INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

SOCIAL ANATOMY OF ACTION:

TOWARD A RESPONSIBILITY-BASED CONCEPTION OF AGENCY

by

Katarzyna Paprzycka

A.B.. Radcliffe College. 1989

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

1997

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Katarzyna Paprzycka

It was defended on

December 4, 1997

and approved by

Nuel Belnap

John McDowell

Nicholas Rescher

Merrilee Salmon

Robert Brandom

Committee Chairperson

SOCIAL ANATOMY OF ACTION.
TOWARD A RESPONSIBILITY-BASED CONCEPTION OF AGENCY

Katarzyna Paprzycka

The dissertation develops a conception of action based on the concept of practical task-responsibility (understood in terms of normative expectations) rather than on the concept of intention. It answers two problems. First, it grounds the distinction between an *action* and a *mere happening*, thus meeting Wittgenstein's challenge to explain what is the difference between an agent's raising her arm and her arm rising? Second, it grounds the distinction between *acting for a reason* and *acting while merely having a reason*, thus meeting Davidson's challenge to give an account of the explanatory force of reasons.

One source of resistance to an account of action in terms of normative expectations rather than in terms of intentions comes from explanatory individualism. The explanatory individualist argues that it is ultimately the intentional attitudes of the agent rather than normative expectations of other people that are relevant to the way in which we explain one another's actions, and that they ought to figure in the account of the nature of action. I defend explanatory nonindividualism, according to which we sometimes act on our own intentions and desires but sometimes on the normative expectations and desires of others (without thereby acting on our own pro-attitudes). Explanatory nonindividualism is fortified by a selectional account of the explanatory force of reasons. I demonstrate that Davidson's challenge can be met by identifying reasons with *selectional criteria* rather than with causes.

In response to Wittgenstein's challenge, I propose that we understand what it is to be an action not in terms of a performance being intentional under a description, but in terms of *it being reasonable to expect* a performance of the agent under some description.

Only one concept (of two) of reasonableness is necessary to make the distinction between actions and mere happenings, and I explicate that concept. In addition to accounting for actions that are intentional under some description, the account also captures cases that are not so straightforwardly captured by that criterion (e.g., spontaneous actions, unintentional omissions). Moreover, cases of so-called basic wayward causal chains are excluded from qualifying as actions.

ACKNOWLEDGEMENTS

Philosophers who have been most influential in the development of the theory are the least often cited in this dissertation for the simple reason that their influence goes far beyond a well-defined cluster of ideas. A series of papers on the philosophy of action by Annette C. Baier is suggestive of much of the flavor of the approach to action and to action explanation presented here. Nuel Belnap's work on *stit* has been a constant companion in the development of my account, sometimes fostering sometimes thwarting its pace, but always inspiring. I am very grateful for the many long conversations.

I owe more than I can say to Bob Brandom. My work on action has started as an attempt to offer a more Brandomian theory of action than Brandom's own. In time, the conceptual framework has departed from Brandom's substantially (necessitated by the different demands of the two theoretical endeavors). However, a careful glance at the structure of the account reveals a deep affinity of the apparatus here developed to Brandom's theory of language. Throughout the process Bob watched carefully that I do not go overboard in making wild claims at times of euphoria, and that I do not give up on too many of them at times of crisis. I am deeply grateful for his unfailing encouragement and enthusiasm, especially at the more gloomy times.

I would like to express my deep gratitude to John McDowell for the time and care he took in reading the multiple drafts and for his firm disagreement with many of the ideas here advanced. His voice is behind many of the objections to nonindividualism I have discussed and will be behind many more I am yet to conceive. I am also grateful to Merrillee Salmon for her critical eye on the empirical claims I have been making as well as for her support, and to Nicholas Rescher for keeping my philosophical feet on the ground. Kurt Baier has helped by providing a guided tour through the difficult topic of responsibility.

An external but constant presence in the conceiving and writing of the dissertation was my father, Leszek Nowak. The project was intended as a way of escaping his enormous influence over my thinking. Toward its end, I am proud that I have not succeeded entirely.

My greatest debt is to my husband, Marcin Paprzycki. I am grateful for the times spent on discussing half- and quarter-baked ideas, for his careful reading of the drafts, but most of all for his unending support and patience. It is no wonder that all the major breakthroughs in the development of my ideas occurred while we were together.
I dedicate the dissertation to him.

TABLE OF CONTENTS

INTRODUCTION

The concept of action, as opposed to mere happening, lies at the intersection of two areas of philosophical interests: ethics and philosophy of mind. Moral philosophers, in particular those interested in questions of moral responsibility, use the concept of action as a given. It is the job of philosophers of mind to analyze the concept for among others such uses. The dissertation proposes such an analysis. It offers a systematic answer to Wittgenstein's question. What is the difference between my raising my arm (an action) and my arm rising on its own (a mere happening)?

1. Action as a Unit of Conduct

Perhaps the most fundamental difficulty in analyzing the concept of action is the fact that it plays a significant role in a number of disciplines as diverse as physics, biology, psychology and sociology. As a result the concept has coalesced a great variety of intuitions. It is thus important to at least try to distinguish some ways in which the concept can be applied.[1]

(i) There is a concept of *inanimate* action. When a billiard ball thrusts into another billiard ball it acts on the other. To its action, by Newton's third law, there corresponds an appropriate reaction of the other ball. At this stage, teleological concepts apply only derivatively. For example, we can speak of the purpose or function of a piece of a thermostat, but its purposefulness is derived from its being designed.

(ii) We speak of the actions of various parts of animal bodies. This is the first stage at which non-derivative teleological concepts find application. The liver's

---

[1] This division is suggested by Harry Frankfurt in "The Problem of Action," in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), pp. 69-79. Frankfurt identifies action with intentional movement.

excreting bile, the heart's pumping are examples of what one might call *purposeful* movements or actions.

(iii) The third level is that of *purposive movement* or action. The subjects of our attributions of purposive movements are no longer parts of bodies but rather agents. The movements a spider produces in spinning a web constitute purposive movements. In this sense also, a drug addict's compulsively taking a shot is purposive.[2] Arguably, sleep-walking, some actions performed under hypnosis, as well as the little movements one performs to alleviate muscle pain in one's sleep are purposive. So are feeding the cat, conversing, looking out of the window, walking through a forest.

(iv) The latter but not the former examples belong to a more restrictive category of *intentional movements*. A movement is intentional just in case there is some description under which it is intentional. The category of intentional movements is an extensional category — it picks out a class of events. As such, it is a very different concept from the concept of intentional action, which does not pick out a class of events.[3] Both intentional and unintentional actions, as they are usually understood, are intentional movements in this sense.[4]

It is not uncontroversial to sharply distinguish the category of intentional movements from the category of purposive movements. One might treat the distinction to be one of degree rather than principle. Yet many of the examples relevant here are at least very different from the ones of purposive movement. So when one deliberately goes to a rally, one performs an action of a different sort than if one went there in one's sleep.

It seems uncontroversial that actions of the first two sorts (i) and (ii) do not constitute the subject of interest to philosophers concerned with understanding the

---

[2] Ibid., pp. 76-77.

[3] The received view is that there is no class of intentional actions (G.E.M. Anscombe, *Intention* [Ithaca: Cornell University Press, 1957]; Donald Davidson, *Essays on Actions and Events* [Oxford: Clarendon Press, 1980]). Rather actions are intentional under some descriptions.

[4] This usage of the term 'intentional movement' is not widespread. (The term is used explicitly by Frankfurt "The Problem of Action," *op. cit.*) It is more usual in the literature to treat the term 'action' in the way I am stipulating to use the term 'intentional movement'. However, since I will advertise a different set of intuitions to coalesce around the term 'action', I shall use the term 'intentional movements' exclusively in the extensional way suggested and contrast it sharply with 'intentional action'.

phenomenon of human action. The examples that have been taken to be paradigmatic examples of action belong to the fourth category of intentional movements. In the present dissertation, our topic will be yet another understanding of the concept of action, action as a unit of our conduct.

(v) The fifth sense of 'action' derives from the idea of an agent's overall conduct. Someone's conduct includes her intentional and unintentional doings but also intentional and unintentional not-doings (omissions). When we inquire after a person's conduct during a rally, say, we will be interested in the things the person said and did as well as the things that he omitted to say or do. The concept of action as part of an agent's conduct has not been at the forefront of philosophers' concern with agency.[5] Most of the debate has centered around the concept of action in the sense of purposive and/or intentional movement. This is among, other things, because intelligence and reason are most clearly manifested in our acting intentionally. But the philosophical focus on "intelligent agency"[6] should not lead one to think that there is nothing interesting about action but for its rational significance. In fact, there are psychological categories that pertain to our conduct rather than merely to our intentional behavior. The most important among them is the concept of character. Character comprises not only agentive voice — active intentional rational excursions into the world — but also idleness, passivity, thoughtlessness, carelessness, forgetfulness — agentive silence, as it were.

I will try to capture the fifth sense of the concept of action in this dissertation. My aim is to acquire a deeper understanding of the sense in which we *do* things when we act intelligently, intentionally, rationally, but also when we act carelessly, when we keep to ourselves, when we *do* not do anything. Henceforth, when I use the word 'action', I will mean action in the last sense, action as a unit of our conduct rather than the way in which it is used in most of the literature — as a unit of our intentional or purposive behavior.

---

[5] The most obvious exception is H.L.A. Hart who frequently speaks of the "philosophy of conduct," intending to cover both actions and omissions (including unintentional ones) by the term. See "The Ascription of Responsibility and Rights," in (ed.) Anthony Flew, *Essays on Logic and Language* (Oxford: Blackwell, 1951), pp. 145-166; *Punishment and Responsibility* (Oxford: Oxford University Press, 1968); *Causation in the Law* (Oxford: Clarendon Press, 1985).

[6] The phrase is Michael Bratman's, see his "Moore on Intention and Volition," *The University of Pennsylvania Law Review* 142 (1994), p. 1708.

This means that one of the immediate criteria of adequacy that are imposed on the account of action here developed is that it apply not only to intentional and unintentional actions (intentional behavior) but also to intentional and unintentional omissions.

Since most philosophers of action do not undertake the task of developing an account of action that would encompass unintentional omissions,[7] and since accordingly few accounts of action apply to unintentional omissions, I should pause to emphasize the nature of my theoretical intention and of others' omission. It is indisputable that if a theorist of action intends to capture the concept of action (understood as a unit of intentional behavior) then unintentional omissions simply do not belong to that theorist's domain of interest. Given how common it is to understand actions as units of intentional behavior (and there are good reasons for it), we should at least foresee the possibility of the following objection arising. Such a theorist might acknowledge that unintentional omissions can be conceived as part of our conduct, but refuse to allow that there is any sense of the concept of action that would cover such cases.

It is very difficult to answer such an objection in a persuasive way since most of the considerations are pre-conceptual or pre-theoretical. There is certainly no argument that would force us to acknowledge that there is a sense of agency involved in our unintentionally omitting something. There is no argument but there are reasons.

For one the concept of conduct plays a significant role in our psychological understanding of the world. This is evident in at least two ways. First, while our understanding of people's characters does include their intentional behavior, it covers more than just their intentional actions and their unintended consequences. Among

---

[7] There are important exceptions. See H.L.A. Hart, "The Ascription of Responsibility and Rights," *op. cit.*; *Punishment and Responsibility, op. cit.*; Steven Lee, "Omissions," *Southern Journal of Philosophy* 16 (1978), 339-354; Patricia G. Smith (Milanich), "Allowing, Refraining, and Failing. The Structure of Omissions," *Philosophical Studies* 45 (1984), 57-67; "Ethics and Action Theory on Refraining: A Familiar Refrain in Two Parts," *The Journal of Value Inquiry* 20 (1986), 3-17; "Contemplating Failure: The Importance of Unconscious Omission," *Philosophical Studies* 59 (1990), 159-176. There are also theorists of action who are simply uninterested in giving an account not only of unintentional omissions, but also of the less controversial negative actions. Carl Ginet declares at the beginning of his book: "...Among the nonactions are such items as not voting in the election, neglecting to lock the door, omitting to put salt in the batter, and remaining inactive. Such things have been called *negative actions*, largely because they can be the objects of choices and intentions. But they are not actions in the sense I am interested in..." (*On Action* [Cambridge: Cambridge University Press, 1990], p. 1).

character traits we mention also traits that comprise the agents' tendency to commit omissions (including unintentional omissions). Carelessness, forgetfulness, idleness, reserve are just some of the examples.[8] Second, we are usually held responsible for the way in which we conduct ourselves, and that includes our being responsible not only for our intentional actions and their unintended consequences but also for our unintentional omissions.[9] When I stand up a friend of mine because I simply forgot that we were to meet in the library, she will rightly hold me responsible for my failure to show up. The fact that my forgetting was unintentional does not make me any less responsible for wasting my friend's time.

These are some pre-conceptual reasons for believing that the concept of conduct can aspire to capture some of our intuitions about agency. The rest of the dissertation ought to provide additional reasons.

2. Two Main Problems

My primary aim in the dissertation is to give answers to two problems that have concerned philosophers of action. The first problem (discussed in Chapters II–VI) has been called the problem of action,[10] and its force is epitomized in L. Wittgenstein's famous question:

> ...When 'I raise my arm', my arm goes up. And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?[11]

The central contrast is that between actions and mere happenings: between what the agent does (in an agentively pregnant sense of 'does') and what merely happens to him.

---

[8] Of course, there are attempts to understand even such character traits as ultimately resting on intentional actions, e.g. past intentional actions that have led to certain habits on the part of the agent giving rise to relevant character traits. This view is proposed by Aristotle in *Nicomachean Ethics*. For an interesting dissenting account: Robert Merrihew Adams, "Involuntary Sins," *Philosophical Review* 94 (1985), 3-31.

[9] There are attempts to reinterpret our practice by arguing that we are not responsible for unintentional omissions but rather for actions that led to the unintentional omission. Proponents of such a view include: Holly Smith, "Culpable Ignorance," *Philosophical Review* 92 (1983), 543-571; Michael Zimmerman, "Negligence and Moral Responsibility," *Nous* 20 (1986), 199-218. Among the dissenters: John C. Hall, "Acts and Omissions," *The Philosophical Quarterly* 39 (1989), 399-408; Steven Sverdlik, "Pure Negligence," *American Philosophical Quarterly* 30 (1993), 137-149.

[10] H.G. Frankfurt, "The Problem of Action," *op. cit.*

[11] Ludwig Wittgenstein, *Philosophical Investigations* (New York: Macmillan, 1958), §621.

The traditional answer to this question aims to capture the idea of action as intentional movement. And thus the fundamental Anscombe-Davidson approach is to take an event to be an intentional movement (action) just in case there is a description under which it is intentional.[12] Anscombe clarifies the idea of an intentional action by suggesting that a special sense of the question 'Why?' applies to it, viz. one to which the proper answer appeals to the agent's reasons for doing what he did. While Anscombe herself aims to clarify this account further in particular by distinguishing cases where the question is refused application (e.g. if the agent says "I did not know I was doing that"), others have attempted to clarify the understanding of what an intentional action is by appeal to causal concepts. Some have suggested that one can understand what it is for an action to be intentional (under a description *d*) by appealing to the fact that reasons that rationalize the action (under *d*) have caused the actions. This is the central thought of the causal theory of action.[13] Causal theorists of action aim to ground the distinction between actions and mere happenings by appealing to the idea of reasons causing action. Anscombe, by contrast, offers a non-causal theory of action (she also advances a non-causal teleological theory of action explanation), where the distinction between action and mere happening is ultimately grounded in the ways in which the special sense of the "Why?" question is applied.

The second central problem that will occupy us (partly mentioned in Chapter I and properly addressed in Chapter VII) concerns the force of ordinary action explanations. Ordinary explanations of human action are teleological in nature. We act in order to accomplish goals. In Aristotle's terms, actions have final causes paradigmatically embodied in their reasons. Reasons explain actions by showing what the agent aimed to do. They explain the action by rationalizing it, by showing what it made sense for the agent to do. Since Aristotle's distinction of four types of causes

---

[12] The view has been first proposed by G.E.M. Anscombe in *Intention, op. cit.* It has been taken up by most theorists of action, among them Donald Davidson ("Agency," in *Essays on Actions and Events, op. cit.*, pp. 43-61).

[13] The most explicit recent defense of a causal theory of action (rather than just a causal theory of action explanation) is presented by John Bishop, *Natural Agency. An Essay on the Causal Theory of Action* (Cambridge: Cambridge University Press, 1989). The causal theory of action ought to be distinguished from the causal theory of action explanation, of which Davidson is a proponent, see p. 7, below.

(among them final and efficient causes), and since the modern day scientistic emphasis on efficient causes, the question that notoriously arises is how final and efficient causes are related to one another. This is a problem of the force of teleological explanations in general. The problem has its special application to the domain of human action, and it became the center of discussion in the philosophy of action since Davidson's famous paper "Actions, Reasons, and Causes."[14]

Davidson has claimed that teleological notions themselves are not sufficient to capture the force of ordinary action explanations.[15] That this is so is evident from the fact that we make a distinction between an agent's acting and his action being rationalizable by his reason, and the agent's acting *because of* that reason. The distinction is most vividly drawn in a case where the agent has at least two reasons for performing an action,[16] and acts because of one but not because of the other. For example, someone may have a reason not to go to the movies (not to meet his arch-enemy) but not go because he decided to watch TV instead. To coin some terminology, he acts *on* his desire to watch TV but merely *with* (or in the presence of) the desire not to meet the enemy. Davidson's argument for the causal theory of action explanation, according to which reasons must be construed as also causes of actions, takes the form of a challenge. He claims that only the causal theory of action explanation can account for the distinction between acting for and acting with reasons.

It is worthwhile to emphasize a terminological point. There is no consistent usage of the term 'causal theory of action' in the literature. Because of Davidson's contribution in reviving the use of causal concepts in philosophy of action, it is sometimes supposed that 'causal theory of action' is simply synonymous with 'Davidson's theory about action'. The problem with identifying the term with whatever position Davidson holds is that there are in fact two ways in which the idea of reasons as causes can be employed. If

---

[14] Reprinted in *Essays on Actions and Events, op. cit.*, pp. 3-19.

