

THE SELECTIONAL FORCE OF REASONS

Katarzyna Paprzycka

The debate between the causalists and the teleologists has reached something of a standstill. In the 1950s, it was widely believed that the proper way of thinking about action (reason) explanations is in exclusively teleological terms and that the very idea of causality is misplaced in a systematic thinking about the relation between actions and reasons (e.g.: Anscombe 1963; Melden 1961; Peters 1958; Ch. Taylor 1964; R. Taylor 1966). This atmosphere was disrupted by Donald Davidson's famous paper "Actions, Reasons and Causes" (1963). He argued that without the invocation of the idea that reasons are causes, one cannot account for the idea of reasons' efficacy, which is manifested in the distinction between acting for reasons and acting while merely having reasons. The teleologists have answered that teleological explanations do too support the distinction (e.g. Collins 1987; von Wright 1971; Wilson 1989). But other challenges ensued. For example, Frederick Stoutland (1976; 1989) objected to G.H. von Wright's version of the teleological theory that a teleological explanation leaves it mysterious why a behavior occurs when the agent intends it to occur. More recently, William Child (1994) argued that reason explanations must be capable of explaining why an action occurs just when it occurs and only a causal explanation can do so. Such challenges are usually met either by demonstrating that teleological explanations are capable of meeting them or that they are not really general features of ordinary reasons explanations (see, for example, Hursthouse 2000).

This dialectical pattern reveals a basic division in our intuitions on these matters. In particular, it reveals a schism in the degree of comfort (discomfort) we feel vis a vis teleological explanations more generally. In such a situation, it might be useful to turn for guidance to the other area, aside from human agency, where teleological relations play a prominent role, viz. biology. One of the virtues of Darwin's theory of natural selection is

that it demystifies the teleological relations in biology by letting them emerge from causal and nonmysterious relations.

In this paper, I will explore a selectional way of looking at the relation between actions and reasons. I will not only show that the account can meet the challenges set forth, but that it has the considerable advantage of reconciling the intuitions driving the causal and the teleological accounts.

In §1, I characterize the notion of a selectional system and describe some general features of selectional explanations. In particular, I introduce Elliott Sober's (1984) 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 selection. (This distinction will be the root of the idea of a reason being operative in action and so it will ground the distinction between acting for and acting while merely having a reason.) I then show how to understand an agent as a selectional system (§2). This provides the requisite background for drawing the distinction between acting for and acting while merely having reasons (§3). To that extent, Davidson's causalist challenge is met. In §4, I consider Stoutland's challenge. 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 §5, I give a reason for thinking that the selectional account is to be preferred over the causal account. I end with considering some objections to the account (§6).

1. Selectional Systems

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.

In response, the teleologists question our scientific inheritance which underlies the custom (or prejudice) of explaining phenomena in terms of efficient causes. Such an account questions the very impulse of 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. To substantiate this claim, they offer examples of teleological phenomena (usually from biology).

As can be expected, the causalists find such a response deeply unsatisfying. They claim that the reason why we are quite satisfied to employ teleological relations in biology, say, is that we have a deeper understanding of the causal processes involved in these phenomena. Such examples would not be quite as satisfying if we did not have a framework of the theory of evolution or of the particular causal workings of an organism to look to. Appearances to the contrary, then, teleological relations do too require us to have at least a sketch of the causal goings on in order for us to have the sense that the relations are not mysterious.

The causalists indeed go on to provide such a demystifying account of the teleological relations characteristic of action. They claim that certain mental states¹ are the efficient causes of actions whose content is suitably related to the future end. The intention-state is not the end itself, but it reflects, represents or embodies the end of the

¹ What these mental states are will depend on the particular brand of the causal theory one accepts. They can be belief-desire pairs, prior intentions, intentions-in-action, volitions, etc.

action. Let us call this the “simple causalist” or the “causalist” interpretation of teleological relations.²

But this is certainly not the only broadly causal understanding of the teleological relations that is available. While the “simple causalist” interpretation of teleological relations has found quite a comfortable niche in the domain of human action, it has also been proposed in the 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 simple 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

² William Child’s (1994) proposal is an exception. He denies that reasons are causes but suggests that a causalist only needs to stand by the claim that reason explanations are causal explanations. The problem with this is that it is notoriously not clear how to understand what a causal explanation is. Without going into the cumbersome details here, I should only like to observe that if the proposal advanced here is right

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 to which the individuals are subject.

In the next section, I will try to make more concrete the proposal to exploit this “selectional causalist” interpretation of teleological relations 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 Sober’s distinction between selection-for and selection-of 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. This includes standard examples of cases of natural and artificial selection but also cases of more simple selectional process such as using a sieve (Dawkins 1986). A slightly more elaborate example is presented by Elliot Sober (1984, p. 99). He describes a cylinder-shaped *selection toy* with four horizontal levels. 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.

Selectional explanations of phenomena rely on what one might figuratively call two focal points. One of the focal points is a “selecting mechanism” which effects a

then reason explanations turn out to be both causal explanations and teleological explanations at the same time.

selection of individuals according to some selectional criterion. The second focal point is a “generating mechanism” which supplies individuals on which the selection operates. And so, in the case of natural selection, the generating mechanism comprises 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). In the case of artificial selection, the generating mechanism is likewise genetic variation and reproduction. The selecting mechanism, on the other hand, lies in the hands of the breeder. It is the breeder who chooses which organisms to reproduce further. In the case of the selection toy, all the individual balls enclosed 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 letting the balls of appropriate size pass.

Another useful idea is the characterization of a selectional process as an adaptational process. What makes a selectional process adaptational is the fact that the generating mechanism tends to supply those individuals that would be selected by the selecting mechanism. 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.³ 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. 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 *only* those individuals that would be selected by the selecting mechanism. 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

³ 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.

