relation between the expectation and the action emerges from the particular configuration of causal and other relations. This is why the selectional theorist can agree that the expectation or reason "causes" the action (in this special sense) while denying that it does so in the same sense in which the belief causes the selection, say.

Second, I have emphasized that the agent's mere belief about what fulfills another person's expectation is not sufficient for us to think that the agent acts on another person's expectation. The case where the other person does not actually hold the agent to the expectation, while the agent believes that she does, is parasitic on the case where the agent is both held to the expectation and believes that he is. In such a case, where all the other conditions are realized, we would be inclined to say not that the agent acts because of what the other expects him to do, but rather that he acts because of what *he thinks* the other expects of him.[67] This means that we intuitively recognize the fact that the other person's holding the agent to the expectation is part and parcel of the idea of the agent acting on that expectation. This is not to say, however, that the other person's expectation must causally generate any performance of the agent. This is precisely the picture that makes the nonindividualist view implausible. But I have argued that we can make sense of it using a different picture of the relation between reasons, expectations and actions.

D. The Account does Not Refute Explanatory Individualism

Finally, one may argue that the account proposed does nothing to dislodge the comfortable niche of the explanatory individualist. All that I have given, the objector claims, are conditions under which it would be appropriate to say that the agent acted

---

[67] One might argue that this only shows that the "non-parasitic" cases can be analyzed as the agent acting because of what he thinks the other expects of him together with the fact that the other actually does. This is a familiar reversal of mental concepts characteristic of phenomenalism, identified and opposed by Wilfrid Sellars in "Empiricism and the Philosophy of Mind." Sellars focussed on the relationship between "looks" and "is" of perceptual reports. His analysis has been extended to cover other concepts, in particular the relationship between "tries" and "does" (Robert Brandom, *Making It Explicit* [Cambridge: Harvard University Press, 1994]).

because of another person's expectation. But I have not definitively shown that for such actions the agent fails to act on *some* expectation of her own. The individualist will point out that his claim is indeed very weak: he requires only that there be *some* expectation of the agent's on which she acts. A nonreductive individualist does not quarrel with the claim that an agent may act on another person's expectation, as long as it is also true that she acts on *some* expectation of her own. He concludes that since I have not shown on any of the examples discussed that there is no expectation of the agent's on which she acts, the position of the nonreductive explanatory individualist stands unshaken.

I agree that I have not shown in either of the examples that there is no expectation of the agent's on which she acts. If I were able to demonstrate it, this would constitute a conclusive argument for rejecting explanatory individualism. I have not, however, claimed to offer any such arguments in the first place. My sole intent in the entirety of the dissertation has been to lay some groundwork for the development of two sibling-thoughts: nonintentionalism in the theory of action and nonindividualism in the theory of action explanation. I have not been trying to offer conclusive arguments for these views and against the alternative positions. Rather, I have been trying to show that despite initial appearances to the contrary, these positions are coherent and defensible.

It is in this spirit that the selectional account of the explanatory force of action explanations ought to be taken — not as refuting explanatory individualism, but rather as vindicating the coherence (not the truth) of explanatory nonindividualism. And this it does. I have shown how an explanatory nonindividualist can coherently claim that an agent acts because of another person's expectation without thereby acting on the agent's own expectation suitably directed to the other's expectation (see section 4, in particular section 4.C). What I have not shown is that there is any one particular example of an action that would convince the individualist to abandon his position. Moreover, there are good reasons to believe this task to be very difficult if not impossible. It would be very hard to give a complete list of reasons and expectations the agent has for performing a particular action, thus making it hard to set up the grid of counterfactual situations required for the account to apply. For no particular action do we actually know (though we may have a good idea based on what we know about the agent's past performance, say) what the agent would have done in these counterfactual situations. It is thus no

wonder that the individualist may always point to just another reason the nonindividualist has not considered.

This might lead one to wonder whether anything has been accomplished at all. The fact that explanatory nonindividualism has been shown to be at least coherent is not a small achievement in itself. Moreover, explanatory nonindividualism offers a prima facie more straightforward understanding of those ordinary explanations that appeal to other people's desires, wishes or expectations of the agent. Instead of requiring that such explanations be enthymematic and so in need of an individualist reconstruction (supplementing the agent with appropriate pro-attitudes), the nonindividualist lets them stand at face value. In this way, explanatory nonindividualism may be appealing as a more faithful representation of our practices.

● ● ● ● ● ● ●

This completes our discussion of the force of ordinary action explanations. I have proposed that we understand the idea of acting for reasons on the model of conceiving the agent as selecting actions to fit his or her reasons. An ideal interpreter, equipped with a knowledge of what beliefs caused the agent to select a given performance, can tell whether the agent acted because of a reason rather than merely with it.

We have seen that the causal theorist of action explanation may claim residual superiority for his account by suggesting that only that theory can account for the problems of congruence. In section 5 we have seen, however, that the problems of congruence can be answered without appealing to the thought that reasons are causes. I have not argued that reasons are never causes. I have merely argued that they need not be thought of as causes, and that it is intelligible to deny that some reasons are causes.

Ultimately, I have made a reconciliatory move toward a kind of causalist view. I have allowed that the selectional model of acting for a reason could be seen as elucidating what we mean when we say that the reason "causes" the action. I have only insisted that it is important to recognize that the emerging "causal" relation is of a different kind, that it operates at a different level, from the causal relations that are part of the model. Moreover, I have argued that on this interpretation of the "causal" relation it is unproblematic to say that another person's expectation of the agent "causes" the agent

to act, without thereby implying that the agent acts because of some of his own expectations. I have thus offered a further reason for holding explanatory nonindividualism, a position for which conceptual space was opened in Chapter I.