[15] Davidson does not speak of teleological concepts per se but rather of relations of rationalization. He is concerned with all the resources (except causal ones) that are available to an interpreter of an agent's action. Some teleologists have in fact criticized Davidson by suggesting that teleological notions are stronger than he supposes (see George M. Wilson, *The Intentionality of Human Action* [Stanford: Stanford University Press, 1989]).

the idea that reasons are causes is used to account for the explanatory force of reasons, it is part of a *causal theory of action explanation*. But the idea can be, and has been, used to account for the distinction between actions and mere happenings.[17] In such a case, it forms a foundation of a *causal theory of action*. To add to the terminological confusion, there are good grounds for believing that Davidson espouses only a causal theory of action explanation but not a causal theory of action. In "Freedom to Act"[18] he seems to denounce the feasibility of offering an analysis of action in causal terms by pointing out cases of wayward causal chains, where actions are caused waywardly by reasons, as a standing counterexample to any such attempt. He concludes that the best one can do is to say that actions are caused by reasons "in the right way." And that is hardly illuminating as an account of action. We should not, however, be overly impressed by the terminological turmoil. All it shows is that we sometimes misuse the term 'causal theory of action' when we suggest that Davidson is its author. Davidson espouses a causal theory of action explanation but not a causal theory of action.[19]

3. A Preview

One objective of a theorist of action is to give an answer to the question, What is action? It is rarely appreciated that this task already carries with it an ambiguity. David Velleman has pointed out that we may try to give an answer either to the question what actions really are, or to the question what we ordinarily conceive actions to be.[20] The distinction is not well put, however. It is not that there is a distinction between what a phenomenon captured by the term 'X' really is and what we conceive *X*s to be. This way of putting the distinction either dooms us to cognitive failure or makes the distinction

---

[16] This is a case where the distinction is most vivid but the core of the explanatory relation between a reason and an action applies equally when only one reason is involved.

[17] J. Bishop, *Natural Agency, op. cit.* This view has been suggested by some of Davidson's comments in his early paper "Actions, Reasons, and Causes," in *Essays on Actions and Events, op. cit.*, pp. 3-19.

[18] Reprinted in *Essays on Actions and Events, op. cit.*, pp. 63-81.

[19] In his recent book, *Causality, Interpretation and the Mind* (Oxford: Clarendon Press, 1994), William Child makes the contrary terminological stipulation: he takes a causal theory of action to be a theory about action explanations rather than about the nature of actions in contradistinction to mere happenings. In his terminology, Davidson is an author of the causal theory of action. In what follows, this usage is shunned.

[20] "What Happens When Someone Acts," in (eds.) John Martin Fischer, Mark Ravizza, *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press, 1993), pp. 188-210.

trivial. On the first horn, it gives the appearance that we could never know what *X*s really are (because we would always have access only to our conception of them). On the second horn, the distinction seems trivial because there is a sense in which whether the term 'X' applies to *X*s or to *Y*s is insignificant.

But there is another way of construing the distinction. We partake in certain practices in which we use some concept *X* in various ways. But we may also have formed a theory about the practices and about *X*s. The philosopher may then undertake either of two tasks. He may wish to explain and systematize the theory about *X*s that we have already developed, but make it more sophisticated, cognitively better, more unified, and so on. (This is the task Velleman undertakes.) Alternatively, he may wish to propose a different theory as to the nature of *X*s that, in the first place, is not guided by the theory we have formed but rather treats that theory merely as part of the data for his new theory. (This is the task I undertake below.) It is important to note here that a theorist undertaking the second approach must offer an answer to the question *why* we have developed the particular theory of *X*s that we have developed rather than the theory that he proposes. This criterion of adequacy of the second approach is particularly important if the theory of action proposed by the theorist of action were to differ substantially from the theory we have developed. Otherwise, failing such an explanation of why we have come up with the theory we have come up with, the proponent of the second approach is open to the objection that he has simply changed the topic. At the end of Chapter I, I will in fact suggest an explanation why individualist and intentionalist tendencies have been so prevalent in the philosophy of action. I will advocate nonindividualism in the theory of action explanation (Chapter I and VII) and nonintentionalism in the theory of action (Chapters III-VI).

Corresponding to these two general methodological attitudes, we might distinguish two methodological strategies specific to answering the question of the nature of action. Since on the first approach the purpose is primarily to understand our conception of action, the primary data are the practices of our explaining one another's actions. The reasoning behind such an approach might be reconstructed as follows. The purpose is to understand our concept of action. The best way of doing so is to see how, in ordinary practices, we understand actions. Such an investigation will yield the kinds of

explanatory categories to which we ordinarily appeal, in terms of which we understand actions. The task for the theorist will then be to use these categories in understanding the nature of action. We shall call it the *explanation-based* approach to understanding the nature of action.[21]

One of the main differences between the explanation-based approach and what I shall call the *responsibility-based* approach to understanding the nature of action, is that the former places a much greater faith in the ways in which we conceive of actions.[22] On the latter approach, the purpose is to understand not only how we understand actions but also to understand what we treat as actions in our practices. One of the main indicators of our treating an agent as having performed an action is to hold her responsible for it. On the second approach, the theorist tries to understand our concept of action relying primarily on our practices of ascribing responsibility to one another rather than on our practices of explaining each other's actions.[23]

Chapters II-VI sketch a version of a responsibility-based approach to the understanding of the nature of action. I will employ a traditional responsibility-based strategy for accounting for the difference between actions and mere happenings. Responsibility-based accounts such as H.L.A. Hart's (discussed in Chapter II) as well as contextualist accounts usually define what counts as a mere happening (and so a non-action) in terms of the presence of certain conditions, henceforth referred to as defeating conditions, e.g.: the agent suffering a spasm, being in a coma, being pushed by the wind, moved by another person, and so on. Actions, as a class, are then defined negatively as

---

[21] In general, a variety of intentionalist approaches belong to the category. See e.g., G.E.M. Anscombe, *Intention, op. cit.*; Roderick M. Chisholm, *Person and Object. A Metaphysical Study* (La Salle, IL: Open Court, 1976); D. Davidson, *Essays on Actions and Events, op. cit.*; H.G. Frankfurt, "The Problem of Action," *op. cit.*; C. Ginet, *On Action, op. cit.*; Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970); Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980); John R. Searle, *Intentionality. An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983); J. David Velleman, *Practical Reflection* (Princeton, NJ: Princeton University Press, 1989); G.M. Wilson, *The Intentionality of Human Action, op. cit.*

[22] I am not claiming that the two approaches have to stand in competition. Rather, my claim is that they differ in emphasis. That there need not be a conflict between these two approaches to understanding the nature of action will in fact be evident in that the account of action I will give could be seen as resulting from pursuing both strategies.

those performances that occur in the absence of relevant defeating conditions. The major task that faces a theorist of action following the just outlined route of accounting for the difference between actions and mere happenings, lies in giving an account of the variety of defeating conditions. I develop such an account in Chapters III-VI.

Chapter II discusses H.L.A. Hart's responsibility-based account of action. I assemble objections that have been raised against it and take them to constitute criteria of adequacy for the account to be developed. The major problem concerns the fact that any account of action that would capture not only legal, not only moral, but all actions, must appeal to a notion of responsibility that is appropriately wider than legal or moral responsibility. Chapters III-V clarify such a concept of practical responsibility in terms of reasonable normative expectations. In Chapter VI, I show how an account of practical responsibility developed on these lines helps in giving an account of the distinction between actions and mere happenings. I will argue that an agent's performance is an action just in case there is a description under which it would be reasonable (in a special sense discussed in Chapter V) to expect of the agent that he perform the action.

This is the gist of the answer to the problem of action. One may be concerned, however, that normative expectations at large, with the exception of self-directed expectations perhaps, ought not to enter an account of action in the first place. After all, what kind of connection could another person's expectation of me have with my action? The resistance to this idea reaches very deeply. In the introductory Chapter I, I will try to identify one aspect of what may be seen as troubling. One source of resistance to this notion may derive from an action theorist's adherence to the explanation-based approach to the problem of action. The standard view on the nature of folk-psychological explanations is individualistic: it conceives of actions as being explained by the agent's desires, intentions, beliefs, hopes, etc. (intentional explanations). But this is a simplification of our ordinary practices. We also explain one another's actions in terms of others' requests, commands, wishes, expectations (nonintentional explanations).

---

[23] Ascriptivist and contextualist theories of action exemplify this strategy: H.L.A. Hart, "The Ascription of Responsibility and Rights," *op. cit.*; A.I. Melden, *Free Action* (London: Routledge & Kegan Paul, 1961); R.S. Peters, *The Concept of Motivation* (London: Routledge & Kegan Paul, 1958).

According to explanatory individualism, intentional explanations are privileged over nonintentional explanations: we can only explain an action as done because of another person's desire if the agent acts on some pro-attitude of his own suitably directed toward the other's desire. The bulk of Chapter I is directed toward arguing that there are no conclusive reasons for explanatory individualism. To adopt an explanation-based strategy is thus not to disavow nonindividualism. It does not threaten the use of normative expectations in an account of action.

I pick up the issue of individualism in Chapter VII. There I make the case for explanatory nonindividualism stronger by showing how we can be thought to act because of other people's expectations of us (without thereby acting on our own expectations). In so doing, I respond to Davidson's challenge and show how to account for the distinction between acting for and acting with reasons. An action can further one end (satisfy one reason) as well as another end (satisfy another reason), and yet be done for one reason rather than another. Teleological relations are not sufficient to render the distinction. So, Davidson suggests that we must appeal to the idea that reasons are causes in order to understand the distinction. I show how we can account for the distinction without supposing that reasons are causes, but rather by thinking of reasons as selectional criteria.

CHAPTER I.

IS EXPLANATORY INDIVIDUALISM CONCEPTUALLY NECESSARY?

In Chapters III-VI, I will argue that we can understand the distinction between actions and mere happenings by appealing to the concept of normative expectation rather than to the concept of intention. One source of resistance to this notion may derive from an action theorist's adherence to the explanation-based approach to the problem of action. Such theorists typically start with a theory of action explanations and then build a theory of action in terms of the elements singled out in the theory of action explanation. Philosophical discussions centering around action explanations typically mention as explanatory elements intentional attitudes of the agent not expectations to which the agent is held by other people. It may accordingly appear as if the conceptual distance of the concept of normative expectation from the concept of action is too great for the former to be used in the theory of action.

In section 1, I identify the positions of explanatory individualism, nonindividualism and anti-individualism about action explanations. Roughly, according to explanatory individualism actions are explained in terms of the agent's pro-attitudes; according to explanatory anti-individualism, actions are explained in terms of other people's pro-attitudes toward the agent; according to explanatory nonindividualism, some actions are explained in terms of the agent's pro-attitudes, others in terms of other people's pro-attitudes toward the agent (without being explained in terms of the agent's own pro-attitudes). In section 1.B, I distinguish between explanatory and normative individualism, each coming in a reductive and a nonreductive version. I offer two preliminary considerations in support of a nonindividualist position: the testimony of our practices (section 1.A) and an evolutionary consideration (section 2). The bulk of the

chapter (section 3) is devoted to showing that a variety of arguments for explanatory individualism fail. I will conclude that explanatory nonindividualism is not incoherent. In the final section 4, I shall endorse normative individualism. I suggest that much of the appeal of explanatory individualism derives (though it ought not to) from the appeal of normative individualism.

1. Individualism vs. Nonindividualism about Action Explanations

"He went to the store because his mother wanted him to bring her some butter for the cake she is baking." "I went to the library because my friend is in the hospital and she asked me to get her some good book to read." "I just told him I forbid him to come near my house, and he stopped bothering us." — These are just some examples of ordinary explanations of actions. What they all have in common is that they relate, in one way or another, how one person's wish, desire, expectation (pro-attitude) influences another person's action. Indeed, the thought that we can *sometimes* affect what others do by wanting, asking or telling them to act in certain ways is rather common-sense. We do this all the time. This is reflected in our practice of ordinary action explanations — we do allow explanations of one person's actions in terms of the pro-attitudes of another person.

This fact is a little jarring if one looks through the extensive literature on action explanation, folk psychology, and our ordinary practices for attributing mental states. With literally a few exceptions,[1] only intentional explanations, i.e. explanations that appeal to the *agent's own* pro-attitudes, are discussed. Some authors speak as if our ordinary action explanations are intentional explanations; others discuss only intentional explanations. This suggests that nonintentional explanations (explanations that appeal to other people's pro-attitudes) tend not to be considered as being on a par with intentional explanations, and that intentional explanations tend to be privileged in one way or

---

[1] Annette C. Baier, "Rhyme and Reason: Reflections on Davidson's Version of Having Reasons," in (eds.) Ernest LePore, Brian P. McLaughlin, *Actions and Events* (Oxford: Basil Blackwell, 1985), pp. 116-129; *Postures of the Mind. Essays on Mind and Morals* (Minneapolis: University of Minnesota Press, 1985); Leszek Nowak, *Power and Civil Society. Toward a Dynamic Theory of Real Socialism* (New York: Greenwood Press, 1991). Georg Henrik von Wright, "Explanation and Understanding of Action," in *Practical Reason* (Ithaca: Cornell University Press, 1983), pp. 53-66.

another. Below, I will distinguish some ways in which intentional explanations can be thought to be privileged over nonintentional explanations (section B). In general, the position according to which the action explanations that appeal to the agent's own pro-attitudes are privileged over explanations that appeal to other people's pro-attitudes will be termed individualism about action explanation. Correspondingly, nonindividualism about action explanation will allow the appeal to the pro-attitudes of people other than the agent himself to have, in certain circumstances, import similar to the appeal to the agent's own pro-attitudes.[2]

---

[2] There are at least three debates that one ought not to confuse with the one intended here. First, there is a debate between individualists and anti-individualists in the philosophy of language and mind, where the question is whether our concepts can be individuated solely in terms of the states of the individual person who possesses the concept (the individualist position) or in terms that reach beyond the states of the person into the surrounding world (the nonindividualist position). See e.g.: Daniel C. Dennett, *The Intentional Stance* (Cambridge, MA.: Bradford Books, 1987); John R. Searle, *Intentionality. An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983); Tyler Burge, "Individualism and the Mental," in (eds.) Peter A. French, Theodore E. Uehling, Jr., Howard K. Wettstein, *Studies in Metaphysics* (Minneapolis: University of Minnesota Press, 1979), pp. 73-122; Hilary Putnam, "The Meaning of Meaning," in (ed.) Hilary Putnam, *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975), pp. 215-271.

Second, there is a debate in the philosophy of action concerning the question whether only individual persons can be properly thought to be agents (individualism about the unit of agency) or whether certain kinds of collectives (e.g. firms, states, groups, institutions) can also be thought to act (collectivism about the unit of agency). Although this debate is orthogonal to the opposition between individualism and nonindividualism, as I will understand it, the dissertation has as its consequence the denial of individualism about the unit of agency. But this is a welcome consequence. A prominent reason for asserting that only individuals can act is if one shapes one's concept of action on the model of a bodily movement caused by a pro-attitude, which pro-attitude is in turn understood as a physiological state of the agent's body. The understanding of action advanced here departs from any such model, and we shall see that there remains no impetus for denying the natural view that groups, families, states, schools can all act.

Third, another question that one may ponder is whether it is possible for a human agent (in the proper sense of the term) to exist without society. Atomists (sometimes called individualists) hold that there is nothing incoherent in the supposition of a solitary agent; holists (sometimes called nonindividualists) hold that our relations with others are constitutive of our nature as agents.

The final and closest question concerns the extent to which the existence of social regularities compromises our picture of ourselves as intentional agents. Individualists deny, while collectivists affirm, that social regularities challenge intentional psychology. This debate (and in particular its distinction from the atomism vs. holism debate) has been put into sharper focus in Philip Pettit's *The Common Mind. An Essay on Psychology, Society, and Politics* (Oxford: Oxford University Press, 1986). It is closest to our concerns although nonindividualism, as it is understood here, is not exclusively tied to the thought that social regularities compromise the individualistic picture of ourselves. It does, however, challenge that picture. (I discuss some connections between collectivism as Pettit understands it and nonindividualism, as it is here understood, in "Collectivism on the Horizon: A Challenge to Pettit's Critique of Collectivism," forthcoming in the *Australasian Journal of Philosophy*.)

A. A Variety of Folk-Psychological Action Explanations

Although it might appear a little pedantic to make this point, it is clear that there is more to our ordinary explanations than explanations in terms of intentions, beliefs, pro-attitudes, hopes, etc. Without pretending to offer an exclusive or exhaustive list of types of ordinary explanations of actions, there are at least six kinds of explanations we offer. (1) We give explanations in terms of the agent's pro-attitudes (intentions, wishes, convictions, pro-attitudes, hopes, etc.), but (2) we also offer explanations that cite a feature of the agent but not a pro-attitude (explanations in terms of the agent's character or personality or habits). Moreover, (3) we give explanations that invoke pro-attitudes but not those of the agent (others' pro-attitudes, wishes, commands, requests, expectations of the agent). (4) We also offer explanations that neither invoke pro-attitudes nor cite a feature of the agent but rather cite a situation in which the agent finds herself, the social role she plays, the customs or norms that bind her. (5) We frequently invoke explanations in terms of goals, aims, aspirations, as well as (6) explanations in terms of facts, such as that it is raining, or that the school year has begun. And so on.

No one doubts that the first class of explanations constitutes a part of our folk-psychology. Also explanations in terms of character or personality traits are becoming more of a part of the picture of our ordinary explanations. This is partly due to the revival of virtue ethics and partly due to the fact that those explanations have been at the forefront of explanations of actions offered by social psychologists.[3] Also explanations in terms of goals and facts (such as those cited above) have been thought to be part of the explanatory enterprise, though many authors took them to be enthymematic forms of intentional explanations.

The matter presents itself differently for explanations of the third and fourth kind. The main source of resistance to acknowledging them as genuine explanations of actions

---

[3] Barbara von Eckhardt has recently argued that the philosophical conception of folk psychology seems to be completely blind to explanations in terms of personality traits: "The Empirical Naiveté of the Current Philosophical Conception of Folk Psychology," *delivered at the Central Division of the American Philosophical Association and the Third Meeting of the Pittsburgh-Konstanz Colloquium in the Philosophy of Science* (1995). It is worth mentioning here that Donald Davidson includes not only pro-attitudes but also something like character traits in the category of pro-attitudes (see the quote in footnote 4).

comes from the adherence to causalism, i.e. to the view that the explanatory elements mentioned in the explanations of action must be causally related to the actions themselves. On this picture, the mention of another person's pro-attitude or of the social situation in which the agent finds himself seems hopelessly distant from the actual causal chain that generated the event we count as action. The explanation of action must proceed via some attitude of the agent.[4] (I will explore this point later.) By contrast, on the alternative (teleological) construal of explanations according to which action explanations do not explain by appealing to causes but to ends or goals of the agent, the third and fourth kinds of explanations seem prima facie less suspect.[5]

G.H. von Wright is among the few who clearly acknowledge those kinds of explanations.[6] He distinguishes two broad classes of reasons: internal (inner) and external (outer). He characterizes inner reasons as ones that are necessarily reasons for action: "no-one who is familiar with action discourse could, without committing an inconsistency, deny that aiming at something and thinking a certain action promotive of this aim is *a* reason for doing it."[7] Aside from actions done on inner reasons, there are ones done in response to "orders, requests, questions, ... [and] other signals" as well as to "(prescriptive) rules or norms, and ... customs, fashions, or traditions within a community."[8] These varied circumstances von Wright jointly calls outer reasons, which in contrast to inner reasons are related to action only contingently. This is to say that

---

[4] Donald Davidson gives the following characterization of the inclusive category of pro-attitudes which includes: "pro-attitudes, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values *in so far as these can be interpreted as attitudes of an agent* directed toward actions of a certain kind. The word 'attitude' does yeoman service here, for it must cover not only permanent character traits that show themselves in a lifetime of behavior, like love of children or a taste for loud company but also the most passing fancy that prompts a unique action, like a sudden pro-attitude to touch a woman's elbow." ("Actions, Reasons, and Causes," in *Essays on Actions and Events* [Oxford: Clarendon Press, 1980], p. 4, emphasis added.)