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, supplied by the generating mechanism, subject to a selecting mechanism (according to some selectional criterion). Such a selectional system can support selectional processes. The selectional processes can differ among others in the number of cycles to which the individuals are subject. A selectional process is perfectly adaptational (adaptational) with respect to a characteristic *C* only if the generating mechanism generates only (tends to generate) those individuals that will be selected by the selecting mechanism (i.e. only individuals that are *C*).

One crucial fact about the process of selection has been emphasized by Elliot Sober (1984) and brought out in his distinction between *selection for* (properties) and *selection of* (objects). 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.

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

balls were selected was coincidental. It 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.

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.

2. The Agent as a Selectional System

An agent can be thought of as a simple selectional system. The generating mechanism consists in the agent’s ability to produce a variety of actions. The selecting mechanism consists in the agent’s ability to recognize whether a action fits a selectional criterion, whether a action satisfies a reason.

The reason why the suggestion that an agent is some sort of a selectional system may appear 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.⁴ With respect to those types of actions, we are reliable in generating the actions that we would select as realizing what we mean to do. It is also in cases where our reliability is disturbed that it becomes clearer how we could think of ourselves as selecting actions in accordance with our reasons.

It is useful to distinguish two ways in which an agent could fail to be reliable. One possibility (semi-reliability) is that rather than generally succeeding in producing the action right away, the agent could succeed eventually but only after a couple of trials. Another possibility (anti-reliability) is that the agent generally does not succeed in producing the action at all. Most of us are reliable, semi-reliable and anti-reliable with respect to various types of actions. There are actions with respect to which we are *reliable*. If I intend to raise my arm, I most likely will succeed in so doing since I am reliable in producing actions 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 actions 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 actions before the right kind of action is produced. Occasionally, of course, we might produce the wanted action right away but not usually. Shooting baskets is something most of us are only semi-reliable at. Finally, there are types of actions with respect to which many of us are *anti-reliable*: we generally do not succeed in producing the wanted action within a recognizable period of time. Juggling four balls is something most of us are anti-reliable at.

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 actions. Nor indeed would the

⁴ 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 we 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.

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 action types, is to suppose that his behavior is not interpretable (in terms of reasons) and that no concept of action is applicable (e.g. Anscombe 1963; Davidson 1980, 1984; Dennett 1981, 1987; von Wright 1971).

The situation is different, however, 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* action, the action that fits what we want, or in other words, the action that is *selected* from among the other (unsuccessful) actions. For instance, the agent who produced an action 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 action until he reached the one that fit what he intended⁵ (the selectional criterion) in which case he could say or think “There” or simply stop producing further actions.⁶

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 actions. Many broadly conceived

⁵ In Eastern-European languages, the non-Latin word for ‘intention’ is actually illuminating in this regard. The Polish ‘zamiar’ (intention, aim) and ‘zamierzyć’ (to intend, to aim), 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 understood as ‘to take for a measure’.

⁶ And it might be not only a matter of the agent’s discretion to decide which of the actions 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.

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 actions of which he selects the action 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 actions that he is disposed to select. This means that a reliable agent produces only actions that he would select.

At this point, one might be inclined to wonder what causes the agent to generate the actions. But I will leave this question open for now and return to it later (§4). 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 actions he produces, we can account for the 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 §4, 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.)

3. Acting for Reasons

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 that we allow for the distinction between acting for and acting with reasons), how should 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 this 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 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 (as selecting her actions in accordance with her reasons, understood as selectional criteria), we can accommodate the distinction between acting for a reason and acting with a reason.

We need one more concept, however — the concept of a particular action realizing, fulfilling or *satisfying* a reason. Suppose that Joe has two reasons to drink coffee. One of the reasons (*S*) is that coffee is a stimulant and he needs to stay awake for a couple more hours. The other reason (*L*) is that he simply likes its taste. Reason *S* will be satisfied by Joe's drinking coffee and becoming stimulated. It would be violated by Joe's drinking decaffeinated or very weak coffee and not becoming stimulated.⁷ Reason *L* would be fulfilled by Joe's drinking coffee that resulted in a pleasurable sensation (consistent with or constitutive of his liking coffee). It would be violated by Joe's drinking foul tasting coffee.

Given this notion of an action satisfying a reason, we can sketch what it would mean for an agent to act for a reason. Suppose that an agent α performed an action *A* for reason *R*. This means, roughly, that the agent α had a reason *R* to perform the action *A*, and had *A* not satisfied *R*, α would not have performed *A*. In other words, the agent selected the action *A* to accord with the reason *R*. The reason *R* functions, as it were, as a selection criterion for the actions the agent would perform. But we need to be a little more careful.

⁷ What actions are categorized as satisfying the reason will depend how we construe the reason. If the reason were construed as Joe's belief that the coffee he drinks is a stimulant then the actions that would satisfy that reason would include all those actions of drinking coffee where Joe holds the relevant belief whether truly or not.

Let us first assume that (r) the agent is reliable in responding to the reason R by satisfying it, and that (t) the agent has true beliefs regarding what actions satisfy reason R. Given those assumptions:

(A)_{(r)(t)} An agent α ϕ s because of R just in case (a) reason R is among α 's reasons to ϕ , (b) α ϕ s and his ϕ ing satisfies R, (c) α would have ϕ ed had his ϕ ing satisfied R but α would not have ϕ ed had his ϕ ing not satisfied R.

Since the idea of acting for a reason is always clearer when at least two reasons are involved. Let's consider a corollary characterization. Assuming that (r') the agent is reliable in responding to reasons R1 and R2 by satisfying them, and that (t') the agent has true beliefs regarding what actions satisfy reasons R1 and R2, we can say that:

(A')_{(r')(t')} An agent α ϕ s because of R1 rather than R2 just in case (a) reasons R1 and R2 are α 's reasons to ϕ , (b) α ϕ s, his ϕ ing satisfies R1 as well as R2, (c) α would have ϕ ed had his ϕ ing satisfied R1 (even if it would not satisfy R2) but α would not have ϕ ed had his ϕ ing not satisfied R1 (even if it would satisfy R2).