[5] G.M. Wilson makes the point that this is also a reason why the teleological conception of explanations is a more natural rendition of our ordinary action explanations than the causal one (*The Intentionality of Human Action* [Stanford: Stanford University Press, 1989].) Also, many of G.E.M. Anscombe's examples have this form (*Intention* [Ithaca: Cornell University Press, 1957].). Not to mention the fact that among Wittgenstein's favorite examples is that of ordering or commanding (*Philosophical Investigations* [New York: Macmillan, 1958]).

[6] G.H. von Wright, "Explanation and Understanding of Action," *op. cit.*

[7] Ibid., p. 54, original emphasis.

"even though the agent recognizes the challenge [a certain situation such as a request] and has learnt or otherwise knows how to respond to it, he need not *acknowledge* it as a reason *for him* to act upon."[9] It should be stressed, however, that although von Wright thinks that the connection between outer reasons and actions is less tight than the connection between inner reasons and action, he nonetheless thinks that the agent can be said to act on outer reasons thus understood.

> In the individual case, it may ... be difficult or even impossible to tell whether the agent obeyed an order *because* he had been ordered, or *because* he feared punishment for disobedience. His motives might have been "mixed." But it would be a distortion to think that his action *must* have had internal reasons and *could not* have taken place on purely external grounds.[10]

Whether von Wright's position is justified as a position in the theory of action explanation we will have to explore further. For now, the point is, however, that (insofar as our ordinary ways of thinking about actions can be seen as embedded in the ways in which we explain actions) it is part and parcel of the way we ordinarily *think* about actions that it is possible for us to act not only on our own intentions, desires, or wishes, but also on others' expectations, desires, or wishes, on others' requests or commands, or on norms explicit or implicit in our social lives. Of course, there might be good reasons for thinking that this liberal and literal attitude toward our practices of explaining actions is too liberal and literal. In particular, one might argue that our practices are subject to certain pragmatic pressures and that is why we offer explanations that we do not really mean to be offering. I shall discuss various reasons for holding such a view later (section 3). At present, I merely want to register the *possibility* that the fact that we do explain our actions in these diverse ways is in fact integral to our understanding of action explanation.

Before proceeding, I should point out that although we shall see in Chapter VII how to account for at least some of the plethora of ordinary action explanations, for present purposes, two kinds will be singled out: intentional explanations, i.e. explanations

---

[8] Ibid., p. 54.
[9] Ibid., p. 54.
[10] Ibid., p. 55, original emphasis.

that appeal to the agent's pro-attitudes, and what I will call nonintentional explanations, i.e. explanations that appeal to another person's pro-attitudes.[11]

B. Normative and Explanatory Individualism

So far, I have characterized individualism as a position on which intentional explanations are privileged over nonintentional explanations, and nonindividualism as a position that rejects individualism. The characterization is wanting, however. I will accordingly distinguish four different individualist positions and two respective nonindividualist positions. Before doing so, let me say a few words about the concept of a pro-attitude.

Davidson gives a very inclusive characterization of the category of pro-attitudes. Included are: "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind."[12] This inclusive category can be illuminated by appeal to Anscombe's notion of direction of fit.[13] Anscombe distinguishes two kinds of mental attitudes. There is a group of mental attitudes such that if the world does not accord with them, they are at fault (beliefs belong to this category). And there is a group of mental attitudes such that if the world does not accord with them, the world is at fault (intentions, desires belong to this category). Beliefs (that *p*) are in general attitudes such that if it is the case that not-p, the belief has to be abandoned. They have a mind-to-world direction of fit.[14] Intentions, desires, pro-attitudes in general (that *p*) are attitudes such that if it is the case that not-p, they dispose the agent toward making it the case that *p*. They have a world-to-mind direction of fit. I should emphasize that the category of pro-attitudes

---

[11] I should stress that the use of the term 'nonintentional explanation' is dictated by terminological convenience. Strictly speaking, of course, all the explanations that are not intentional explanations could be thought to be nonintentional. However, since our discussion until the end of Chapter VII will focus on just one kind of nonintentional explanations, viz. those that appeal to the pro-attitudes of persons other than the agent, I shall reserve the term for those kinds of explanations.

[12] "Actions, Reasons, and Causes," *op. cit.*, p. 4.

[13] This thought underlies Michael Smith's theory of desire ("The Humean Theory of Motivation," *Mind* 96, 1987, 36-61). I consider the implications of Smith's account for nonindividualism in section 3.C, below.

understood as those attitudes having a world-to-mind direction of fit will include not only desires (the paradigmatic motivating attitudes of the Humeans) but also noninstrumental beliefs (the paradigmatic motivating attitudes of the Kantians).[15]

Four individualist positions can be obtained by crossing two criteria. First, the individualist thesis for privileging the agent's pro-attitudes in relation to that agent's action may concern either the rationalizing or the stronger explanatory relation between reasons and actions. Second, the individualist positions may differ with respect to the force that they attach to the thesis. The position may be merely committed to the thesis that in all cases, the intentional explanations are privileged in one way or another. But they may also hold the stronger thesis that the privileging holds at the expense of any nonindividualist positions. Let us consider them in turn.

On the weakest position of (inclusive or non-reductive[16]) *normative individualism* (which will be discussed in greater detail in section 4, where the reason for giving it the name will be clearer), for any action that we intuitively explain nonintentionally, it is possible to attribute to the agent some pro-attitude toward the action:

(NI) For any action, the agent α has some pro-attitude toward the action.

In other words, normative individualism is a position according to which every action can be rationalized in terms of the agent's pro-attitudes. To say that it is possible to attribute a pro-attitude toward the action is to say that the action is justified by the pro-attitude.

---

[14] Note that there is a certain ambiguity in how to read the phrase 'mind-to-world'. It is customary to read it "mind-(ought)-to-(fit)-world" rather than, as might be tempting, "(from)-mind-to-world."

[15] The category of "moral views" also figures in Davidson's list (see the quote above). Smith's paper in which he originally proposed this broad understanding of desire was meant as a defense of a Humean theory of motivation. It has been argued against Smith, however, that his argument against Kantians is unsuccessful. For there is nothing on Smith's account to prohibit certain attitudes (noninstrumental beliefs) from having a dual direction of fit: world-to-mind (characteristic of motivational attitudes) and mind-to-world (characteristic of cognitive attitudes). For more on this see e.g., I.L. Humberstone, "Direction of Fit," *Mind* 101 (1992), 59-83; Philip Pettit, "Humeans, Anti-Humeans, and Motivation," *Mind* 96 (1987), 530-533; G.F. Schueler, "Pro-Attitudes and Direction of Fit," *Mind* 100 (1991), 277-281; Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994). For us the important point is that the issue between the individualists and the nonindividualists is orthogonal to the issue between Kantians and Humeans.

[16] I shall adopt the convention of speaking of reductive normative (respectively, explanatory) individualism but omitting the adjective 'inclusive' or 'non-reductive' when speaking of (inclusive or non-reductive) normative (respectively, explanatory individualism) unless to emphasize the point made.

that the agent *may* have performed the action because of that pro-attitude. This is not to say, however, that the agent actually acted *because* of that attitude.

The force of the 'because' in question is the one made famous by D. Davidson. We may rationalize an action by giving the reasons that the agent had for acting in that manner but this is not yet tantamount to our explaining his action. For to say that a reason explains the agent's action is to say not only that the agent had the reason while acting but that the reason was efficacious in bringing the action about, that the agent acted *for* that reason. In these terms, normative individualism is committed merely to the thesis that for any action, it is possible to list reasons the agent had for acting, but it is noncommittal with respect to a stronger thesis, that for any action, some reason of the agent explains the action.

A stronger position would exclude any pro-attitudes of other people from rationalizing the agent's action. The *reductive normative individualist* not only holds (NI) but also:

(rNI) Only α's pro-attitudes can rationalize α's actions (can be reasons for α to act).

Reductive normative individualist will thus oppose the thought that my friend's wanting me to go to a concert with her can possibly rationalize my going there with her. Only *my* wanting to go to the concert or *my* wanting to oblige my friend could rationalize my action. A non-reductive normative individualist will see no problem in allowing my friend's wishes to rationalize my action as long as there are pro-attitudes of mine that rationalize my action as well. In either the non-reductive or the reductive flavor, normative individualism concerns only the rationalization relation that holds between the agent's pro-attitudes and his action. The stronger explanatory relation is of concern to explanatory individualism.

The weaker (non-reductive) *explanatory individualism* holds not only that every action is rationalizable in terms of some pro-attitude of the agent but that every action can be *explained* in terms of some pro-attitude of the agent.

(EI) All α's actions are explained by some pro-attitude of α.

According (EI), when an agent acts, not only can we always attribute some belief and pro-attitude that will rationalize the action, but, in addition, the agent acts *because* of some of his beliefs and pro-attitudes. A position that is stronger still is *reductive explanatory individualism.* Reductive explanatory individualism not only endorses the thesis that any action is explained by some pro-attitude of the agent but it puts forward the stronger thesis that no action of one agent can be explained by a pro-attitude of another person:

(rEI) Only α's pro-attitudes can explain α's actions.[17]

Explanatory individualists oppose the thought that my friend's wanting me to go to the concert caused me to go to the concert with her unless some of my desires (at the very least to oblige my friend) was an intermediary. They allow another person's pro-attitude to explain the agent's action provided it is *mediated* by some pro-attitude of the agent. Reductive explanatory individualism, on the other hand, rejects the thought that somebody else's pro-attitude can be explanatorily relevant to the agent's action.

Given these four individualistic positions, all of which in some respect privilege intentional explanations, we ought to ask which of the positions the nonindividualist must reject. There is no question that any nonindividualist must reject the reductive theses (rNI) and (rEI). An *explanatory nonindividualist* might reject in addition (EI) without rejecting (NI). Such a nonindividualist will hold that while some of our actions are explained by our own pro-attitudes, there are also actions that are explained by others' pro-attitudes without being mediated by the pro-attitudes of the agent. Such a nonindividualist would not deny, however, that the actions that are properly explained nonintentionally ("caused" by others' pro-attitudes) can still be rationalized in terms of the agent's pro-attitudes. By contrast, a *normative nonindividualist* might adopt the

---

[17] We may summarize the relation between the various theses as follows:
(NI) $(\forall x)\,[Ax \supset (\exists y)\,(Ryx \,\&\, Py)]$ (rNI) $(\forall x)\,[Ax \supset (\forall y)\,(Ryx \supset Py)]$
(EI) $(\forall x)\,[Ax \supset (\exists y)\,(Eyx \,\&\, Py)]$ (rEI) $(\forall x)\,[Ax \supset (\forall y)\,(Eyx \supset Py)]$
where 'Ax' stands for 'x is α's action', 'Py' — 'y is α's pro-attitude, 'Rxy' — 'y rationalizes x', 'Eyx' — 'y explains x', and (rNI) is the thesis that reductive normative individualism holds in addition to holding (NI), likewise for explanatory individualism.

position that there are actions that are not rationalizable in terms of the agent's pro-attitudes but only in terms of the pro-attitudes of others.[18]

We should reflect on the fact that the nonindividualist positions have been defined relative to individualist positions. The reason for this is that nonindividualism in general constitutes a more faithful representation of common sense. The spirit of *nonindividualism* is *inclusive*: it allows that we act because of our own pro-attitudes as well as because of those of other people. What it opposes is, accordingly, the *exclusive* positions of both *individualism*, and a parallel position of *anti-individualism*.

There are few, if any, anti-individualists, but it may help to appreciate the pluralistic and tolerant spirit of nonindividualism to but briefly characterize parallel anti-individualist positions. (NA) *Normative anti-individualism* would be a position according to which we can rationalize all of an agent's actions in terms of others' pro-attitudes. (rNA) *Reductive anti-individualism* would hold that we can only rationalize an agent's actions in terms of others' pro-attitudes but not the agent's own. (EA) On *explanatory anti-individualism*, all of an agent's actions could be explained by others' pro-attitudes. And finally (rEA), the thesis that the agent's actions could only be explained by others' pro-attitudes is characteristic of the *reductive* version of *explanatory anti-individualism*.

Chapter VII will allow us to understand how it is possible to act on another person's pro-attitude without acting on any of the agent's pro-attitudes (the position of explanatory nonindividualism). In this preliminary chapter, however, my concern is solely to argue that the arguments for explanatory individualism are not conclusive, and so to argue against the supposition that explanatory nonindividualism is incoherent (section 3). In section 4, I shall discuss some reasons that might be responsible for the fervent adherence to individualism, suggesting that they support normative individualism at the very best, not explanatory individualism.

---

[18] One might think that the candidates for such actions are cases that Allan Gibbard has baptized 'social akrasia' (*Wise Choices, Apt Feelings. A Theory of Normative Judgment* [Cambridge: Harvard University Press, 1990]). An excellent example is given by Stanley Milgram's experiments on obedience, where (*Obedience to Authority* [New York: Harper & Row, 1969]) the experimental subjects follow the commands of the experimenter *against* their better judgments.

It may pay to be reminded of Dennett's distinction between two levels at which folk-psychological concepts are used: subpersonal and personal.[19] The dispute between the nonindividualist and the individualist concerns the appropriate way of reconstructing our folk psychology at the personal level. It is thus perfectly appropriate for a nonindividualist to claim that explanations in terms of others' pro-attitudes do not require the involvement of the agent's pro-attitude (if the pro-attitude-talk is understood at the personal level) and yet allow the subpersonal investigations of cognitive psychology to postulate pro-attitude-like states on the part of the agent.

One final point demands emphasis. The individualist, anti-individualist and nonindividualist positions are all characterized in terms of the relation between the agent's action and the agent's and other people's *pro-attitudes*. For an explanatory nonindividualist to allow for the possibility that an agent acts on another person's pro-attitude without the mediation of his own pro-attitudes is not to deny that the agent's beliefs may be involved in his acting. In fact, we will see that a certain kind of beliefs will be relevant to the selectional model of acting for a reason (Chapter VII).

2. Individualism, Nonindividualism and Evolution

Before venturing any further, it might be worthwhile to throw the nonindividualist thought we are considering against the background of our evolutionary development. It has been argued that the evolution of human beings favored and selected rational behavior.[20] Humans who were able to conduct themselves in rational ways were better off than those who were not. Since rational conduct consists (in part at least) in satisfying one's desires[21] to the best of one's knowledge, this gives one reason to believe that humans will, in normal conditions, be rather good at satisfying their desires. While this line of thought is perfectly reasonable (and nothing I say serves to undermine it — except for altering its status), theorists have also come to recognize the evolutionary

---

[19] Daniel C. Dennett, "Three Kinds of Intentional Psychology," in *The Intentional Stance*, *op.cit.*, pp. 43-82.

[20] See e.g. Daniel C. Dennett, "Intentional Systems," in *Brainstorms* (Cambridge, MA: Bradford Books, 1981), pp. 3-22, and "True Believers: The Intentional Strategy and Why It Works," in *The Intentional Stance*, *op. cit.*, pp. 13-35.

[21] I use 'desire' very broadly. It should be taken to be synonymous with 'pro-attitude'.

advantage of another kind of conduct — conformism. It has been argued that there is a distinct evolutionary benefit for us in conforming.[22] It is reasonable to assume that the patterns of behavior adopted by a particular group of people have been tested out in the particular kinds of situations and environment in which the group tends to find itself. It may be beneficial for an individual joining such a group to use the tested out patterns of behavior (thus adopting the wisdom of the past) instead of risking that the behavioral pattern of his invention will be selected out. This is the selectional advantage of conformism — of our acting not on our own minds but rather on other people's minds.[23]

These two parts of the evolutionary story are in no way exclusionary. They simply illustrate the presence of forces supporting, on the one hand, the development of a tendency for us to be independent, acting on our own convictions, and on the other hand, the development of the converse tendency for us to depend on others. Insofar as both forces have been operational, we would expect our lives to be an arena for a struggle between these two tendencies in certain situations. And this thought has a true phenomenological ring to it. The nonindividualist idea that we sometimes act on our own pro-attitudes and sometimes on others' pro-attitudes simply reflects this evolutionary heritage. And just as the 'individualist' part of the evolutionary story (taken on its own) would support the individualist's commitment to the thought that in normal conditions we act on our own beliefs and desires, so the whole story should support the nonindividualist thought that we ought to extend our understanding of what happens in normal conditions to encompass not only our acting on our own desires but also our acting on others' desires.

The two-pronged nature of the evolutionary account also suggests adopting a suspicious attitude toward the reductionist strategy of the individualist. In insisting that all actions done on others' pro-attitudes (in normal conditions) are reducible to actions done on the agent's pro-attitudes, the individualist in effect gives priority to one of the prongs in the evolutionary story. This would be understandable if there were a

---

[22] Robert Boyd, Peter Richerson, *Culture and the Evolutionary Process* (Chicago: Chicago University Press, 1985).

[23] Note that this is not tantamount to saying that there are evolutionary grounds for our acting *against* our own minds (the impression to the contrary may be dictated by the ambiguity discussed in section 3.B).

conceptual competition between the two evolutionary tendencies. If they were incompatible with one another, that would give a reason for being an *individualist* (and preferring the individualist prong, thus conceiving of all actions as done on the agent's own pro-attitudes, in normal conditions) or for being an *anti-individualist* (and preferring the nonindividualist prong, conceiving of all actions as done on others' pro-attitudes, in normal conditions). But there is no conceptual competition between the two parts of the evolutionary account. The only competition there is (if there is any at all) concerns the question which of the forces takes precedence in the agent's action in particular circumstances. But if so, then we find no evolutionary reason to suspect that the reductionist strategy should be the one to hold the most promising — whether in its individualist or anti-individualist form. In fact, we find every reason to believe that the two parts of the evolutionary story will be reflected in the way in which we are 'designed'. This supports[24] the *nonindividualist* (in contrast to the anti-individualist) thought that we sometimes act on our own minds and sometimes on those of other people.