Example. 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 signaling by raising his arm is a viable option to do so. The second reason, *G*, is that by raising his arm he will be greeting his friend, which he wants to do also. Both reasons are satisfied by his raising the arm. More precisely, the first reason is satisfied by his signaling a turn, the second reason is satisfied by his greeting a friend, but raising an arm (in the appropriate circumstances) can count as doing both things. Why did the agent act, supposing that these were the only two reasons in play?

We need to consider appropriate counterfactual situations. The actual situation [SG] is one where the driver's raising his arm satisfies both reasons *S* and *G*. The three counterfactual situations will include: [S–G] a situation where only the reason 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), [–SG] a situation where only the reason to greet a friend would be satisfied by the agent's raising his arm (e.g., the driver passes his

friend long before the approach of the turn), [-S-G] a situation where neither of the reasons would be satisfied by the agent's raising his arm (e.g. the driver does not either pass his friend or approach a turn).

We assume that the driver has and would have true beliefs in all these situations and that he is reliable in responding to both reasons, i.e. that he is reliable in raising his arm.⁸

(1) If he were to raise his arm only in situations where he were approaching the turn [SG] and [S-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 [S-G]), but would not have raised it if he were not nearing the turn ([-SG], [-S-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 [-SG]), but would not have raised it if he were not passing the friend ([S-G], [-S-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 ([-SG], [S-G]), but would not have raised it if he were neither passing a friend nor nearing a turn [-S-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 [SG], but would not have raised his arm either if he were approaching a turn but not passing a friend [S-G], or if he were not approaching a turn though passing a friend [-SG], or neither [-S-G].

⁸ One might object at this point that the idea that the agent is reliable in responding to a reason already hides causalist content. I address this objection more fully in §4. At present, it is useful to bear in mind that reliability is not in fact necessary to make the distinction between acting for and acting with reasons, semi-reliability suffices.

(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 [-S-G], then he acted for neither of the reasons.

Let us generalize further by removing the simplifying assumptions. If we do not assume (t) (and (t')) that the agent has true beliefs concerning what actions satisfy the reason in question, then we need to consider the agent's beliefs. Suppose that John has reason to wake himself up and that he makes some coffee. What would it mean to say that John made the coffee in order to wake himself up? We can say that John acted for that reason just in case he would have made the coffee if *he believed* that it would help him stay awake but he would not have made the coffee if he believed that it would lead to the frustration of the reason (if he believed that yet another cup would make him drowsy).

Furthermore, we need not assume (r) that the agent is reliable in satisfying the reason. We may include cases where the agent is only semi-reliable in satisfying the reason in question. In such a case, rather than considering what action the agent would produce in appropriate counterfactual situations, we must consider explicitly which of the actions the agent might produce he would select.⁹

- (A) An agent α ϕ s because of reason R just in case (a) α has reason R to ϕ , (b) α selects his action p of ϕ ing, p satisfies R, and α believes that p satisfies R, (c) α would have selected p had he believed that it satisfied R but α would not have selected p had he believed that it would not satisfy R.¹⁰

Note that what is required by clause (b) is not only that the agent ϕ but that he actually select his ϕ ing. 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 action type. Just as it would make no sense to

⁹ To say that the agent is reliable is to say that he generates the performances he will select.

¹⁰ One might mistake clause (c) for a different clause: (c') α would have selected p had he believed that it satisfied R but α would not have selected p had he *not believed* that it would satisfy R. Clause (c'), but not (c), amounts to the thought that the agent selects the action *because* he believes it satisfies R.

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 action that he did not select (see also §6.3.).

Example. 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.¹¹ 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 action, he thinks to himself “Yes!” and goes on to the next scene. It may be useful to classify the actions he produces into four categories: *dt* — actions that are ambiguous between disdain and terror, *d-t* — actions that display more disdain than terror, *-dt* — actions that display more terror than disdain, *-d-t* — actions 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 action 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 actions 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 action *-dt* is to emphasize terror.

¹¹ It does not matter for our purposes whether the actor can learn to make the expressions at will. It will matter for the success of the play at least that he can learn to realize one of the three interpretations.

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 action 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 action dt , that has been actually selected by the agent, were different: if instead of dt the agent produced either of the three other action types. [DT] The actual last action satisfies D as well as T . [D-T] One counterfactual situation to consider is whether the agent would select the action if it expressed disdain but not terror ($d-t$), i.e. if it satisfied D but not T . [-DT] Another is to consider whether the agent would select the action if it expressed terror but not disdain ($-dt$), i.e. if it satisfied T but not D . [-D-T] The final question is what the agent would do if the action failed to satisfy both reasons (action of type $-d-t$).¹²

We can now see how we can determine what reason the agent acted for. We would say that he brought about the action dt to realize interpretation D , if he would select his action were it to satisfy D (i.e. dt or $d-t$) but he would not select the action were it to fail to satisfy D (i.e. $-dt$ or $-d-t$). In such a case, if he produced $-d-t$ or $-dt$, he would continue producing further actions until he managed either dt or $d-t$. Similarly, he acted (dt) to realize interpretation T , if he would select his action were it to satisfy 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 action were it to satisfy both T and D (i.e. dt) but not otherwise (i.e. $d-t$, $-dt$ or $-d-t$).

There are certain limitations to the applicability of the account. One limitation is brought out by considering cases involving two reasons. The account will apply as long as that the conditions of satisfaction of one of the reasons do not include those of the other. In cases, where this is not so only limited distinction of the efficacy of the reasons

¹² Note that in order to be interpretable, in case the agent were to produce a action of type $-d-t$, he would have to produce another action.

will be possible. Another limitation concerns the role of beliefs. I have considered what happens when an agent holds correct beliefs about whether an action satisfies a reason but I have not discussed what happens when the agent holds erroneous beliefs. While the exploration of these and other questions is important, I want to set them aside as they would take us too far off our main course. I want to but sketch an alternative way of looking at the relation between actions and reasons but there remain quite foundational questions to be answered.

4. “Yes, but why does he act in accord with the reason?”

The suggestion that we ought to conceive of the agent as selecting actions to fit 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 action of an agent is a causal process, it seems absolutely mysterious to suppose that the reason for which the agent acts, the goal that he intends to realize, is not somehow causally involved in the generation of the action. One way of putting the point is that while the selectional account simply has to assume that the agent is disposed to generate certain actions, 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 action of the right sorts of actions in the right sorts of circumstances.¹³

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*? How is the action generated? This is the first problem, the *problem of spontaneity*. The

¹³ This is essentially an objection that has been launched against G.H. von Wright’s account by his otherwise sympathetic reader Frederick Stoutland (1989, p. 323): “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’. 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.”

second problem is the flip-side of the first: If reasons are not causes then what causes the agent to act *in accord with them*? This is the *problem of congruence*. Here is another way of putting it: 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. However, there are many types of actions (like raising an arm, walking, etc.) for which we are reliable not just semi-reliable. How can one account for the fact that there are many types of actions with respect to which we are reliable unless one supposes that reasons are causes?

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 actions, just the actions 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. Let me begin with the first problem.

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. John Dewey has suggested long ago that we not forget, in our psychological thinking, what happened in physics:

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.
(Dewey 1957, p. 112)

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 frequently act even while possibly staying motionless – when we perform negative actions (omissions, for example).

Bearing this thought in mind ought to relieve the impression that the very notion of action becomes inert unless we conceive of reasons as causes. But it is not unreasonable to ask how actions are generated. I think that the most plausible answer is that there is no unified account to be given. Consider two kinds of generating causes.

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:

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. (Stoutland 1976, p. 319)

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 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 now why should we in addition identify the event that caused his action, which we already identified as his hearing the request, as his wanting to be (or perhaps as his being) polite?¹⁴ It seems that the only reason for this is that the causalist

¹⁴ This point has been also made, among others, by Child (1994) and recently, quite forcefully, by Hursthouse (2000). Child’s claim is actually too strong. He concludes that such events cannot be identified with the agent’s reasons for acting on the ground 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. It is sufficient for my purposes to draw a much weaker conclusion. Such cases do not *demonstrate* that not all reasons are causes, for they do not demonstrate that the generating causes could not

envisages the simple-causal model of the relation between actions and reasons as the only account of the relation. But it is not the only account of the relation and it is certainly not the only account of the relation which illuminates the causal processes involved.

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 in. 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. I could resist it but I let it generate the sequence of events that follows. This is a case where it is still plausible to think that I jumped in to swim my few laps, but where my desire to jump into the water is plausibly not thought to be the immediate cause of my action. (And here again the causalist will most certainly add that we could identify the desire with the agent's state. Perhaps we could. But we must ask why we should.)

This leads to a worry that the selectional account by leaving so much freedom in accounting for the generation for action, by not having a unified account of the generation process, misses out on an important virtue of the simple-causal picture. In view of our natural inclinations sketched above, it is unclear whether this "virtue" is not really a display of too cavalier an attitude at best. Simplicity is a virtue except in cases where it lets us forget about the complexities that still need to be understood. This is the virtue of the selectional account. It allows us to see how the possibly various generating explanations of the actions can come together with the processes of the agent selecting the actions to constitute the selectional complex explanation of an agent acting for a reason. In addition, there is a sense in which what unifies the various accounts of action generation is the agent. We sometimes think of *the agent* as causing, generating or

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. This is also the conclusion to be drawn from Hursthouse's discussion. The burden of proof accordingly lies with the causalist to argue why we would need such an identification.

producing the actions. This provides a general, though blanket, way of thinking about the generating mechanism. 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. 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 agentic function to the whole 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.

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.

Let me address the second problem, the problem of congruence. The selectional account of acting for reasons presupposes that the agent is either reliable or semi-reliable in responding to his reasons. However, 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 (R).

(R) For most agents, there are many actions for which the agent is disposed to produce bodily actions that she is disposed to select.

The question is *why* (R) holds.

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.¹⁵ 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.

¹⁵ “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” (von Wright 1971, p. 132).

Why do we by and large generate the bodily actions 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.¹⁶ But there are other ways of accounting for (R). In fact, I would like to suggest that there is not, nor need there be, a unified account of why (R) holds. Instead, we can invoke various considerations that support (R) and are compatible with the selectional (and more generally teleological) construal of reasons.

a) *Skill and Learning*. It is a contingent fact about us that we are capable of developing skills and of 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. It is simply a contingent fact about the actor who tries to produce the ambiguous grimace that whatever causes these actions, more and more of them come to resemble the target action. It is a contingent fact about people less talented in this area that no matter how hard and long they try they will not be able to do it. It is a contingent fact about most of us that we manage to acquire many skills that are basic to our culture (in part for that very reason). Still, if that were not a contingent fact about us we *could* act, actions would be possible, they would just look different – we would have to tolerate agentive stutter more than we do.

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

¹⁶ 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 ϕ ing) without relying on any finite number of types of physical causal relations. (The worry was first formulated by Stoutland 1976; 1980. For further discussion, see Heil & Mele 1993.)

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 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.

b) *Some Causes May be Reasons.*¹⁷ 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, 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.

c) *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

¹⁷ I am assuming here that reasons are to be identified with reason-states. It is possible to offer a theory of reasons (as facts, for example), where this does not make sense.