3. Arguments for Explanatory Individualism

Explanatory individualism is incompatible with explanatory nonindividualism. According to explanatory nonindividualism, it is possible for some actions of the agent to be explained in terms of somebody else's pro-attitude without being mediated by any pro-attitudes of the agent. Explanatory individualists deny this possibility. My sole aim in this section will be to disarm arguments that might favor explanatory individualism.

It should be emphasized that the goal is to cast doubt on the conclusiveness of such arguments, but not to suggest that there are no reasons for adopting explanatory individualism. Rather, I claim that the reasons for developing an explanatory individualist

---

[24] This is not a definitive argument for nonindividualism and against individualism. It is also not the kind of consideration that someone who is already convinced of the truth of individualism is going to find appealing in any way. Someone like that will embrace everything that is said here and simply reinterpret the idea of acting on another person's mind in individualist terms (acting on one's own mind which is open to what another person might want, for example). This consideration might be appealing, however, to someone who suspended his commitment to either position and declared himself open to considering the

position are not strong enough to render explanatory nonindividualism incoherent. I will argue that it is not necessary, not that it is not possible, to believe (EI) to be true. It would be erroneous to give the impression that the assembled arguments exhaust the reasons for (EI). But prima facie they give rather powerful support to the position.

In section A, we will see how, contrary to appearances, the natural conception of pro-attitudes as internal states of the agent does not threaten the nonindividualist interpretation of folk psychology. In section B and section C, I consider two arguments designed to show that it would be incoherent to think that an agent may act on another person's pro-attitude without acting on a pro-attitude of her own. We will see that neither the argument from breakdown cases (section B) nor M. Smith's argument (section C) establishes that conclusion. In either case, there is conceptual room for the nonindividualist position.

A. Internal States and Individual Action

Perhaps it is best to begin by dissipating a worry that may be responsible for a certain incredulity with which a nonindividualist understanding of folk psychology might be met. It is customary to construe pro-attitudes as internal states of an agent.[25] It is also customary to construe actions as events that are caused by the agent's, among others, internal states. But if so, then it might seem that whatever other person's pro-attitudes may be relevant to the agent's performing the action, the agent's pro-attitudes are necessarily involved, for they cause the very event in question. To deny the involvement of the agent's pro-attitudes is to deny the involvement of the agent's internal states, and this is unintelligible.

The argument begs the question against the nonindividualist in an important way. Just as it is customary to construe pro-attitudes as internal states and actions as events

---

result of such an evolutionary argument as at least a reason (albeit not a decisive one) for adopting a position on the matter.

[25] There are important exceptions, among them: Lynne Rudder Baker, *Explaining Attitudes. A Practical Approach to the Mind* (Cambridge: Cambridge University Press, 1995); D.C. Dennett, *Brainstorms, op. cit.*; D.C. Dennett, *The Intentional Stance, op. cit.*; Jennifer Hornsby, "Which Physical Events are Mental Events," *Proceedings of the Aristotelian Society* 81 (1980-1), 73-92. The point of the argument survives

caused *inter alia* by the agent's internal states, so it is customary to understand our attributions of pro-attitudes as part of a holistic attempt to understand an agent's behavior.[26] According to this last position, any particular internal state of a person counts as the agent's pro-attitude that *p*, for example, only insofar as an attribution of a pro-attitude that *p* would maximize our understanding of the person's behavior. Such an attribution is regulated by our adherence to certain claims about human behavior, in particular, the claim that people act on their beliefs and pro-attitudes. In other words, the identification of our pro-attitudes presupposes a certain understanding of our folk psychology.

At this point, the nonindividualist must claim that the nonindividualist understanding of folk psychological explanations will affect the very identification of our pro-attitudes. If we allow at the outset that aside from acting on their own pro-attitudes, people also act on others' pro-attitudes, then we might seek the maximization of our understanding of a person's action not by attributing a pro-attitude to that person but rather by attributing a pro-attitude to another person. Think of a scenario when one person exhibits a certain behavioral pattern only in the presence of a certain person. While, of course, defeasible, this would count as a prima facie evidence that the person does what she does because of the involvement of the other person.[27]

But if this is so, then a nonindividualist can also uphold all three customary positions. He may hold the customary view that pro-attitudes are internal states. He may hold that actions as events are caused *inter alia* by the agent's internal states. And he

---

even if one does not identify pro-attitudes with internal states, as long as pro-attitudes are conceived to be causally efficacious states of the agent.

[26] D. Davidson, *Essays on Actions and Events, op. cit.* D.C. Dennett, *Brainstorms, op. cit.* and *The Intentional Stance, op. cit.*

[27] That we do as a matter of fact exhibit the tendency to interpret actions in such terms is a matter of common sense. Those a little more skeptical will benefit from a reminder of what Allan Gibbard has called the phenomenon of social akrasia, the paradigmatic example of which is Milgram's experiments (see *Wise Choices, Apt Feelings, op. cit.*). Such cases appear to be most naturally explained as the agents acting on the experimenter's wishes, commands or requests. In these cases, the individualist (who holds that we act on our beliefs and pro-attitudes) experiences the same sort of conceptual discomfort he experiences in cases of akrasia. For just as in cases of akrasia, we seem forced to interpret the action in terms of the agent succumbing to a temptation, acting on a weaker pro-attitude, so in the cases of social akrasia, we seem forced to interpret the action in terms of the agent succumbing to another, acting on someone else's pro-attitude against her own.

may hold that pro-attitudes are attributed as part of a holistic attempt to understand an agent's behavior. The fact that actions are caused by the agent's internal states does not mean that actions must be caused by the agent's pro-attitudes since not all internal states of the agent are the agent's pro-attitudes. Only those states of the agent that we would have holistic reasons to understand as pro-attitudes are pro-attitudes. And by accepting the nonindividualist reconstruction of folk psychology, a conceptual space opens for not understanding all performances of an agent in terms of that agent's pro-attitudes.

B. The Argument from Breakdown Cases

One way of supporting the individualist would be to show that for any nonintentional explanation of action (i.e. an explanation that does not mention the agent's pro-attitudes) there must be an intentional explanation of the action (mentioning some pro-attitude of the agent). The argument from breakdown cases purports to do just that.

The line of thought is quite simple. It becomes evident that for any nonintentional explanation (citing another person's pro-attitude) there exists (even if it is not explicitly mentioned) an intentional explanation of the action, when we imagine an appropriate counterfactual situation. Let some nonintentional explanation why an agent performed an action be given. Imagine now what would happen were that agent *not* inclined (in one way or another) to perform the action in question. It seems clear that *ceteris paribus* had she not wanted to perform the action (under some description), she would not have. But since she did perform the action she must have wanted to perform it (under some description).

Consider an example. Let us suppose that someone asks you for directions to Sydney. You give him the directions. Why did you give the directions? Because he asked for them. We understand your behavior by appealing not to your pro-attitude to give directions to the person but rather by appealing to that person's having asked you for directions. But, the objector continues, the fact that this explanation is natural (if not obvious) does not yet show that there is no intentional explanation accompanying it. And she wants to suggest that in fact there *must* be an accompanying intentional explanation. This is because *had* you *not* wanted to give the person directions you *would not* have. So, since you did give the directions, you must have wanted to after all.

But what makes us think that you would not give directions if you did not want to? Well, you might have thought the driver looked suspicious and you did not even want to come near the car. You might have been upset by the daily events, or someone just running into your groceries, and did not want to help any member of the human race. Many events like this, or even spur-of-the-moment viciousness might have made you not want to give him directions and not give the directions *even though* you were asked.

We should, however, reflect on the fact that we easily tend to skip over a scope ambiguity with respect to negation.[28] It is one thing to *want not* to do something (in the sense of having a con-attitude toward it), it is another *not* to *want* to do something (in the sense of lacking a pro-attitude toward it, possibly being neutral with respect to it). This difference is very easy to overlook. Consider the announcement: "I have no intention of complying with the court's order." The claim is certainly not that suggested by the surface grammar — the speaker is not expressing a lack of an intention. To the contrary, she is announcing an *intention not* to comply. Or when a child says "I don't want to play with him," she is not expressing a lack of attitude.

Bearing this distinction in mind, it is clear that in order to argue that a pro-attitude is a necessary part of any action explanation, the objector has to show that had the pro-attitude been missing (rather than had the con-attitude been present) the agent would not have done as he did (*ceteris paribus*). But if we look again at the sorts of examples that made us think that you would not give directions to the stranger if you "did not want" to do so, we will discover that they are ones where you *wanted* to *avoid* doing so, where you *wanted not* to do so. In neither of these hypothetical cases do you lack a want to give directions to the driver. You do not merely lack a pro-attitude when you think the driver suspicious and "do not want" to come near the car — you actually have a con-attitude: you *want to avoid* coming near his car. Likewise, you have a negative attitude toward helping others if you are angry. And so on. But if so, then the argument does not show what it purports to show. It does not show that for any nonintentional explanation there

---

[28] The ease with which we fall prey to this kind of ambiguity has been emphasized in the recently developed logic of agency. See e.g. Nuel Belnap, Michael Perloff, "Seeing to It that: A Canonical Form for Agentives," in (eds.) H.E. Kyburg, Jr., R.P. Loui, G.N. Carlson, *Knowledge Representation and Defeasible Reasoning* (Dordrecht: Kluwer, 1990), pp. 175-199.

must be an intentional one because it skids over the scope ambiguity in its fundamental premise.

In fact, little reflection is required to see that the individualist could not have hoped to make use of this argument. For intentional psychology can only predict or explain what the agent would do given that he *has* some pro-attitude.[29] The theory offers no insight into what happens when the agent *lacks* an pro-attitude.[30] So, the argument from breakdown cases not only does not but could not show that an intentional explanation must accompany any nonintentional explanation. Since the argument does not prove that it is necessary to invoke an agent's pro-attitude to explain her action, it does not show the nonindividualist interpretation of folk psychology to be incoherent.

C. The Argument from Smith's Theory of Desire

The refutation of the argument from breakdown cases indicates that there is some conceptual room for the claim that the agent need not have acted on any of his pro-attitudes. Or, at any rate, we must not suppose on such grounds that the agent must have acted on some of his pro-attitudes when performing the action. Recently, Michael Smith[31] has argued on different grounds not only that desires must be present in every action but that they are the source of all motivation. Smith presents an extremely simple argument in support of the contention that every motivating reason must include a desire and so that every instance of an action for a reason must have had its source of motivation in a belief-desire pair. He argues[32]:

(1) Having a motivating reason *is, inter alia,* having a goal.

---

[29] For this reason also, it will not do to replace the idea of "wanting" in the original argument with a "weaker" pro-attitude like "being inclined to" or "thinking that something was to be said for." Moreover, these cases do not exhibit any special features with regard to the formal structure that underlies the argument. As long as it is possible to drive a wedge between the having of a respective con-attitude and the lacking of a pro-attitude, the argument will go through. It might get more clouded in view of the "weaker" nature of the pro-attitudes. See also section 3.E, below.

[30] One could argue that while intentional regularities indeed allow us to predict only what the agent would do if he had some pro-attitudes, intentional psychology as a whole allows us to do more. For it to do so, it must be assumed that intentional psychology offers a *complete* picture of human behavior. This assumption would render the argument question-begging against the nonindividualist.

[31] "The Humean Theory of Motivation," *op. cit.*, and *The Moral Problem, op. cit.*

[32] "The Humean Theory of Motivation," *op. cit.*, p. 55.

(2) Having a goal *is* being in a state with which the world must fit.

(3) Being in a state with which the world must fit *is* desiring.

Hence a motivating reason includes, among other things, a desire. We will see that even if this argument is sound, it does not tell us *whose* desire must be included in the motivating reason. In fact, I will argue that Smith's argument cannot offer a non-question-begging way for showing that it must be the agent's desire.

The best place to begin is with the notion of direction of fit which guides Smith's account. Following Anscombe, Smith conceives of desires as states with which the world must fit. My desire that I pick up a piece of paper aims at its realization, and is realized when I pick it up. So, it is plausible to suppose that my desire that you pick up a piece of paper also aims at its realization and is realized when you pick it up. Since both my and your actions are part of the world there is no prima facie reason why only my and not your actions must fit my desires.[33]

One may object, at this point, that while the extension of the metaphor of the direction of fit prima facie makes sense, Smith's account does not rest with the metaphor. For Smith explicates the guiding metaphor in terms of a dispositional account. He identifies a desire to φ with 'that state of a subject that grounds all sorts of his dispositions: like the disposition to φ in conditions *C*, the disposition to [ψ] in conditions *C'*, and so on (where, in order for conditions *C* and *C'* to obtain, the subject must have, *inter alia*, certain beliefs)'.[34] Since a desire thus conceived is the agent's disposition to act and so to change the world according to the desire, it has the distinctive world-to-mind direction of fit.

Can our extension survive this explication? It will need to be modified, of course. Just as Smith identified α's desire to φ with the state of α that grounds α's dispositions to φ in *C*, ψ in *C'*, so we might identify β's desire that α φ with that state of β that grounds

---

[33] Of course, there has to be some explanatory connection in play. I give an account of the connection in Chapter VII.

[34] Ibid., p. 52. The original formulation is misleading since it does not allow erroneous beliefs. A pro-attitude to φ (drink gin and tonic) might be realized where the conditions *C* are such the agent believes of what is petrol that it is gin, and is accordingly motivated to ψ (drink petrol and tonic). Smith corrects it in *The Moral Problem, op. cit.*, p. 113.

all sorts of α's dispositions to φ in *C*, to ψ in *C'*. So my desire that you pick up a piece of paper is that state of mine that grounds your dispositions to, among others, pick up pieces of paper when I ask you to do so, when I expect you to do so, etc.

Whether this characterization makes sense depends on what we understand by "grounding." This idea must be cast in counterfactual terms. To say that the disposition to dissolve in water in conditions *C* is grounded in properties *P*, is to say something to the effect: were a substance with properties *P* immersed in water in conditions *C*, *ceteris paribus* it would dissolve. But the idea of "grounding" also involves an appeal to an explanatory connection between the state and the dispositions. Thus, we want to say that the microstructural properties of water which we describe as solubility can explain why a piece of salt immersed in water (in the right conditions) would dissolve.

If so, then we can cast the idea of desire in the following form

> α's desire that α φ is the state *d* of α that explains α's dispositions to φ in *C*, to ψ in *C'*,

which implies, among other things, that were α not in *d ceteris paribus* α would not φ in *C* or ψ in *C'*.[35] To require that an explanatory relation be invoked is not immediately to say anything about the explanatory relation in place. This leaves us room to understand desires directed toward others accordingly as explaining another person's dispositions:

> β's desire that α φ is the state *d'* of β that explains α's dispositions to φ in *C*, to ψ in *C'*,

which implies that were β not in *d' ceteris paribus* α would not φ in *C* or ψ in *C'*.[36]

It could be objected that Smith's theory commits us to the thought that even if in the case where β's state is explanatorily involved, we need to interject α's desire. How

---

[35] Or, at any rate, that α's φing in *C* or ψing in *C'* would be accidental.

[36] One could object at this point and argue that this is too liberal an understanding of the idea of "grounding." What Smith intends is surely to pick out some state of the individual agent that explains her dispositions to act. To this, one can respond amicably. It may very well be that this is what Smith intends since he is only concerned with pro-attitudes directed to the agent's own actions. But this accidental focus on the pro-attitudes directed at the agent's own actions hardly constitutes a reason against the nonindividualist. If Smith's account were to be used against the nonindividualist, there would have to be actual reasons for thinking that there is something wrong in thinking of grounding in this liberal manner. Smith, for one, does not produce any.

so? Well, presumably what β's state explains is α's dispositions. But for all these dispositions of α, there is going to be a state of α that is going to explain them. This state, on Smith's account, just is α's desire. So, even in the cases where the agent responds to somebody else's desire, he still acts on his own desire. And this is just what the individualist claims.

But this claim is not as innocent as it seems. We should note, first of all, that explanation can occur at different levels. It would be hard not to grant the objector that even in cases where we claim the agent's disposition to φ is naturally explained by some state of another person β, there is a level of explanation at which some state of the agent α explains α's disposition to φ.[37] Presumably, this is plausible for some physiological level of explanation. But the question is why this fact should affect the nonindividualist identification of desire. There are two options here. Either the individualist will find reasons to restrict the explanatory attention to the agent's state at the level of action (qua action, rather than qua physiological event) explanation or not. If the individualist does find such reasons then the suggested nonindividualist extension of Smith's account of desire to include others' desires directed toward an agent's actions will be unwarranted. But in such a case the employment of Smith's argument against the nonindividualist relies on having arguments against the nonindividualist already. For to suppose that there are reasons (at the level of action explanation) to restrict the search for explanatory states to the states of the agent is already to have an argument for an individualist position. Smith's argument gives no additional resources to the individualist.

If the individualist does not find reasons to restrict attention to the states of the agent at the level of ordinary action explanations then it is not clear why the nonindividualist should be in any way impressed by the insistence on the fact that the agent's body must have been in a physiological state disposed to the production of certain bodily motions. The nonindividualist should not be impressed any more than he would be by the fact that the agent's arm must have been in the right kind of causal disposition to cooperate in the carrying out of the action. The nonindividualist does not deny that the

individual's states must have been causally involved in the action, but he will object to identifying those states as the agent's desires (see section A, above). Once again, Smith's argument does not advance the individualist cause.

The upshot of the discussion is this. Smith's argument shows that when an action is done for a reason, there is a desire in play understood as having a distinctive world-to-mind fit. What Smith's argument does not differentiate between is whose mind is in play. It can be the agent's mind that the world must fit. But there is conceptual room for the thought that it can be another person's mind.

D. The Problem of Mere Happenings

It is customary to suppose that a performance is an action just in case it is intentional under some description. If we take it that a performance is intentional under some description only if it has been caused by the agent's pro-attitude and a suitably related belief, we have a straightforward problem for the nonindividualist. To the extent that a performance is an action at all, it must have been caused by the agent's pro-attitude, period.

The argument is valid, but it is not clear that its premises must be accepted. For one, there is no consensus on the precise shape of the second premise, though perhaps enough consensus could be forced against the nonindividualist. One might also reject the first premise. I will show how to do so in Chapters III-VI. It may be worthwhile, however, to sketch the shape of the account.