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. 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. 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.

d) *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 (Cohen 1978; Nowak 1975, 1989). 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 desires but because it increases their chances of surviving in competition with rulers who do not hold such beliefs.

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 actions 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 actions that are appropriate in such a situation.

This does not establish the truth of (R) 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.

I have thus argued that other explanations are available to account for the problem of spontaneity and the problem of congruence than the simple causalist explanation relying on the thought that reasons are causes. One thing ought to be clear. In all (or

most¹⁸) of the above cases, the picture that emerges is thoroughly *causal*. My passing salt to you is *caused* by my hearing your request. It is an exercise of my *causal* skill to pass salt (or other objects). My passing salt will be selected by me as I *causally* form the belief “Yes, that is what I wanted to do.” And, crucially, the action would have been selected by me in appropriate counterfactual situations and would not have been selected in others. My jumping into the water will be *caused* by my mother’s gentle push. It will have triggered a sequence of motions which constitute an exercise of my *causal* skills of jumping into the water and swimming. And I will have selected my total action by recognizing it “Oh, all right! Here I go.”¹⁹ Again, crucially, the agent would have selected the action in appropriate counterfactual situations and not in others. To underscore that the processes involved are causal we might call this picture the *selectional-causal picture* of acting for a reason.

Two things should be clear, however. First, although the processes involved are causal, they are not simple causal relations between the reasons and actions envisaged by the simple causalist picture. The relation between actions and reasons emerges from the operation of the selectional apparatus which involves actual and counterfactual causal relations between various states of the agent. Thus, the simple causalist has been right to insist that causal relations are involved but has been wrong to suppose that they are as simple and straightforward as he suggests. Second, if one compares the teleological and the causal accounts of acting for reasons, the teleological accounts are closer to the truth. Reasons are properly understood as the aims involved in the agent’s selection of a action. What they have been unnecessarily shy about is admitting the importance of causal relations. The virtue of the selectional account is that it provides a framework where the intuitions driving both opposing camps can be allotted appropriate place.

¹⁸ The hesitation here pertains to what one thinks about the functional explanations.

¹⁹ As it stands, the case is one where the agent wanted to jump into the water and it does not matter to her what exactly the triggering cause may be. But one could describe a case where it does. The agent my jump into the water as a result of the push and then quickly climb out and start an argument with the mother. When the mother points out that the daughter could have not jumped, the daughter may still object “I just did not pay attention at this moment. I let myself jump in willy-nilly. But I wanted to do it all by myself!”

5. Reasons as Selectional Criteria rather than Causes

Let me close with the following consideration that favors the selectional-causal over the simple-causal view. The causal and the selectional view have at their center certain counterfactual claims. On the causal view, we must attend to what would happen had the agent not had a certain reason to act. On the selectional view, we must attend to what would happen had the agent's reason not been satisfied by the action generated. These are very different counterfactual claims. I will try to show that the selectional counterfactuals are in fact closer to the counterfactuals we would intuitively attend to in trying to determine whether an agent acts for a given reason.

Considering an intuitive example where two reasons constitute possible explanations of why a certain action was taken. Suppose that one academic department has hired one person out of many candidates. The candidate was a qualified woman. As it turns out, the department had both a reason to hire a woman candidate (it was under pressure from the EO/AA office) as well as a reason to hire a qualified candidate (the department aspires to being a good department). Both constitute reasons for the department to hire the successful candidate.

What we are trying to understand is what it means to ask the question, Which of the reasons was operative?, On what reason(s) did the department act? Let us begin by sketching intuitively what we ordinarily take the questions to mean. (To simplify, I am assuming that only those two reasons are in play. I will also skip over the possibility that the department acted on both reasons.) Intuitively, the department will have chosen the candidate simply because of her gender if it would have chosen her over more qualified male candidates. The department will have chosen the candidate because of her qualifications if it would not have chosen her had there been better qualified male candidates. So:

- (f) (Even) if there were a more qualified male candidate, the department would have chosen the female candidate it has chosen.
- (q) If there were a more qualified male candidate, the department would have chosen him.

Consider the selectional understanding of the case. I will assume that the department has true beliefs about whether the candidates satisfy the reasons. I will assume that the department is reliable in responding to the reasons.²⁰ We can then say that the department chose the candidate because of her gender if:

(F_S) had the candidate not been female (in which case the department's possible choosing of such a candidate would not have satisfied reason F), the department would not have chosen to hire her.

(F_S) does capture the intuition expressed in (f) that the department will choose women over men.²¹ The department will have chosen the candidate because of her qualification if:

(Q_S) had the candidate not been qualified (in which case the selection would have violated reason Q), the department would not have chosen to hire her.

But (Q_S) does not quite capture all that (q) says. (Q_S) is more restrictive because it talks about the candidates that were actually chosen, while (q) mentions other candidates that there could have been. But it seems clear enough that there is a way of bringing (Q_S) to bear out the content of (q) by adding another clause. So:

(Q_{S1}) had the candidate C not been qualified, and had there been another candidate C' who was qualified, the department would not have selected C but C' instead.

(Q_{S2}) had the candidate C not been qualified, and had there been no other qualified candidate, the department would not have selected C or anyone else.

²⁰ This assumption seems inherent in the very process of departmental decision-making (given the usual rules). The decision that the members finally agree on will be the one that is selected as the appropriate action. It is hard to imagine how the department could "stutter" (unless the rules of conduct were violated: the decision was made by only a few members behind the others' back, for example).

²¹ In fact, it is arguable that (F_S) is stronger than (f). According to (F_S), the department will not have chosen any candidates at all unless they were women. And this is indeed what it would require for the department to act solely because it wanted to hire women candidates. The reason why (f) is weaker is that it expresses our intuitions about real cases, and in reality the reason to hire women is never the only reason the department acts on, it is always superimposed on some criterion of qualification. In most real cases, while the department might prefer a woman candidate to a man candidate, the preference will be exercised on the pool of candidates that went through a preliminary screening of their qualifications, for instance.