The core of any theory of action is the account of the distinction between an action (the agent raising an arm) and a mere happening (the arm rising on its own). There are two traditional strategies of approaching the problem. On one hand, one may characterize what it is for a performance to be an action, by appealing to the performance's intentional etiology. Alternatively, however, one may characterize what it is for a performance to be a mere happening, appealing to conditions that interfere with our agentive involvement (defeating conditions), and characterize actions as those

---

[37] Though, perhaps, one might be more wary in supposing that there is one state of the agent that explains all the relevant dispositions.

performances of the agent that are produced in the absence of defeating conditions.[38] On the former strategy, an agent's raising his arm is an action to the extent that the performance has been caused (in the right way) by some of his pro-attitudes. On the latter strategy, an agent's raising his arm is an action of his to the extent that the arm movement has not been caused by a spasm, by someone else's grabbing it upward, etc.

The appeal of the latter strategy to a nonindividualist should be clear. It allows us to drive a conceptual wedge between a performance being explainable by the agent's pro-attitudes and its status as an action. The performance's status as an action is determined by the absence of defeating conditions; thus as long as being explainable by other people's pro-attitudes does not count as a defeating condition, the threat to nonindividualism is averted.

E. Explanatory Individualism: Innocent Pro-Attitudes

So far, our main target has been the explanatory individualist's claim that the agent's pro-attitudes must explain the agent's actions. But I have also provided some reasons to open the conceptual space for explanatory nonindividualism, which allows that there are some actions that can be explained without reference to the agent's pro-attitudes. In this section, I want to consider an argument that may be taken to show that explanatory nonindividualism is false if one properly understands the attribution of pro-attitudes.

It may be suggested that I have misconstrued the role of pro-attitudes. Would it not be possible to construe pro-attitudes innocently? It is after all so natural and immediate to suppose that when someone asks me for directions, I will give directions only if I believe that he asked me for directions and only if I have a pro-attitude to comply with his request or to give him directions, etc. As long as one agrees that the belief and pro-attitude do not usurp explanatory power from the request itself, and as long

---

[38] This strategy has its roots in Aristotle's characterization of voluntary action in terms of what is not involuntary (*Nicomachean Ethics*, 1111a22-24), and has been pursued by contextualists (e.g. H.L.A. Hart, "The Ascription of Responsibility and Rights," in (ed.) Anthony Flew, *Essays on Logic and Language* [Oxford: Blackwell, 1951], pp. 145-166; A.I. Melden, *Free Action* [London: Routledge & Kegan Paul, 1961]).

as one holds a sufficiently non-phenomenological conception of beliefs and pro-attitudes, this position should be unobjectionable. The fact that it is so natural to attribute pro-attitudes to the agent in the explanation of every action, coupled with the provisos about the innocence of such attribution, might appear to support the position of explanatory individualism.

The position is not as innocent as it appears, however. The argument relies on the fact that it is very natural for us to attribute pro-attitudes to the agent. And I think this is largely right for reasons I discuss in section 4. However, this is insufficient as a ground for supporting explanatory individualism. In the absence of further considerations, the fact that it is so natural for us to attribute pro-attitudes to the agent no matter what the agent does would support normative individualism not explanatory individualism. In view of the flexibility of our intentional framework, we can always attribute zillions of pro-attitudes and beliefs, and taking into account various sorts of constraints, select a couple of attributions that fit the behavior best. But, as Davidson has reminded us, this is not sufficient to argue that the pro-attitudes thus attributed actually *explain* rather than merely *rationalize* the agent's action. An argument for explanatory individualism would require an argument that it is always possible to attribute pro-attitudes to the agent that actually *explain* the agent's behavior. The fact that it is so natural to attribute pro-attitudes to the agent merely supports the position that all actions can be rationalized in terms of the agent's pro-attitudes, i.e. the position of normative individualism. And that position is perfectly compatible with explanatory nonindividualism.

F. The Common-Sense of Nonindividualism

The distinctive nonindividualist claim is then that we can act on others' pro-attitudes just as we can act on our own pro-attitudes. We have seen that our practice does appear to support the nonindividualist picture, and that at the same time many of the arguments that might have been expected to show the nonindividualist position to be incoherent, fail. I want to close by considering once more the individualist strategy for accommodating actions that we intuitively explain by appeal to others' pro-attitudes.

The individualist has two options. First, he can consider such actions as occurring under 'normal' conditions, in which case he must suppose that the action is mediated by

the agent's pro-attitude to perform it. Second, he can consider them to be cases of the aberrant type, in which case the agent acts on another's pro-attitude and against her own. Here are two examples paradigmatic of the categories. Suppose I ask you to tell me to switch on the light. You tell me to switch on the light and I faithfully do so. It is very natural to describe such a case as one where *I* wanted to obey your command, and did so for this reason. Suppose you tell me to switch on the light, and I really want not to do so, but do it anyway for "reasons" I do not myself understand very well. Such a case belongs to the aberrant class of cases.

The cases that do not fall neatly in either of these categories are cases where the agent acts on another's pro-attitude, not against her pro-attitude but without having a pro-attitude of her own at all. (These are the cases which are conveniently obliterated by the ambiguity mentioned in section B.) Suppose that an agent rides in a bus, has no particular pro-attitude to stand one place or another, is in fact not very concerned with the ride at all. Within limits, she does not care what happens in the bus. A person comes in and asks politely "Could you, please, move over a little." The agent, of course, moves over — after all she does not care one way or another.

It is intuitively implausible to construe the agent as now having to consult her pro-attitudes as to what to do, to construe her as now having to decide whether or not she should move over. The individualist might argue that the relevant pro-attitudes need not be construed as coming into the foreground but may operate in the background.[39] We may first stomp our foot and ask, Why do we need to suppose that? Why go against the natural way of thinking about such a case? What reason does one have for insisting on this? Surely, it is not that had the agent not wanted to move she would not have done it. This argument, as we saw, relies on an equivocation on the idea of the agent not wanting to move and is quite compatible with the nonindividualist picture. And Smith's argument will not help here either because its employment would be question-begging at this point. So, why not simply adopt the natural picture? The agent moves over because the other person wants her to move over, period.

---

[39] Philip Pettit, Michael Smith, "Backgrounding Desire," *Philosophical Review* 99 (1990), 565-592.

Moreover, the proposal that the relevant pro-attitudes reside in the background seems to contradict our supposition that the agent genuinely does not care what happens in the bus. And if the individualist insists that her not-caring attitude is only an expression of her not having any pro-attitudes in the foreground, he is dangerously close to asserting that for *any* state of affairs, we have either a pro- or a con-attitude toward it — at least in the background.

It is natural to think that our commonsense understanding of ourselves involves the supposition that unless we really want not to comply with others' requests (and are strong-willed enough to carry out our wants), we generally will comply with them. This thought could be seen as embodied in the idiom of "not-minding," for instance. Sometimes when asked why we have, say, complied with another person's request, rather than answering that we wanted to do so, we say that we did not mind. This is an interesting phrase because quite literally what it expresses is not the presence of a pro-attitude but rather the absence of a con-attitude.

The individualist interpretation of folk psychological explanations abstracts from normal everyday interaction between people and begins exclusively with the perspective of the agent. Insofar as it then takes into account any interactions, it always does so through that perspective. But what exactly justifies such an abstraction in the first place? The individual perspective is no doubt very important, but why should we in thinking about ourselves abstract from our ordinary interactions? As Annette Baier reminds us, "My first concept of myself is as the referent of 'you', spoken by someone whom I will address as 'you'."[40]

4. Normative Individualism

In section 3.E, I have noted how natural it is for us to attribute pro-attitudes to the agent. We have also seen that although this fact fails to support explanatory individualism it makes the position of normative individualism very plausible. In fact, one of the reasons why nonindividualism might appear to be so implausible at first sight is because of the intrinsic plausibility of normative individualism. We have already seen

that there is no conflict between accepting even the position of explanatory nonindividualism (which allows that some of an agent's actions are explained by another person's pro-attitudes but not the agent's own pro-attitudes) and accepting the position of normative individualism (according to which it will be still possible to attribute some pro-attitudes to the agent, thus rationalizing the action). In this final section, I want to consider some further reasons that support our adherence to normative individualism, and that may lie behind a certain kind of prejudice against nonindividualism.[41]

Why do we resist the thought that there are any genuine nonintentional explanations of action? The reasons are not far to find. They lie in what we value in people. The picture of us as agents whose actions would be genuinely (irreducibly) explained nonintentionally is (by and large) not very flattering. Occasional politeness is one thing, but just imagine a housewife answering the question why she cleans the house, mends the socks, cooks the food, and so on, by (seriously) explaining that it is her social role as a housewife, and that the social role is a part of the on-going patriarchal order of things. There is something wrong (we think), even though many (and perhaps by now most) of us believe that the facts to which she would appeal are *true*, and are more than likely to indeed *explain* why she cleans house, mends socks, cooks food. So why is our explanation of her action not all right when she offers it? Why should not her saying it simply confirm our explanation?

Her explanation of her actions in terms of the patriarchal structure of the society is not all right because it is not the kind of explanation that *we want from her*. What kind of explanation do we want? G.E.M. Anscombe[42] was surely right — *we* want to know *her* reasons. While the factual claim that ordinary explanations of action always cite the *agent's* reasons is questionable, it seems nonetheless true that the reason why we find the housewife's sociologically sophisticated explanation hard to accept is that it does not give *her own* reasons to so act. To the contrary, it seems to offer reasons for her not to so act. After all, who would want to continue living in servitude?

---

[40] "Cartesian Persons," in *Postures of the Mind, op. cit.*, pp. 89-90.

[41] I say 'prejudice' in view of the fact that no reason that supports normative individualism is a reason for rejecting explanatory nonindividualism.

And it is here that we finally touch on the most important point. For why is it that what we want from the housewife is an explanation in terms of reasons, in terms of *her* reasons — for we would (hopefully) be just as dissatisfied if she explained her actions by appealing to her husband's really wanting her to clean house, mend socks, cook food? After all, we have a perfectly good explanation in sociological terms, and a really powerful one, for it explains not only one individual action but the whole tendency of women to stay at home and perform house duties (despite their potential dislikes and aversions). (No assembly of reason-explanations could claim to carry so much explanatory power.) The reason why we want an explanation in terms of reasons is not so much that we think of ourselves as creatures that *do* act on our own reasons, as that acting on our own reasons is what we *value*, that we think of ourselves as creatures that *should* act on our own reasons. A person who always can give good reasons for her actions, whose reasons make up a systematic whole, who is not easily swayed by interpersonal and social pressures, who exhibits autonomy and integrity, has, as we honorifically say, a personality, or is a *person* or an *individual*. Acting on one's own reasons is an ideal to which we aspire, and to which we expect others to aspire as well — on pain of our not valuing and respecting them as much.

It is for this reason that queries that look like ordinary why-questions, allegedly seeking an explanation of action, play quite a different function in our ordinary discourse. "Why did you do this?" when uttered in most circumstances does not necessarily seek an explanation of the action. For an explanation of action need not be offered in terms of reasons, while this is exactly what is expected in answer to the question. We might call such questions "challenges," for they do not so much inquire after the best explanation of someone's action (explanations in terms of social roles and structures are among the best) as they challenge the agent to give her reasons for performing the action. They are challenges because in case of failure to offer adequate reasons, we will be prima facie justified in not treating the agent as a personality, in respecting her less.

In the picture that emerges, there are two levels to our understanding of the concept of action and the discourse surrounding it. On the first ("base") level, our actions

---

[42] *Intention, op. cit.*

are explained by a variety of factors, which are reflected in the way in which we (disengagedly[43]) explain our actions. On the second ("superstructure") level, agents are required to present their own reasons for their actions; the agent's inability to do so might then be sanctioned by our less respectful attitudes toward him. Our reasons-discourse is thus not as friendly as it might appear at first sight. We continually challenge one another's conduct. But this is partially compensated by the great deal of charity we exhibit in our ways of thinking about actions. The very way our agentive language functions is geared toward making the individual appear in the best light vis à vis his independence of others, his strong will. Let us have a cursory look at some evidence for this suggestion.

Consider the ambiguity we have noted in rebutting the argument from breakdown cases. We will remember that we are notorious for confusing the lack of a pro-attitude with the presence of a con-attitude. There is a good pragmatic reason why this equivocation has survived: a form of words that allows us to inform others of lacks of attitude is simply not very useful since there are just too many attitudes that we lack. But whatever the reasons for its survival, we should inquire into the significance that its survival has. Perhaps the reader will not be surprised to learn that the equivocation plays a profound role in helping us aspire to the ideal of a person. How so?

Consider the law of excluded middle as applied to the having of a pro-attitude. It is presumably true that:

(LEM) for any action, either it is the case that the agent wants to perform it or it is not the case that the agent wants to perform it,

More idiomatically:

(lem) for any action, the agent either wants to perform it or does not want to perform it.

Our slick equivocation allows one to render (lem) as the false (PPA):

---

[43] In contexts where the question of respect does not arise or is subdued. For instance, this will happen if one tries to explain someone else's actions (his or hers, not yours or mine), or in settings where people trust each other, or where the actions involved do not carry much significance — politely moving over on a

(PPA) for any action, the agent either has a pro-attitude toward performing it or the agent has a con-attitude toward performing it.[44]

We might call this rendition of (LEM) the principle of polarization of attitudes, for what it licenses us to do is to attribute to the agent *some* attitude (whether pro- or con-) for *any* action. This is important for in view of the ideal to which we aspire, the worst that could happen is if the agent had no attitude, was indifferent. — You ate spinach, so you must have liked it, because had you not liked it you would not have eaten it; you did not eat spinach — so you must have disliked it, because had you liked it you would have eaten it.[45] The possibility of your having simply eaten the spinach, without having shed one thought, like or dislike, vanishes under the universal reign of (PPA). (PPA) makes sure that you stand behind your actions, that your attitudes *reflect* your actions.

While this ambiguity is helpful in supplying you with attitudes you might not have had, and so forms an integral part of the charitable discourse, ambiguities are not usually free and give rise to various kinds of troubles and tensions. Such is the case also here. The most common type of conceptual tension is that our intentional vocabulary sometimes stands in the way of our describing phenomena we are quite familiar with. Let us mention three of them: altruism, weakness of will, and servitude.

Perhaps the most famous conceptual tension lies behind the debate between those who believe that we are capable of altruistic actions and those who believe that we are not. Although much more is involved, one argument nicely summarizes the issue:

> With regard to altruism, the ... intuition is that since it is I who am acting even when I act in the interests of another, it must be an interest of mine which

---

bench (just because someone asks) is something we understand, but politely killing someone (just because someone asks) is not something we understand.

[44] Matters are slightly more complicated. A con-attitude toward performing an action *A* can be interpreted as a pro-attitude toward performing a (negative) action not-*A*. Thus, excluded middle holds for any (positive or negative) action: for any *A*: the agent either wants to perform *A* (has a pro-attitude toward performing *A*) or does not want to perform *A* (lacks the pro-attitude toward performing *A*); for any not-*A*: the agent either wants to perform not-*A* (has the con-attitude toward performing *A*) or does not want to perform not-*A* (lacks the con-attitude toward performing *A*).

[45] Of course, it is possible for you to offer *another* reason (like the fact that you did not want to be rude), i.e. to exhibit another attitude, but exhibit an attitude you must.

> provides the impulse. If so, any convincing justification of apparently altruistic behavior must appeal to what *I* want.[46]

The very attempt to formulate what an altruistic action is, viz. action done for the sake of another, seems doomed because we must understand the action as done because of what the agent wants or intends. (After all, had he not wanted to...) And if so then his action must be conceived as furthering the agent's end (even if that end will be to further another's end), and must ultimately be conceived not as an altruistic action as might have been thought but as an egoistic one. The paradox of altruism is interesting because it arises out of nowhere, out of the very way that the vocabulary functions, and yet contrary to the thoughts that are to be conveyed. Of course, one may take this fact to show that we indeed are egoists, or one may try to specify the kinds of wants that could be candidates for confirming that we are egoists. But one may also try to look back at the phenomena and juxtapose a greedy businessman and someone who stakes his life for the life of another. It is when one does the latter and hears someone insisting that both are egoists in *some* sense that Wittgenstein's diagnosis of our language sometimes going on a holiday seems the most appropriate. But it is more than a holiday. There is a deeper purpose that this function of the intentional language is designed to play, viz. to present the individual agent as autonomous master of his actions.

A similar tension has been involved in the conceptualization of the very phenomenon of weakness of will. When we imagine an akratic agent who resolves not to φ, is fully motivated not to φ, and then φs, we are almost immediately drawn into supposing that he must have wanted to φ in some sense. (After all, had he not wanted to...) Perhaps a momentary desire to φ, a momentary change of mind, governed his action, so that his action was not weak-willed after all. And indeed if one looks at particular cases of akratic actions, it is very tempting to reconstruct them in ways that turn the weak-willed into strong-willed actions. As a result, we are more confident in the existence of akrasia as a phenomenon than in the existence of particular instances of akratic actions. Once again, our skill in interpreting actions as strong-willed is remarkably consistent with our charity toward the individual.

---

[46] Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970), pp. 80-81.

One last example of a conceptual tension involves cases of undue influence of others on the agent. On one conceptualization of such cases, most of us have a tendency to respond with hostility if exposed to continued acts of malevolence on the part of another. However, there comes a point where if the acts of malevolence increase in intensity our tendency to respond with hostile actions becomes broken and we tend to respond with benevolent acts.[47] The telling examples here involve cases of people who have been "broken" (the best literary example being Winston Smith[48]): prisoners, soldiers, mental patients, women, slaves, subjected to mental, physical, and situational torture. When a person in such a situation behaves with benevolence toward her oppressor, we want to interpret the action as servile. But when we try to understand the action intentionally, the agent acting because she *wants* to be benevolent or even because she *wants* to be servile, the characterization of the action as servile seems threatened. It is almost as if we want to say that she is within her rights to do as she wants, and if *she wants* to behave in that way toward her oppressor that is *her* privilege. But if this is the psychological portrait of the agent then she seems to be a strong-willed person, not servile at all. Once again, the intentional explanation seems to turn around the intuitions that we harbor about the phenomena.

In these three cases, of altruism, of akrasia, and of enslavement, we see a tendency for intentional explanations of actions to falsify our intuitions about the phenomena. It is as if our intuitions are hard to express in intentional terms. That this is so is only to be expected if the very ways of our language were geared toward furthering the picture of ourselves as autonomous, strong-willed, independent agents. In truth, we do not always conform to this picture, or perhaps even only rarely do so. We do things out of habits, on others' orders and expectations. Of course, this is not to say that we

---

[47] L. Nowak, *Power and Civil Society, op. cit.* and "Man and People," *Social Theory and Practice* 14 (1987), 1-17. Nowak suggests that aside from the relatively "normal" areas of human interaction where the agent responds with malevolence to malevolent actions and with benevolence to benevolent actions, there are two "abnormal" areas: of enslavement, where the malevolence of the other is sufficiently large that the agent responds with benevolence, and of satanization, where the benevolence of the other is sufficiently large that the agent responds with malevolence. His model is indirectly confirmed by constituting the foundation for his general theory of real socialism.