The sense of (q) is captured by (Q_{SI}), where it is in addition supposed that the counterfactual would hold true even if the other candidate were male. We can thus see that the counterfactuals characteristic of the selectional construal of reasons are close to the counterfactuals we would consider intuitively. This is not the case for the counterfactuals dictated by the simple causal view.

If the efficacy of reasons were to be construed causally, presumably the following counterfactuals would have to hold. The department chose the candidate because of her gender (reason *F*) if

(F_c) had the department not had reason *F*, it would not have chosen her.

The department chose the candidate because of her qualifications (reason *Q*) if

(Q_c) had the department not had reason *Q*, it would not have chosen her.

The causal construal not only does not illuminate (f) and (q) but is too strong.

Let us first note that the counterfactuals are actually false. Let us suppose that the department acted on reason *F*. In such a case, it is too strong to require that had the department not had the reason to hire women candidates, it would not have hired the female candidate. After all, it could so happen that the candidate it chose for other reasons (for instance, because she was qualified) also *happened* to be a woman. Similarly, even if the department did not have a reason to hire qualified candidates, it still could *happen* that candidates it hired were qualified.

The failure of the causal (counterfactual) construction of the efficacy of reasons can be thought to be a consequence of the well-known problem with counterfactual accounts of causality, the problem of overdetermination. But the falsehood of (F_c) can be also thought to rely on a notorious scope ambiguity. There is a difference between not having a reason to do something and having a reason not to do something. And this difference accounts for the falsehood of (F_c). For while it is false to say that:

(F_c) had the department *not had* the reason to choose female candidates, it would not have chosen the female candidate in question,

it is *ceteris paribus* quite true that

(F'_c) had the department *had* a reason *not* to choose female candidates, it would not have chosen the female candidate in question (no matter what her qualifications).

But the latter counterfactual is true only if the reason is actually operative which brings us back to where we started. The problem, of course, is that a causal analysis requires the consideration of what would happen were the event not to occur rather than were some other event to occur.

A perhaps more serious problem is that the causal counterfactuals are the wrong counterfactuals to consider. Our intuition is that the department chose the candidate because she was a woman if, even if there were more qualified male candidates, it would have still chosen her. The possible worlds picked out by (f) are ones where there are more qualified male candidates, but in *all* of them, the department is still supposed to *have* the reason to hire female candidates, the department is still under the pressure from the EO/AA office. By contrast, the possible worlds picked out by (F'_c) are ones where the department *lacks* a reason to hire women candidates. There is a fundamental difference between the causalist rendition of what it means for the reason to be operative and our intuitions on this score. Entirely different counterfactual situations are involved. The causal rendition of what it is for the department's reason to have been operative does not illuminate what the department would do had there been more qualified male candidates at all.

The same holds for the causal rendition of what it means for the department to act for the other reason, to have chosen the candidates because of their qualifications. On the causal rendition, we have to imagine worlds where the department does not have the reason to choose candidates because of their qualifications. But again this amounts to a consideration of a very different sort than our intuitions suggest. Our intuitive considerations suggest that we should consider worlds in which there were male candidates who were more qualified than the successful candidate. If the quality of the candidates were what mattered to the department, if that were the reason because of which the department chose the candidates, in such worlds the department would have chosen candidates that were better qualified (even though they were men).

It may be objected at this point that I have made a straw-case for the selectional account because I have been abstracting from the epistemic rendition. But once it becomes clear that it is the department's *beliefs* that are relevant to the action, the causal account will fall into place. For consider the epistemic rendition of the selectional counterfactual (F_S):

(F'_S) had the department not believed that the candidate is a female (in which case the department's possible choosing of such a candidate would have violated reason F), the department would not have chosen her.

But now the objection to the causal account that it considers the wrong kind of counterfactuals vanishes. For had the department not believed that the candidate is a female it would have no reason to choose her. Thus the antecedent of this counterfactual is close, contrary to the claims I made above, to the causal rendition (F_C). The causal counterfactual does after all pick out the possible worlds that it should pick out.

This is only an appearance, however. A different notion of "not having a reason" is in play. The sense in which we say that if the department did not believe that the candidate was a woman it would *have no reason to* choose her is "generic." It does not require our supposing that the department *lacks the reason to* choose a woman. To the contrary, the department can be still under pressure from the EO/AA office and can still want to hire a woman. It is just that it believes that a given case will not satisfy that reason. In other words, the department *has no reason* (in the generic sense) to act on its reason, on the reason it *does* have.

So, the point that the causal analysis requires us to imagine very different counterfactuals stands.

The causalist might object that this is not his turf. What kind of sense can we make of *a department* having a reason? Indeed what kind of a *causal* sense can we make of it? I take it that to the extent that these questions are hard-pressed for an answer,²² the

²² It is really not so clear whether they are or are not. Too little attention has been paid to the question what the causalist ought to say about agents other than biological units acting for reasons. A rare exception is French (1985).

fact that the selectional account handles these cases without a stretch is a virtue not a vice. Whatever stretches causalist philosophers experience, ordinary folks don't. Still, one may worry that the above example plays into the hands of the selectional account. So, let us consider a different example.

Consider the case of the driver who waves his arm, and has two reasons to do it: to greet the friend and to take the turn. I argued that we should say that he waves the arm to greet the friend just in case

(D_S) he would have waved in the cases when (he believed) the friend was there and would not have waved otherwise in cases when (he believed) his friend was not there but he was nearing a turn.

On the causal account, we ought to say that he waves the arm to greet the friend just in case

(D_C) had he not had the reason (to greet the friend) he would not have waved his arm.

As before, (D_C) is false because he could have waved his arm despite lacking the reason to greet the friend. (Again, we may believe a related counterfactual: had he *had* a reason *not* to greet the friend, he would not have waved the arm – as long as he believed that his waving the arm would amount to greeting the friend.) As before, the causal counterfactual considers the wrong kinds of possible worlds. Rather than picking out worlds where the driver lacks a reason to greet his friend, the worlds that should be picked out are those where the driver has a reason to greet his friend but believes that given the state of the world (his friend not being there, for example) his reason will not be satisfied.

So far I have been talking about two intuitive cases. It should be clear, however, that there is but a short step to a stronger conclusion here. For if the above analysis of the cases is correct then it follows that even if it were the case (implausible though it is) that reasons always *actually* caused actions, this would be insufficient to ground the distinction between acting for reasons and acting while merely having them. That reasons happened to cause an action would be simply a coincidental fact which had little to do with what we meant when we said that the agent acted for a reason. In such a case the

causal account would happen to apply to the right sorts of cases but it would still fail as an *analysis* of them. What we *mean* when we say that an agent acts for a reason *is not* that the reason causes the action even if it actually does happen to cause the action.

6. Some Further Objections

6.1. *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 actions 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 actions, they must be construed as causing the agent to select the right actions.

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, we do not need to think of the agent's selecting the action as the agent's action done for a reason. The agent who selects the action merely *recognizes* it as fitting his reasons. The only resources required for an agent to select a action as according with a reason are for him to have a conception of a action satisfying a reason. To the extent that the agent has a conception of a action fitting a reason (i.e. is capable of correctly sorting actions that fit the reason and those that do not), he is able to select a given action 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 action is simply to apply his conception of the reason.

6.2. *The problem of Action.* It might be objected that one should not dispense with the idea that reasons cause actions since it is necessary to account for the distinction between action (raising one's arm) and mere happening (arm rising). – I believe that the presupposition here is false. The idea that reasons cause actions is not necessary to account for that distinction either (see my 1998; for an alternative account of the distinction, see 1997). That aside, however, a selectionist need not reject the idea that reasons cause actions (or even the idea that reasons *always* cause actions).²³ His only commitment concerns grounding the distinction between acting for and acting while merely having a reason. And it is here that he claims the idea that reasons cause actions is not necessary.

6.3. *Trying for a Reason.* The account seems to preclude the very possibility of our trying to do something for a reason. – No. The account of acting for reasons I offered presupposes only that the action is actually selected. It therefore does not apply to the actions the agent rejects (if any) before selecting one. It seems intuitive to describe such actions as the agent's attempts or tryings. But these are not all the tryings there are. Some tryings are *selected* and to those the account does apply. Very frequently, tryings can be redescribed in success terms. "I tried to call last night to let you know that I will not come." The trying here can be presumably redescribed in such terms as "I dialed your number, but I could not reach you since it was busy". In such a case, there is no reason at all why the account should not apply to tryings.

However, if trying is merely a failure term (the action is not actually selected by the agent)²⁴ then reasons will not apply to tryings unless in a derivative sense. There is no

²³ I do, however, believe that this is an implausible position to occupy, as I said in §4.

²⁴ Jennifer Hornsby (1995) offers examples of reasons to try to ϕ where the reasons are not the agent's reasons to ϕ . She imagines an agent who is convinced that a certain weight is too heavy to lift for her, so she tries to lift it just to show the impossibility of doing so. Here she has a reason to try to lift the weight but not to actually lift it. But note that this is not really a case of failure trying in the above sense. Here there are certain patterns of actions (which we conjointly call "trying to lift a weight"), which are selected

full-fledged *reason* why the stutterer utters the stuttering syllable ‘a-’ (in ‘a- a- a- a- apple’). It is not appropriate for him to *justify* this action by appealing to a *reason*. One might object here that we can surely say “But the reason why the stutterer uttered ‘a-’ is that he was trying to say ‘apple’.” It seems clear, however, that the use of ‘reason’ here is derivative at best. At any rate, it is not the same sense of ‘reason’ as above. This new piece of information does not show us how to justify the action at all – it does not provide any reason why we should think that the stutterer *ought to* have uttered ‘a-’. Rather, it simply explains the action as a failed attempt in sequence of actions which ends with a successful utterance of ‘apple’ for which utterance the stutterer may have had good reasons.

7. Conclusion

It has been by and large accepted since the appearance of Davidson’s classic paper that only a causal theory of action explanation can give an account of the explanatory force of reasons (codified in the distinction between acting for a reason and acting while merely having one). The teleological attempts to demonstrate that the distinction is born out by teleological explanations are sometimes dismissed as not providing *any account* at all. I have given some reasons to support the following claims. First, I have argued that, contrary to appearances, whatever the causal account is an account of it is *not* an account of the explanatory force of reasons. If we trust our intuitions about the counterfactuals involved in deciding whether an agent acted for a reason or not, they are not the causal counterfactuals (§5). Second, I have proposed a selectional-causal rather than a simple-causal understanding of the teleological relation between actions and reasons (§2 and §3). On this view, reasons are seen as selectional criteria in accordance with which the agent selects (and would select) the actions she

by the agent. These are cases where they would constitute success. As such, the agent can have reasons for doing so and act on them. One might wonder why we would not be inclined to redescribe such cases in success terms. The most plausible reason is that these are cases where the agent believes that the world will not (not just might not, as is usual) cooperate in the action – she believes that the weight is too heavy for her to lift.

generates. One of the virtues of this account is that it allows one to preserve the virtues of both the teleological and the causal views. If reasons are selectional criteria then their function is akin to the end for which the action is undertaken, not to the cause that generates the action. At the same time, the selectional model allows one to identify the various causal elements that are presupposed for the proper functioning of selectional systems that we are: the generating mechanism, the selecting mechanism and the adaptational mechanism. The existence of the first and the third, in particular, could be and has been thought to provide reasons for preferring the idea that reasons are causes. I have argued, however, we can account for generation and for adaptation on other grounds as well (§4).