[48] George Orwell, *Nineteen Eighty-Four* (New York: Harcourt, 1949).

never do things because we want to do them; just that we do not always do things because we want to do them (under whatever description).

Since our practices (including the relevant linguistic practice) are geared toward being most charitable to the individual, to making him appear in the best possible light, we can at least understand why the intentional categories seem inescapable. This is because the intentional categories are inescapable for us in most of our interactions with others — because the way that our intentional categories function is geared toward the greatest charity to the individual and his point of view.

But one might wonder why that is so. So far, it looks as if our individualism is a cultural phenomenon, specific perhaps only to our cultural milieu. However, one might, with some right, think that individualist theories of action explanation, in more or less explicit ways, have a genuine claim to universality. But if this is so, if different cultures to a greater or lesser extent privilege the individual's perspective, then such consistent privileging must seem like a cosmic coincidence from our point of view. Indeed, the question why this is so is among the most difficult, and it will be hard to do justice to it here. But let us glance at the answer.

The reason why the individual and the individual's perspective is so privileged in our thinking about action is the fact that the nature of action is nonindividualist. What actions are is frequently determined by circumstances external to the agent. This is so in what we call unintentional action, where the agent is forced to recognize an entirely unintended deed as his deed nonetheless. Intentionalists will claim that this is all right and shows little because although the agent does not intend *that* deed (does not intend to perform the action under *that* description), he nonetheless does intend some other deed (intends to perform the action under some other description) which, on the occasion, happens to be identical with the unintended one. However elegant, this attempt to accommodate unintentional actions as intentional actions under other descriptions obliterates something very fundamental to agency. For being an agent, first and foremost, involves *taking responsibility* for what one does or does not do and for what happens as a result of one's doing or not doing. But taking responsibility is frequently at odds with one's intentions, and with one's (fore)knowledge of the consequences. This is very clear in what one might — from this point of view — see as a paradigm of our

agentive involvement with the world: in unintentional omissions. When one simply fails to show up at a meeting with one's friend, perhaps because one forgot in the rough and hectic time, perhaps because one was so tired that one simply fell asleep and missed it, one *does* something that affects one's friend — one wastes his time in the very least, perhaps upsets him as well. One is rightly held responsible for doing so. And one is rightly expected to recognize it as something one did by taking and acknowledging responsibility for it — whether by apologizing or by excusing oneself.

If this indeed is the angle from which we ought to look at agency then what will be noted immediately is that the individual does not hold a prominent position in determining what counts as an action of his, in determining for what he may be held responsible. When one sleeps sweetly after a hard day, it seems almost silly to think that one can *do* something (*perform* an action) at the time — and a nasty one at that, of wasting a friend's time. And yet, the friend's time *is* wasted. And not by chance, but by one's carelessness. It is the recognition of the way in which we affect the world, in particular the social world, that constitutes the core of our idea of agency. As such, the very idea of action (as part of conduct) is not geared toward the individual but rather toward the *responsibility* the individual has toward others.

It will be equally obvious, however, that the very idea of action so understood will seem *unfair* to the agent. And it is to compensate for this that our *thinking* about agency is geared toward the individual and the individual's point of view. It is because the idea of action is geared toward others and the way that they are affected, that we try to make the individual look in the best possible light vis à vis his actions. It is because we require so much of the agent, because we require of her first and foremost to think about how others are affected (and if not to think about it then at least take responsibility for it), that it seems only fair then to take her point of view as central in our evaluating the action.

But if this is the case then, as theorists of action, the worst we can do is to become impressed by the apparent inevitability of our intentional understanding of actions. The worst guide for understanding the nature of action lies precisely in what we find inescapable in our understanding of it. For this urgency with which intentional categories swarm our picture of action is due to the fact that they must compensate for what we really take actions to be, for what they really are. The intentional picture of action

constitutes the best-intentioned false consciousness but a consciousness that is false nonetheless. And however painful it may be, as theorists of action, we must recognize that *our* primary obligation is to the truth not to the individual.

•

The purpose of this chapter was to clear some of the initial resistance one might have toward the account of action presented in Chapters III-VI, as the account does not exclusively refer to the pro-attitudes of the agent, but also to the pro-attitudes of others, viz. their normative expectations of the agent. I have suggested that some of the resistance against such a view might stem from what I have called individualism about action explanation, i.e. the view that intentional explanations of action (that appeal to the agent's pro-attitudes) are privileged over nonintentional explanations of action (that appeal to the pro-attitudes of people other than the agent).

I have distinguished two kinds of individualist positions that may be advanced in a reductive or non-reductive spirit. First (non-reductive normative individualism), the individualist may assert that all of an agent's actions must be rationalized in terms of the agent's pro-attitudes. Second (non-reductive explanatory individualism), the individualist may make the stronger claim that all of the agent's actions may not only be rationalized but also explained in terms of the agent's pro-attitudes. Both positions may be advanced in a reductive spirit asserting in addition that only the agent's pro-attitudes may rationalize the agent's actions (reductive normative individualism), or that only the agent's pro-attitudes may explain the agent's actions (reductive explanatory individualism).

I have further characterized two different nonindividualist positions, declaring that the dissertation is written in the spirit of explanatory nonindividualism, which is incompatible with either version of explanatory individualism but is compatible with non-reductive normative individualism. My aim in this chapter has been accordingly to show that there are grounds to resist arguments that support explanatory individualism. In Chapter VII, I will argue for the tenability of an explanatory nonindividualism (by showing how it is possible to act on others' pro-attitudes). I will need to demonstrate that the explanatory relation proposed will obtain not only between an agent's expectations of

herself and her action but also that it can obtain between others' expectations of the agent and her action.

The discussion ought to have given some reason to reject the popular notion that any adequate account of action must appeal to intentions, beliefs, desires, in one form or another. In this way, the idea of a responsibility-based approach to action ought not to appear as intuitively foreign as it may have at first sight.

CHAPTER II.

THE CHALLENGE OF HART'S THEORY OF ACTION

In Chapter I, we have seen that according to explanatory individualism explanations in terms of the agent's pro-attitudes are privileged over explanations in terms of others' pro-attitudes toward the agent. This may be thought to support the position that the agent's intentional attitudes ought to be used as the primary categories in understanding the nature of action.[1] We saw that while there are some strong arguments supporting such an interpretation of folk psychology, they do not in fact force it on us. I argued that an alternative nonindividualist understanding of folk psychology can at least be seen as a contender.

The main aim of the present chapter is to prepare some ground for the responsibility-based account of action developed in Chapters III-VI by drawing some lessons from H.L.A. Hart's account of action in terms of responsibility ascriptions. I consider and address in a preliminary way the major objections that have been raised against Hart's theory, and take others as challenges of adequacy for the account to be developed.

Section 1 sketches two traditional strategies that theorists of action can employ to draw the distinction between actions and mere happenings. In section 2, I present the main theses of Hart's theory. The further sections will be devoted to the discussion of the major criticisms of Hart's view. In section 3, I shall consider an objection that might be

---

[1] This transition, which is in effect a transition from a theory of action explanation to a theory of action, is very common. This is to say nothing about its legitimacy. In particular, I do *not* claim that anyone subscribing to individualism about action explanation is thereby *committed* to analyzing the nature of action in terms of the intentional attitudes of the agent. I owe this point to J. McDowell.

thought to undermine a responsibility-based approach in its very foundations. The charge is that a responsibility-based theory of action reverses the proper logical order of the concepts of action and responsibility — responsibility is a concept that is logically secondary to the concept of action (we are responsible *for* actions, after all) and so cannot be thought to precede it. We will see that there are at least three different ways of disarming the objection while acknowledging the thought that underlies it. In section 4, I consider Geach's famous criticism directed against the ascriptivist nature of Hart's theory.

1. Two Kinds of Action Theories

> What comes about by force or because of ignorance seems to be involuntary. What is forced has an external origin, the sort of origin in which the agent or victim contributes nothing — if, e.g. a wind or human beings who control him were to carry him off.[2]

As far back as Aristotle, it has been recognized that there are certain circumstances that interfere with our agency, like being pushed by someone or something, being physically forced to do something by someone, something, or the state of one's own body or mind, etc. Aristotle described those cases as ones where the principle of action is not in the agent.[3]

Aristotle's account is suggestive of a certain natural picture of what it means for a performance to be a mere happening rather than an action:

(e) The agent's φing was a mere happening (non-action) iff external forces caused him to φ.

This may be thought to generate a corresponding picture of what it means for a performance to be an action:

(i) The agent's φing was an action iff internal forces caused him to φ.

---

[2] Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis: Hackett, 1985), 1110a1-4.

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

In fact, however, (i) does not follow from (e), nor (e) from (i). What is indisputable is the fact that a performance is an action just in case it is not a mere happening. If so, then what follows from (e) is:

(i^e) The agent's φing was an action iff it was not caused by external forces.

Analogically, what follows from (i) is

(e^i) The agent's φing was a mere happening iff it was not caused by internal forces.

This allows us to see at least two strategies a theorist of action can follow. One might begin with the idea of what it means for external forces to cause an agent's performance (e) and then explain what it means for the agent to act by appealing to the absence of such forces (i^e). This is the strategy of responsibility-based accounts of action. From this point of view, the idea of internal forces causing the performances is a hypostatization of the absence of such causation by external forces. Or, alternatively, one might begin with the idea of what it means for internal forces to cause an agent's performance (i) and then explain what it means for the agent's performance to be a mere happening in terms of (e^i). This strategy is typical of explanation-based accounts of action.

Of course, on neither strategy must one begin with the notion of internal forces causing a performance or of external forces causing a performance. A theorist may seek to *explicate* these concepts further. And so, causal theorists of action[4] aim to understand what it means for a performance to be "caused by internal forces" in terms of the idea of being caused by mental states in the right way. Some teleological theorists of action may seek to understand what it means for a performance to be "caused by internal forces" in terms of the performance being suitably teleologically related to the agent's intentions

[4] Recall that I use the term to cover those who aim to understand the concept of action in terms of the concept of being caused by mental states, not to cover those who (like Davidson) argue that the force of action explanations is causal. The latter are causal theorists of action explanation, not necessarily causal theorists of action.

and not in causal terms at all.[5] The idea of a performance being "caused by internal forces" is thus taken to be a metaphor that is further explicated. Similarly, for responsibility-based approaches the idea of being "caused by external forces" need not be taken as a given but may be explained further.

The account I will offer is a responsibility-based account. Rather than analyzing the notion of action in terms of its relation, causal or otherwise, to the agent's reasons, I will analyze it ultimately in terms of whether it was reasonable (in a special sense to be explained) to expect a performance of the agent under some description (Chapter VI). As we shall see, the presence of defeating conditions makes it unreasonable to expect of the agent that she perform the action under any description (Chapter V). In the present chapter, we will have a closer look at a responsibility-based account proposed by H.L.A. Hart. I will respond to some of the criticisms launched against it and formulate challenges that the account to be proposed will have to answer.

2. H.L.A. Hart's Theory of Action

It is one of the main criteria of adequacy for any theory of action that it should account for the distinction between actions and mere happenings. This is usually done by conceiving of the distinction in ontological terms. While Hart does not deny that there is a distinction between actions and mere happenings, he proposes to change its status. Rather than thinking about the distinction as pertaining to two kinds of entities (events), he suggests that we ought to think about it as being normative in nature. It is a distinction between two ways in which it is appropriate to treat certain events.

It is customary to interpret a claim like "John broke the glass" as describing an event, a very special kind of event — an action. The special kind of event, the action, is sometimes considered to be ontologically distinct from another kind of event that, on its surface, may appear to be very similar, John's spasmatic movement of the arm breaking

---

[5] An example of a theorist who defends a teleological account of the intentionality of action is George M. Wilson, *The Intentionality of Human Action* (Stanford: Stanford University Press, 1989). It would be inappropriate, however, to take his theory as a theory of action. While Wilson does believe that all action is intentional under some description, this cannot be taken to analyze the concept of action. One structural reason is that his account actually presupposes the distinction between voluntary and involuntary behavior (i.e. in our terminology: action and mere happening).

the glass, for example. The latter is not John's action, it is a mere bodily movement, a mere happening. Hart, by contrast, proposes that claims like "John broke the glass" not be interpreted as describing an action but rather as ascribing responsibility to the agent (here: for the glass breaking). Action claims are ascriptive rather than descriptive. They are never true or false; they may only be appropriate or inappropriate in view of relevant conditions. Their function is to ascribe responsibility to the agent. Transposed from the formal into the material mode, there are no actions among the ontological furniture of the world.

What distinguishes actions from mere happenings, on Hart's view, is not any ontological fact, but rather the appropriateness of ascribing responsibility for events in certain conditions (when we intuitively think of them as actions) and the inappropriateness of ascribing responsibility for events in other conditions (when we intuitively think of them as mere happenings).[6] This is what it means to say that the distinction between actions and mere happenings is normative in nature.[7] But this is not yet to give an account of the distinction. In fact, Hart never does give a complete account of the distinction but rather notes that there are conditions that contribute to it being appropriate or inappropriate to ascribe responsibility to the agent.

The structure of action attribution is characteristically defeasible. First, there are, in Hart's terminology, positive conditions that establish the prima facie applicability of the responsibility attribution. In our example, such conditions include John's arm moving in such a way as to break the glass. Second, there are negative (defeating) conditions that defeat the prima facie appropriateness of ascribing responsibility to the agent. Such conditions include John's arm moving because of a spasm.

This structure allows us to understand the difference between actions and mere happenings or between it being appropriate and it being inappropriate to ascribe

---

[6] He compares Wittgenstein's question "What distinguishes the physical movement of a human body from a human action?" to the question "What is the difference between a piece of earth and a piece of [real] property?" See H.L.A. Hart, "The Ascription of Responsibility and Rights," in (ed.) Anthony Flew, *Essays on Logic and Language* (Oxford: Blackwell, 1951), p. 161.

[7] The characterization of the distinction as normative rather than ontological ought not to indicate that it is impossible for such a distinction to be construed as being both. Hart does appear to be denying, however, that we ought to construe the distinction in ontological terms.

responsibility to an agent. It will be inappropriate to ascribe responsibility to the agent if either no positive conditions are present or while the positive conditions are present some defeating condition occurs. It will be appropriate to ascribe responsibility to the agent if the positive conditions occur and no defeating conditions are present.

It is important to bear in mind that Hart's view reverses a natural way of construing the relation between the concept of action and that of responsibility. This is nicely brought out by considering how easy it is to misunderstand Hart's project. George Pitcher's critique, apart from making many valuable points, is a nice and helpful illustration of how not to understand Hart. Hart tells us that we should understand action claims as ascriptions of responsibility. Pitcher asks, But responsibility for what?

> Let us look more closely at Hart's example: we are told that when one says "Smith hit her," he ascribes responsibility to Smith. But for what is Smith supposed to be responsible? ...At the beginning of his article, Hart tells us what Smith is responsible for, namely, his action.[8]

And Pitcher cites Hart where he indeed uses this unfortunate form of words.[9] It is telling, however, that Pitcher has to look at the very first page of Hart's article, where Hart describes his venture for the first time. Hart never again speaks of the agent being responsible for an action. Instead he simply speaks of ascriptions of responsibility, and does not really tell us for what the responsibility is ascribed. Since we usually think that one is responsible *for something* Pitcher's query is well justified. However, the suggestion that Hart must mean "responsible for actions" (on the grounds of the non-committal statement on the first page) is a fundamental misunderstanding of Hart's project.[10] In fact, Pitcher and another of Hart's critics, Joel Feinberg, offer as their

---

[8] George Pitcher, "Hart on Action and Responsibility," *The Philosophical Review* 69 (1960), p. 226.

[9] The quote reads: "...sentences of the form 'He did it' have been traditionally regarded as primarily descriptive whereas their principal function is what I venture to call *ascriptive*, being quite literally to ascribe responsibility for actions." (H.L.A. Hart, "The Ascription of Responsibility and Rights," *op. cit.*, p. 145.)

[10] Pitcher's actual objection concerns the fact that Hart is mistaken about what we can be responsible for (namely our actions), and he goes on to argue that we can only be responsible for the consequences of our actions. This position has been challenged by among others Feinberg who claims that we can also be responsible for our actions ("Action and Responsibility," in (ed.) Alan R. White, *The Philosophy of Action* [Oxford: Oxford University Press, 1968], pp. 95-119).

suggestions for an improved account, not an account of action but an account of responsibility for action.

Hart's project is to understand the very idea of action in terms of the idea of responsibility. This reverses the natural way of proceeding. For it is rather natural to think that ascriptions of responsibility are founded on our knowledge whether a person acts or not. Hart's position undercuts this thought. Whether a person acts or not is not a matter of checking whether a particular event that is the agent's action exists. He argues extensively that in the law, the question whether action is intentional or not is settled by appealing to the legal code for ascribing responsibility and defeating prima facie descriptions by appeal to prototype cases. By contrast, Hart proposes to begin with the notion of responsibility and to construct the notion of action from it. It might be helpful to resort to a sketch of the metaphysical structure of his project. At the first stage, let us imagine that there are events or facts. Some of these events or facts will matter for the ascription of responsibility, i.e. for the attribution of action. In order to bring out how they matter for action attribution, we first need to postulate actors (people, firms, etc.). Actors participate in the causal order of the world, their limbs move, their mouths close. They are not yet capable of performing actions, however. It is only when we take them to be embedded in complex normative practices that we are given the tools to understand what actions are. To attribute an action to an agent is to tie the occurrence of a certain event under the presence of some positive conditions (and the absence of defeating conditions) to the agent. The tie in question is normative, it is the responsibility relation. The agent is responsible for a certain state of the world — if appropriate conditions hold. The agent's being responsible for that state of the world just *is* the agent's having performed an action. So it is sloppy language at best for Hart to say that the agent is responsible for his action (at least in his theoretical voice). But as we will remember Hart does not say this in his theoretical voice, he says it only in the introduction to his paper.