My general suspicion is that we are in fact operating at various levels when we try to give a causal interpretation of action. What speaks for causality is the mechanical nature of the cogs in our system but at the same time we operate with so abstract a notion of reasons that it is almost arrogant to think that one can easily relate these two. The dialectic of the causal theory is a constant conversation of these two levels without our knowing which speaks when. In presenting this possibility I offer it as a sophisticated *causal* alternative. (Although I have been putting it as an alternative to a causal interpretation of the force of reasons, it is only opposed to what I have been calling *simple causalism*.) And I offer it as a sophisticated *teleological* alternative since reasons here are fundamentally seen as ends. Selectional explanations demonstrate that there is no incompatibility here and provides one model of effecting such a reconciliation.

I have argued that the causalist perspective ingrained in us tends to obscure the fact that we are semi-reliable with respect to many more actions than we would expect. But there is a flip-side to this point. Suppose that we were perfectly reliable with respect to all actions. If this were the case, then the causal theory of action would look even more natural than it does nowadays. If, on the other hand, we were almost never reliable but mostly semi-reliable then it is hard to see how the causal theory of action could have any appeal. It would look either as if the reason causal state were lucky in producing the action or as if the theory were methodologically lucky (not to say ad hoc) to identify as reason the state that produces the successful action not the ones that produce the

unsuccessful performances. If knowledge were power one could say that it is in the vested interest of the causal theory of action to focus our attention on just those actions that we are reliable for. It would not be surprising perhaps that round and about the time when the causal theory of action has come to popular favor, philosophers have paid much attention to basic actions, to bodily actions, and have turned their attention away from omissions, collective actions, consequences of actions, social actions. Sometimes the stutterer has a better chance at expressing the truth.

REFERENCES

- Anscombe, G.E.M. (1963) *Intention. 2nd edition.* Cornell University Press, Ithaca.
- Child, W. (1994) *Causality, Interpretation and the Mind.* Clarendon Press, Oxford.
- Cohen, G.A. (1978) *Karl Marx's Theory of History. A Defence.* Princeton University Press, Princeton.
- Collins, A.W. (1987) *The Nature of Mental Things.* University of Notre Dame Press, Notre Dame, IN.
- Davidson, D. (1963) "Actions, Reasons, and Causes," in (1980), pp. 3-19.
- Davidson, D. (1980) *Essays on Actions and Events.* Clarendon Press, Oxford.
- Davidson, D. (1984) *Inquiries into Truth and Interpretation.* Clarendon Press, Oxford.
- Dawkins, R. (1986) *The Blind Watchmaker. Why the Evidence of Evolution Reveals a Universe Without Design.* Longman, Harlow.
- Dennett, D.C. (1981) *Brainstorms. Philosophical Essays on Mind and Psychology.* Bradford Books, Cambridge, MA.
- Dennett, D.C. (1987) *The Intentional Stance.* Bradford Books, Cambridge, MA..
- Dewey, J. (1957) *Human Nature and Conduct.* The Modern Library, New York.
- French, P.A. (1985) "Fishing the Red Herrings Out of the Sea of Moral Responsibility," in LePore, E. and McLaughlin B.P. (eds.) *Actions and Events. Perspectives on the Philosophy of Donald Davidson,* pp. 73-87. Basil Blackwell, Oxford.
- Heil, J. and Mele, A., eds. (1993) *Mental Causation.* Clarendon Press, Oxford.
- Hintikka, J., ed. (1976) *Essays on Wittgenstein in Honor of G.H. von Wright. Acta Philosophica Fennica 28.* North-Holland, Amsterdam.
- Hornsby, J. (1995) "Reasons for Trying," *Journal of Philosophical Research,* 20: 525-539.
- Hursthouse, R. (2000) "Intention," in Teichman (2000), pp. 83-105.
- Melden, A.I. (1961) *Free Action.* Routledge & Kegan Paul, London.
- Nowak, L. (1975) "Theory of Socio-Economic Formations as an Adaptive Theory," *Revolutionary World,* 14: 85-102.

- Nowak, L., ed. (1989) *Dimensions of the Historical Process. Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 13.* Rodopi, Amsterdam.
- Peters, R.S. (1958) *The Concept of Motivation.* Routledge & Kegan Paul, London.
- Schilpp, P.A. and Hahn, L.E., eds. (1989) *The Philosophy of Georg Henrik von Wright. The Library of Living Philosophers XIX.* Open Court, La Salle, IL.
- Sober, E. (1984) *The Nature of Selection. Evolutionary Theory in Philosophical Focus.* The MIT Press, Cambridge, MA.
- Stoutland, F. (1976) "The Causation of Behavior," in Hintikka (1976), pp. 286-325.
- Stoutland, F. (1980) "Oblique Causation and Reasons for Action," *Synthese*, 43: 351-367.
- Stoutland, F. (1989) "Von Wright's Theory of Action," in Schilpp & Hahn (1989), pp. 305-332.
- Taylor, C. (1964) *The Explanation of Behaviour.* Routledge & Kegan Paul, London.
- Taylor, R. (1966) *Action and Purpose.* Prentice-Hall, Englewood Cliffs, N.J..
- Teichman, R., ed. (2000) *Logic, Cause and Action. Essays in Honour of Elizabeth Anscombe.* Cambridge University Press, Cambridge.
- von Wright, G.H. (1971) *Explanation and Understanding.* Cornell University Press, Ithaca.
- Wilson, G.M. (1989) *The Intentionality of Human Action.* Stanford University Press, Stanford.