At least four further objections can be, and have been, directed against Hart's view. (1) Peter Geach has argued that action claims cannot be construed as having exclusively ascriptive uses. This objection shapes one criterion of adequacy for any responsibility-based account of action: to be able to show that action claims are not precluded from having descriptive uses. I will discuss Geach's objection and point to a

way in which the account of action I will propose avoids it in section 4. (2) A number of critics have complained that Hart says very little about the notion of responsibility involved. It seems clear that the notion of responsibility that is to constitute a foundation for a theory of action must not be the notion of legal responsibility (lest it ground a theory of legal action) or of moral responsibility (lest it ground a theory of moral action). It is thus imperative to develop a notion of practical responsibility. I will clarify such a notion in Chapters III-V. (3) Hart says rather little about both the positive and negative conditions that underlie the propriety of responsibility attributions.[11] The object of Chapter V will be in part to remedy this failing. (4) Last but not least, I must address an objection that can be thought to be fatal to any responsibility-based account of action, viz. that it reverses the logical order of the concepts of responsibility and action. Let us begin with considering just this objection.

3. The Fundamental Problem: The Concept of Action is Prior to the Concept of Responsibility

Although the concept of action is closely connected with the concept of responsibility, the latter is usually not thought of as having the potential to illuminate the former. Indeed, if anything it is the other way around. After all, in most instances, we base our judgments of responsibility on our judgments about actions. To claim that John is responsible for breaking the glass we must know that it is John who broke it, that he *did* it. This conceptual order seems to be also reflected in the very way we use the concept of responsibility: we are paradigmatically responsible for our actions. The problem then is this: How can a person's action be understood in terms of whether it is

---

[11] In a later paper ("Acts of Will and Responsibility," in *Punishment and Responsibility* [Oxford: Oxford University Press, 1968], pp. 90-112), Hart characterizes involuntary movements as those "which occurred although they were not appropriate" (p. 105). He cashes this idea out in terms of whether the movements are "subordinated to the agent's conscious plans of action," whether they occur "as part of anything the agent takes himself to be doing." This is, however, too weak, for the agent can do something involuntarily though it may (by accident) fit what the agent planned to do. I may want to drop a spoon by way of giving an agreed on signal to my partner, but when I set about dropping the spoon, my fingers may tremble and the spoon may fall out by accident. While the falling of the spoon is certainly consistent with my plans, it has nevertheless fallen by accident through no action of mine.

appropriate for her to be held responsible if whether or not it is appropriate for her to be held responsible depends on whether or not she has acted?[12]

There are at least three different (though mutually compatible) ways for a responsibility-based approach to temper the intuitive burden of the objection, two of which I shall endorse. First, one might deny that our judgments of responsibility are based on our judgments about actions. This is the road taken by Hart. On Hart's view, judgments about responsibility are not based on judgments about actions but rather on judgments about the presence of positive and absence of negative (defeating) conditions. In view of the fact that Hart does not say very much about these kinds of conditions, this way of resolving the problem might not seem too inviting. It does, however, show the objection not to be fatal.

But there are at least two other ways of showing that responsibility-based accounts are not based on a fundamental error, while preserving the natural belief that judgments of responsibility are based on judgments regarding actions. The first of these begins with the observation that the concept of responsibility comes in various flavors, three of which I have already mentioned: legal, moral and practical. I have claimed that a responsibility-based theory of action (rather than of legal or moral action) must appeal to the notion of practical (rather than legal or moral) responsibility. If so, then it is no longer clear that the problem is indeed as fundamental as it seems at first sight. There is prima facie nothing incoherent in thinking that our judgments about moral or legal responsibility are in part based on our judgments about actions and that our judgments about actions are based on our judgments about practical responsibility. The one disadvantage of such a response is that the account of practical responsibility would have to be different from (and not modeled on) the account of either moral or legal responsibility. While the latter can presuppose that the agent acts, the former cannot. Chapters III-V are devoted to developing a concept of *practical* responsibility that does not depend on the concept of action.

---

[12] Christopher Cherry formulates a version of this objection directed specifically at Hart's view: "Hart's account is incoherent to the extent that it is framed in terms of ascribing responsibility for *actions* — as it mostly is. For the upshot is that a non-responsible action is a contradiction-in-terms" ("The Limits of Defeasibility," *Analysis* 34, 1974, p. 106).

Third, the concept of responsibility not only comes in various flavors but there are in fact very different senses of the concept. Broadly speaking, there are three different categories of concepts of responsibility: all-encompassing, forward- and backward-looking. The concept of accountability[13] is an all-encompassing responsibility concept. When we speak of normal adults as people who can be held responsible and contrast them to the mentally ill or the minors, we take them to be accountable. The concept of task-responsibility is a forward-looking responsibility concept. The agent who is held task-responsible for the performance of an action is held to the task of performing the action at some future time.[14] For example, a captain is held task-responsible for the safety of the passengers. K. Baier distinguishes three different concepts belonging to the last category of backward-looking responsibility concepts: answerability, culpability and liability. What they all have in common is the fact that they presuppose that there is something the agent has done (or not done) for which she is held responsible. They look back toward the action. The agent is answerable for an action as long as she has performed the action. She is culpable for the performance of the action if she is answerable for it and no excuses apply. If the agent is culpable, she is liable to punishment, condemnation or payment of compensation.

Given this three-fold classification of responsibility concepts, we can immediately tell that the fundamental problem involves only one of the categories. If we agree that responsibility judgments are based on judgments about actions (and so disagree with Hart's solution to the fundamental problem, see above), we take responsibility in its backward-looking sense. It is thus not open to such a responsibility-based theorist of action to construe the concept of action in terms of any of the backward-looking responsibility concepts. But this is to say nothing about the other kinds of responsibility concepts. The fundamental objection is so much as an objection only for an account that

---

[13] I follow the terminological distinctions made by Kurt Baier in "Moral and Legal Responsibility," in (eds.) Mark Siegler, Stephen Toulmin, Franklin E. Zimring, Kenneth F. Schaffner, *Medical Innovation and Bad Outcomes* (Ann Arbor, MI: Health Administration Press, 1987), pp. 101-129. See also Kurt Baier, "Responsibility and Action," in (eds.) Michael Bradie, Myles Brand, *Action and Responsibility* (Bowling Green, OH: Bowling Green University Press, 1980), pp. 100-116.

[14] It might be objected that the fundamental problem is merely postponed. After all, what we are task-responsible for is an *action*. In Chapter III, I will show how to circumvent this objection.

understands action in terms of a backward-looking concept of responsibility. In Chapter VI, I shall propose to understand the concept of action in terms of the forward-looking concept of task-responsibility. As such, the objection is no objection to the account to be offered. Nothing will stand in the way of our admitting that ascriptions of moral and legal (backward-looking) responsibility are based on attributions of actions.

Contrary to appearances, the objection that a responsibility-based theory of action is doomed because it reverses the logical order of the concepts need not be fatal. The belief that gives rise to it (denied by Hart's theory) is that our judgments of responsibility for actions are based on whether or not the agent performed the action in question. While Hart's account points out at least a direction of thinking that does not rely on this belief, there are at least two other ways of saving the belief by distinguishing flavors and kinds of responsibility. There is nothing incoherent about the claim that judgments as to whether an agent is morally or legally responsible are based on judgments whether the agent acted or not, which in turn are based on the judgment whether the agent is *practically* responsible. Moreover, the objection does not appreciate that there are many different kinds of concepts of responsibility. As suggested, an account of action that is based on a forward-looking concept of task-responsibility does not stand in conflict with the thought that backward-looking responsibility judgments are based on judgments about actions.

4. Against Ascriptivism

The most piercing, if brief, criticism of Hart's ascriptivism is due to P.T. Geach.[15] Geach argues that whether a claim is ascriptive rather than descriptive is easily settled by appeal to what some authors have since called the Frege-Geach test. The test relies on the distinction between the content of a statement and the force that attaches to the content. Frege noted that there are some contexts where assertoric force attaches to the content, as in free-standing occurrences of a statement, when the statement is asserted. But there are other contexts where the content is stripped of assertoric force, as in the antecedent of a conditional. When one sincerely says "This box weighs 50 kg" one

is asserting the content, claiming it to be true that the box weighs 50 kg. On the other hand, when one sincerely says "If this box weighs 50 kg then I will not be able to carry it alone" what one is asserting or claiming true is the whole conditional not its antecedent. The force that attached to the content "This box weighs 50 kg" when the first assertion is made no longer attaches to it when one asserts the conditional.

Geach's suggestion is that we can check whether a claim is ascriptive or descriptive by seeing not how it behaves in free-standing contexts but how it behaves in embedded contexts. Suppose that action claims indeed do not have any descriptive content but have a purely ascriptive function. Their only role is to express the attitude of the ascriber toward the responsible person. If that were so then the conditions in which they were to play the role of antecedents, for instance, would be senseless. Asserting a conditional like "If she plays the piano, the sky whitens" would be like asserting the conditional "If dfalfkj aldkfajf, the sky whitens." In other words, there ought not to be any meaningful conditionals with action claims in the antecedents. But surely action claims can be embedded. The conditional "If she did it then he did not do it" is perfectly intelligible and important in our practices. Geach observes also that one had better not take the line that a different sense of doing is involved when action claims are embedded in subsentential contexts. For this would lead to the disastrous consequence that one could not apply modus ponens to the conditional without equivocating. And we surely do want to uphold the inference from "If she did it then he did not do it" and "She did it" to "He did not do it."

Geach's argument is simple and persuasive. I will not launch a full defense of Hart, though such a defense might be called for if only for historical reasons since Hart does occasionally mention that there are descriptive uses of action claims. He does not, however, tell us how we are to understand them.[16]

---

[15] "Ascriptivism," in *Logic Matters* (Berkeley: University of California Press, 1972), pp. 250-254.

[16] J. Feinberg ("Action and Responsibility," *op. cit.*, see in particular pp. 110–117) attempts to rescue Hart's insights by suggesting that there is a sense of 'ascriptive' in which action claims are ascriptive, but at the cost of abandoning the connection with responsibility. Feinberg suggests that the feature of ascripitivity that Hart really intended to capture had to do with the fact that there is a certain degree of discretion in making the judgment: the action claim is not forced by facts. Feinberg then goes on to propose that one finds a great deal of discretion in making causal judgments, in singling out the causes of an event that

One possibility of developing Hart's view might be to suggest that the descriptive uses of action claims correspond to assertions of propriety of responsibility ascriptions. On such a view, a speech act of the form "α φed" serves to ascribe responsibility for φing to α. To say that "α φed" is ascriptive rather than descriptive is *inter alia* to say that it cannot be true or false, and as such it cannot enter into embedded contexts. However, while attributions of actions understood as ascriptions of responsibility cannot be true or false they can be appropriate or inappropriate (depending on the presence or absence of positive and negative conditions). If so, then it is open to Hart to propose that action claims have a derivative descriptive content. The force of an action claim of the form "α φed" is to ascribe to α responsibility for φing in circumstances *C*. The derivative content of that action claim could be paraphrased as "It is appropriate to ascribe to α responsibility for φing in circumstances *C*." It is thus possible to paraphrase conditionals such as "If she did it then he did not do it" as "If it is appropriate to ascribe responsibility to her in circumstances *C* then it is inappropriate to ascribe responsibility to him in circumstances *C'*."

Such a supplementation of Hart's view seems to solve the problem without compromising the spirit of Hart's account. It preserves Hart's rejection of the ontological division into events that are actions and events that are not. The descriptive content of action claims does not describe actions but describes the propriety of responsibility attributions.

We have seen that while Geach's objection is powerful, there is at least a direction which if followed could save Hart's account. But responsibility-based theories of action need not be ascriptivist. In what follows, I will develop an approach that is not ascriptivist in aspiration, and as such is immune to Geach's objection.

---

contributed more and less to its occurrence, and in this (causal) way salvages Hart's insight. Aside from the fact that it is most certainly not true to Hart's intention, it seems like a rather far-fetched extension of Hart's view. Moreover, from our standpoint it misses the virtue of Hart's view — the tie of action to responsibility.

• •

In this chapter, I have discussed H.L.A. Hart's view according to which attributions of actions can be understood in terms of appropriate responsibility ascriptions. This discussion, in particular the consideration of the criticisms of Hart's account, allowed us to formulate some criteria of adequacy for any responsibility-based account of action.

In section 3, I have shown how a responsibility-based account of action can avoid the objection that it is based on a fundamental error because the concept of action is prior to the concept of responsibility. We have seen that a responsibility-based account of the notion of action is quite compatible with the thought that whether or not we are to be held responsible (in a backward-looking sense) for actions is to be determined in part by appeal to whether we have acted or not. In Chapters III-V, I will develop a forward-looking concept of practical task-responsibility, which is immune to the fundamental objection. The ascription of moral culpability does not settle the ascription of practical task-responsibility: in fact, the dependence goes in the other direction.

As Peter Geach has argued, an account of action must allow for action claims to have descriptive uses. In section 4, I have argued that responsibility-based accounts of action are not committed to being ascriptivist. In fact, on the account that will be developed the primary uses of action claims are descriptive. I will argue (Chapter VI) that to say that an action has been performed is to say that a practical task-responsibility has been discharged.

Three major tasks lie ahead. First, the notion of practical task-responsibility needs to be clarified in such a way as to prevent the account from being subject to the fundamental objection (Chapters III-V). Second, this involves giving an account of defeating conditions (Chapter V). Third, the notion of practical responsibility that I shall develop must then be shown to help in understanding the nature of action, in particular in rendering the distinction between actions and mere happenings (Chapter VI).

CHAPTER III.

PRACTICAL RESPONSIBILITY I: NORMATIVE EXPECTATIONS

In Chapter II, I have identified a basic objection to any responsibility-based account of action, the fundamental problem. In a nutshell, the concept of action appears to be prior to the concept of responsibility in the logical order of things. If so, then an account of action in terms of responsibility is impossible. I have also suggested that the challenge thus posed could be met with a concept of practical task-responsibility. The aim of this and the next two chapters is to develop such a concept. I will claim that a person is practically task-responsible for φing just in case it would be reasonable (in a special sense I will explain in Chapter V) to expect of her that she φ. Two major conceptual tasks lie ahead. First, the concept of expectation involved must be clarified. Second, the concept of reasonableness must be explained. These are the respective tasks of the present and the next two chapters. As we will see, both tasks are rather delicate. In both cases, we will see that the fundamental problem reappears at various junctures in the natural course of explanation of the concepts.

I begin the chapter by clarifying the distinction between normative and predictive (or descriptive) expectations (section 1). Sections 2-4 proceed to discuss the concept of normative expectations, since the concept of practical task-responsibility is characterized exclusively in terms of normative expectations. After some preliminary conceptual remarks in section 2, section 3 discusses the question what fulfills normative expectations. This is a delicate topic as this is the first place where the fundamental problem reappears. Section 4 briefly discusses the distinction between practical and moral expectations. Finally, in section 5 I will show how to neutralize the perspectival character of the notion of normative expectations.

I should note that the aim of this and the next two chapters is primarily to lay the groundwork for the discussion in Chapter VI. As such, the present considerations will not be dialectical. The aim of the chapters is not to defend the concept of practical task-responsibility but rather to lay down its meaning. Chapter VI will then use thus developed notion to show that it can do some useful philosophical work.

1. Normative vs. Descriptive (Predictive) Expectations

I expect of my mother-in-law that she treat me with respect and yet I expect that she will not. That no contradiction is involved is clear. Two different concepts of expectation are involved: the former expectation is normative, the latter predictive or descriptive.[1] Here is Patricia Greenspan's example. "If someone is known to be unusually lazy, say, or simply to dislike a certain kind of action — cleaning up, for instance — it might not be reason for us to 'expect' that person to perform it, in the sense of predicting that he *will;* but it might still be reasonable to think that the person *ought* to perform it — to expect it *of* him, in the sense of holding him to a standard which requires it."[2]

There are various ways of drawing the distinction between normative and descriptive expectations. As we saw, Greenspan characterizes the distinction in terms of the notion of prediction, on the one hand, and the notion of holding the agent to a demand, on the other. Wallace ties the notion of normative expectation with various reactive emotions we are inclined to feel when the expectation is frustrated (guilt,

---

[1] The distinction has a long standing in sociology, where normative expectations are taken to define social roles (see e.g. Erving Goffman, *Stigma. Notes on the Management of Spoiled Identity* [New York: Simon & Schuster, 1963]). It has progressively come to occupy a more important place in philosophical literature. For example, Patricia Greenspan has used the notion of reasonable normative expectations to define freedom ("Behavior Control and Freedom of Action," *Philosophical Review* 87, 1978, 225-240, and "Unfreedom and Responsibility," in (ed.) Ferdinand Schoeman, *Responsibility, Character, and the Emotions* [Cambridge: Cambridge University Press, 1987], pp. 63-80). A similar distinction (though labeled regularity- and rule-engendered expectation) is at work in Steven Lee's "Omissions," *Southern Journal of Philosophy* 16 (1978), 339-354. R.J. Wallace appeals to the notion of normative expectations in giving a compatibilist theory of moral responsibility (*Responsibility and the Moral Sentiments* [Cambridge, MA: Harvard University Press, 1994]).

However, the distinction between normative and descriptive expectations is not always recognized. Susan Sterrett (unpublished manuscript) shows the limitations of D. Lewis' account of convention due to his failure to take the distinction into account.

[2] P. Greenspan, "Unfreedom and Responsibility," *op. cit.*, p. 72.

resentment).[3] The distinction can be sharpened by appealing to the metaphor of direction-of-fit introduced by G.E.M. Anscombe.[4]

Predictive expectations (that *p*), like beliefs, have a mind-to-world fit: if it is the case that not-*p* the fault lies with the expectation. Normative expectations (that *p*), like intentions, desires, etc., have a world-to-mind fit: if it is the case that not-*p*, the fault is with the world which ought to be changed accordingly. More precisely, we can say that a person predictively expects that *p* when (among other things) he is disposed to dismiss the expectation as having been wrong if not-*p*. A person β expects (in the normative sense) of another person α that *p* when β is disposed to sanction α's failure to bring about *p*.

> β expects (in a normative sense) of α that *p* when β is disposed to impose a negative sanction on α if α fails to bring it about that *p* and a positive sanction if α does bring it about that *p*.[5]

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

Four points deserve a mention.

(i) Sanctions are to be understood very liberally. Negative sanctions in particular ought to include the reactive emotions Wallace speaks about. Being susceptible to feeling guilt, resentment or indignation are all forms of being disposed to sanction oneself or others in case of failure to fulfill the expectation.[6] But it includes sanctions of a lesser moral

---

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

[4] I discuss the distinction in Chapter I, p. 19.

[5] Section 3 clarifies what is meant by 'α brings it about that *p*' and 'α fails to bring it about that *p*'.

[6] Wallace discusses cases of irrational guilt, where one feels guilty without believing that one has frustrated any expectations one accepts. In explaining how this is possible Wallace suggests that we must distinguish between the ends that one values and the ends one is motivated to pursue. In our terms, the distinction is

magnitude. Feeling dissatisfied or disappointed by oneself or by another, criticizing oneself or others, etc. are all forms of negative sanctions. But there are also positive sanctions. Various forms of reward or feelings of satisfaction or accomplishment are forms of positive sanction.

(ii) Robert Brandom[7] argues extensively that to understand normative attitudes in terms of sanctions, one must not attempt to reduce normative attitudes to people's (or communities') behavioral dispositions to sanction. Rather, any understanding of normative attitudes must appeal to an already normative notion of sanction. Indeed, it must be the case not only that a person *does* or *tends* to sanction non-conforming behavior but that the person *ought to* sanction it.

The above characterization of what it means for one person to expect something of another does not attempt a reduction of the normative attitude of expectation to a mere disposition. When β expects of α that *p*, β is required to be disposed to negatively sanction α in very specific circumstances, viz. when α fails to bring it about that *p*.[8] In other words, β is required to be *correctly* disposed to negatively sanction α. Likewise, β is required to be *appropriately* disposed to place a positive sanction on α, when α does bring it about that *p*.

(iii) It may be worthwhile pointing out that it is not uncharacteristic for philosophers writing on responsibility to focus on negative sanctions. While the availability of the negative side is crucial for an account of action, for it will ultimately allow us to capture negative actions, it is also crucial that the positive side not be left out, for if it were we could not account for positive actions. If there were a reason in principle why the concept of responsibility had to be geared toward the negative side this would constitute a

---

one between expecting something of oneself and believing that such an expectation is reasonable. Usually these two attitudes go hand in hand, but it is possible for one to expect of oneself what one believes not to be reasonable, in which case one feels guilty (because one is disposed to sanction oneself) but irrationally or unreasonably because one believes that the expectation is unreasonable.

[7] *Making It Explicit* [Cambridge: Harvard University Press, 1994], pp. 34-46.

[8] The characterization would not be immune to a charge if the condition of negative sanction were "if β believes that α fails to bring it about that *p*."

prima facie reason for finding responsibility-based approaches to action suspect. Wallace comments on just this point:

> It is striking... that the responses of blame and sanction are negative and punitive in character. Of course, there are positive responses to which holding people responsible occasionally disposes us as well: we praise people, for instance, who are outstandingly good and virtuous. But praise does not seem to have the central, defining role that blame and moral sanction occupy in our practice of assigning moral responsibility. [continued in the footnote:] This is not to rule out the very possibility of a system of social reactions organized primarily around the positive responses of praise and reward rather than blame and sanction; such a system might even be superior to our present practice, in some respects.[9]

Two points are clear from this passage. Wallace rightly or wrongly takes the focus on the negative side to be, first, a characteristic of moral responsibility and, second, of *our practices* of holding people to be morally responsible. This suggests that there should be no problem in extending his characterization of the basic concept of holding someone (practically) responsible to cover the positive cases.

One may be tempted to speculate that the fact that philosophers of responsibility tend to focus on negative cases has a not accidental correlate in the fact that philosophers of action tend to focus on positive cases. One could imagine that the concepts of action and responsibility could be in a kind of equilibrium: the concept of action covering cases of positive as well as of negative actions, and the concept of responsibility (primarily in the sense of answerability) correspondingly attaching to them in the right circumstances. As it happens, the concept of action is focused on the positive cases, while the concept of responsibility tends to be focused on the negative cases. In either case, the focus does not appear to have any solid justification. In what follows, I will be trying to treat both concepts as having an equal bearing on both sides: the positive and the negative.

(iv) Another point about the characterization of normative expectations deserves a mention. We have seen in Chapter I how M. Smith used the idea of direction-of-fit to define desires. It will be instructive to consider the difference between these characterizations. Smith understands a desire that *p* as "that state of a subject that grounds all sorts of his dispositions: like the disposition to φ in conditions *C*, the

---

[9] R.J. Wallace, *Responsibility and the Moral Sentiments, op. cit.*, p. 61.

disposition to φ in conditions *C'*, and so on (where, in order for conditions *C* and *C'* to obtain, the subject must have, *inter alia*, certain beliefs)."[10] Thus understood desires are conceived to be intrinsically motivating — they dispose the agent to the desired action. By contrast, expectations are not seen to be intrinsically motivating; they dispose the agent to adopt sanctioning attitudes whether to oneself or to others. That this is not incompatible with normative expectations playing a motivating role will become clear in Chapter VII. But this role is not built in, as it should not be, into the very concept of an expectation.

2. Normative Expectations

In what follows, I will be concerned exclusively with normative expectations. The term 'expect' is henceforth reserved for normative expectations unless it is explicitly noted otherwise. I will assume that it is possible to formulate all normative expectations in the following canonical form:

β expects of α that α φ,

or: β expects of α that α bring it about that *p*,

where β is the expector, α is the agent, 'φ' is an action-verb, '*p*' is a sentence.

Most normative expectations lend themselves to this canonization very well. Thus: Jane expects of Jim that he move his car so that she may drive out of the garage. The teacher expects of his student that she take part in the school play. Jennifer expects of herself that she become another Maria Callas. Other normative expectations may not be explicitly stated in this form, but they can be easily recast. In the simplest case, 'Sam expects Mary to be here in five minutes' can be reformulated as 'Sam expects of Mary that she come here in five minutes'. Likewise, 'The admiralty expects of the captain that the ship arrive safely' can be reformulated as 'The admiralty expects of the captain that he make sure that the ship arrives safely'.

---

[10] Michael Smith, "The Humean Theory of Motivation," *Mind* 96 (1987), p. 52.

We need to say a little bit about the actors involved. Then, we will consider the that-clause in some detail, for it is there that the fundamental problem of responsibility-based approaches, discussed in Chapter II, resurfaces. Three preliminary points first.

(i) I will primarily speak of individual people as holding each other to expectations. However, there is no barrier to thinking that other *social agents* can hold each other to expectations. One state may expect of another state that it not be invaded. A firm may expect loyalty of its employees. A group of people may expect another group to play fair. And so on.

(ii) It is also possible to expect something of oneself, in fact many expectations are *reflexive* or *self-directed.* The concept of a reflexive expectation is indeed a very close kin to the concept of intention. Usually, when I have a prior intention to do something I expect of myself that I do it.[11]

(iii) The that-clause in the expectation specifies the description under which the act is expected of α by β. Its occurrence is thus not transparent. What is expected is never a concrete particular performance but rather a type. So, it does not follow that if Jane expects Jim to greet her friend by waving hello, and if his waving hello happens to be identical to his voting for a challenger, then she expects him to vote for the challenger.

3. Fulfilling Normative Expectations: Actions and Performances

Normative expectations can be fulfilled, frustrated, or neither fulfilled nor frustrated by events. Suppose that Mary expects of John that he bring her his homework by 5pm. The event of John handing over the homework to Mary at 3pm *fulfills* her expectation. So would the event of John's sending his homework by mail if it gets there by 5pm.[12] However, if John sits in the library at 5pm (perhaps intending to bring the

---

[11] This is not meant to imply that the concepts are identical.

[12] One might read the original expectation in a more strict way. One might insist that the only way for the expectation to be fulfilled would be if John physically brings the homework to Mary. I shall not worry about this issue at all. I doubt that there is a way of settling the question by looking at the language used. At the same time, nothing in our account will depend on the issue being resolved in one way or another. It will suffice for my purposes to assume that the content of the expectation is settled by its fulfillment and frustration conditions.

homework at 6pm), Mary's expectation would be *frustrated.* Finally, there are events that will *neither fulfill nor frustrate* the expectation, e.g.: Vesuvius exploding a century ago, the Congress passing a bill at 5pm that day, John procrastinating on the steps of the library at 4pm, and so on.

We need to consider the reappearance of the fundamental problem alluded to at the beginning. Here is the problem in a nutshell. The that-clause appears to be an agentive statement. If so, then in order to make sense of normative expectations, we would have to have a firm grasp on the notion of action. But this seems to render the very project at hand circular, for I have recommended that the concept of action is to be illuminated by the concept of normative expectation. Once again, the order of the concepts of action and normative expectation appears to be the reverse of that needed by the project.

The first step toward a solution consists in noting that the occurrence of 'α φ' in the that-clause does not yet prejudge the issue. What matters is how we think about the performances that fulfill and frustrate the expectations. There are two ways of thinking about the fulfillment and frustration conditions. Either, one may think that expectations are fulfilled (frustrated) only by actions — we may say that the expectations are *agentively fulfilled* (frustrated). Or, one may think that expectations are fulfilled (frustrated) by a more liberal class of performances which includes not only actions but mere happenings. In this case, we may say that the expectations are only *prima facie fulfilled* (frustrated).

Consider an example. A guest at a party becomes annoyed by the hostess's bragging about her authentic Persian rug not just a little too much. He thinks to himself that she would probably be annoyed if something damaged it and immediately thinks of his cup of tea. This is how he comes to expect himself to knock the cup of tea when he reaches forward for some sugar. Just as he is about to do that a muscle spasm shakes his arm thus making it bend in such a way that he knocks the cup of tea from the table where it spills over the bragged about rug. The guest's expectation of himself is prima facie fulfilled, but it is not agentively fulfilled.

Given our task to construct an account of action in terms of fulfillment and frustration of normative expectations, the project would indeed be circular if we took

normative expectations to be fulfilled only by actions (i.e. to be fulfilled agentively). We are accordingly committed to taking expectations to be prima facie fulfilled (frustrated). In other words, normative expectations must be construed as being fulfilled by *performances* (which includes actions and mere happenings). The distinction of agentive fulfillment and frustration conditions will then be made in terms of the standard of reasonableness (in a special sense to be explained in Chapters IV and V).

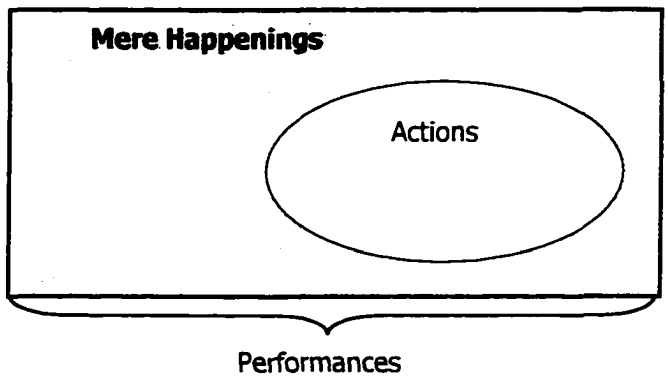


Figure 1. The relation between the class of performances, actions and mere happenings.

I should note a certain delicacy in trying to characterize the category of performances. I do not want to offer any theoretical characterization of what a performance is beyond saying that it includes the category of actions and of mere happenings. This will mean that bodily movements count as performances. However, saying anything beyond that is controversial for it hinges on highly controversial questions in the ontology of action,[13] which I will not address in the dissertation. I will try to circumvent the issue by focussing the category of performances on bodily

[13] Some of the important voices in the debate include: G.E.M. Anscombe, *Intention* (Ithaca: Cornell University Press, 1957); "Under a Description," *Nous* 13 (1979), 219-233; Annette C. Baier, "Ways and Means," *Canadian Journal of Philosophy* 1 (1972), 275-293; Donald Davidson, "Agency," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 43-61; Lawrence H. Davis, *Theory of Action* (Englewood Cliffs: Prentice-Hall, 1979); Carl Ginet, *On Action* (Cambridge: Cambridge University Press, 1990); Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970); Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980).; Hugh McCann, "Volition and Basic Action,"

movements. Most theorists struggling with the question of the proper way of carving up ontological space for actions agree at least on the fact that some bodily movements (raising an arm, reaching out, walking) count as actions.[14] I should point out, however, that this focus is contrary to the spirit of the account here developed. For there is nothing in the account that would dictate the thought that actions must be thought of paradigmatically in terms of bodily movements. Even if this were true for individual actions (which I do not believe), it would hardly be so for collective, institutional, or more generally, social actions. However, it is less clear that we can give as intuitively appealing a characterization of the category of performances for those kinds of actions without getting involved in the very complex issues surrounding the proper way of constructing the ontology for a theory of action. I shall therefore simplify the account here by restricting the category of performances to the category of bodily movements.

4. Moral vs. Practical Normative Expectations

One of the tasks of a responsibility-based account of action is to make sure that the concept of responsibility is broader than the more familiar concepts of moral or legal responsibility. The concept of practical responsibility can be delineated by means of the concept of practical normative expectations.

It may be helpful to follow R.J. Wallace's attempt to develop the concept of *moral* responsibility in terms of moral normative expectations. Wallace finds the concept of responsibility based on normative expectations too inclusive for his purposes. He consequently restricts the expectations relevant to his task to ones that have a specifically moral justification, i.e. to moral expectations. Since our task is to offer a responsibility-based account of action not of moral action, we ought to use a more inclusive concept of (practical) responsibility. It is natural for us simply not to restrict the class of normative expectations to just those that have a moral justification and instead to include all

---

Philosophical Review 83 (1974), 451-473; Judith Jarvis Thomson, "The Time of a Killing," *Journal of Philosophy* 68 (1971), 115-132.

[14] This is a somewhat deceitful depiction because there is a considerable difference in the way in which bodily movements are conceived. Jennifer Hornsby (*Actions, op. cit.*) in particular offers a very esoteric interpretation of the bodily movements that are actions.

normative expectations. They will have to be subjected to normative appraisal (to exclude arbitrary expectations, for instance) but such normative appraisal will show them to be inappropriate or appropriate in particular situations, it will not show them not to be practical.

In what follows, any normative expectation is considered to be a *practical* normative expectation with one exception. Those expectations that have either an empty fulfillment or an empty frustration set are not considered practical. Any expectation of the form 'β expects of α that α bring it about that p', where p is either logically or physically non-contingent, is not practical. If one were to expect of someone that 2+2=4, such an expectation could not be frustrated; similarly, if one were to expect of someone that 2+2=5, such an expectation could not be fulfilled. The intuitive reason behind the refusal to classify such expectations as practical should be clear: there is nothing anyone can *do* to make it the case that what is expected is the case or is not the case.

Expectations with a non-empty fulfillment and a non-empty frustration set are thus considered to be practical expectations. This is to say something about their content, rather than about their propriety. Practical expectations may be appropriate (reasonable) or inappropriate (unreasonable). In the next two chapters, we will see the complex conditions that are responsible for expectations being reasonable in various situations.

5. 'It is (would be) reasonable to expect of α that α φ'

So far we have spoken of one person expecting something of another person. For reasons that will become clearer, we need to introduce another concept that does not explicitly mention the person who holds another to the expectation. Since it is not clear that there is a settled intuitive meaning of the phrase "*it would be reasonable* to expect something of a person," I will distinguish two readings and use one of them consistently throughout the dissertation. I should emphasize that my aim in this section is to understand what it means to say that *it would be reasonable* to expect something of α, not to understand what the reasonableness of an expectation consists in. For the rest of this section, I shall assume that the notion expressed by 'β's expectation of α that α φ is reasonable' is clear.

It may be helpful to begin with a slightly simpler notion. Sometimes, we may say that it would be reasonable *for a person* to expect something of another. For example, it may be reasonable for me to expect of my husband that he does dishes from time to time. The first point to note is that to say so is not yet to say that I in fact *do* expect him to do the dishes. The claim that is made is rather conditional: if I were to expect this of him, my expectation would be reasonable. Thus, more generally,

'*it would be reasonable for* β to expect of α that α φ' means:

> 'if β were to expect of α that α φ, such an expectation of β's would be reasonable'.

Occasionally, we might want to say not only that it *would be* reasonable for β to expect something of α but that it *is* reasonable for β to expect it of α. In some cases, to say that it is reasonable for a person to expect something of another presupposes that that person does hold the other to the expectation. The claim then merely assesses the expectation as reasonable. Thus, knowing that a coach expects of his athletes that they abstain from drinking, we may judge that it is reasonable for him to expect it of them. In other cases, one may say that it *is* reasonable for a person to expect something of another rather than saying that it *would be* reasonable, in order to emphasize that one believes the person *ought* to hold the other to the expectation. So, one might say that it not only would be reasonable for me to expect of my husband to help out with the dishes, but that it *is* reasonable for me to hold him to the expectation, i.e. that I ought to expect it of him. I will use the phrase thus:

'*it is reasonable for* β to expect of α that α φ' means:

> 'if β were to expect of α that α φ, such an expectation of β's would be reasonable, *and* β either holds or ought to hold α to the expectation that α φ'.[15]

---

[15] For our purposes, the difference between the claim that it is reasonable to expect something of someone and the claim that it would be reasonable to expect it of her is negligible. It will have no bearing on any substantive commitments.

Given that we understand what it means to say that *it would be* reasonable *for a person* to expect something of another, we can ask what it means to say that *it would be* reasonable to expect something of another. There are two interpretations one could give. On one reading, to say that it would be reasonable to expect of α that α φ is to say that it would be reasonable *for everybody* to expect of α that α φ. (In other words, for any person ξ, were ξ to expect of α that α φ, ξ's expectation of α that α φ would be reasonable). This is probably what we mean when we say that it is reasonable to expect of a person that she not kill another. Such an expectation would be reasonable no matter who expected this of the person. We might say that it would be *universally* reasonable to expect of α that α φ.

But there is a weaker reading, according to which to say that it would be reasonable to expect of α that α φ is to say that it would be reasonable *for someone* to expect of α that α φ (or: for some person ξ, were ξ to expect of α that α φ, ξ's expectation of α that α φ would be reasonable). To see this as a plausible interpretation, imagine that α has a certain position in a hierarchical organization. Let us suppose that α is a computer programmer and it is part of his job to produce a certain amount of code within a specified amount of time. When we say that it is part of his job (which he accepted of his own will, etc.), we believe that ceteris paribus it is reasonable to expect of him, among other things, to produce this amount of code in the specified amount of time. To believe that *it would be reasonable* to expect this of him is now no longer to believe that it would be reasonable for *everybody* to expect it of him. Rather, it is to believe that it would be reasonable for some person (e.g. his supervisor, his firm, his coworkers) to expect of him that he produce the code.

Henceforth:

> '*it would be reasonable to expect of α that α φ*' means 'For some person ξ, were ξ to expect of α that α φ, ξ's expectation of α that α φ would be reasonable'

while

> '*it would be universally reasonable to expect of* α *that* α φ' means
> 'For every person ξ, were ξ to expect of α that α φ, ξ's expectation of α that α φ would be reasonable'.

The concept of it being reasonable to expect something of an agent will be of central importance for the account of action I shall offer. It has the advantage that it is applicable even in cases where no-one actually holds the agent to the expectation.

Let us note that under the second reading, the concept of it being reasonable to expect something of an agent leaves the possibility of conflicting expectations open. It may be reasonable to expect of α that α φ (because it is reasonable for β to expect of α that α φ) but it may also be reasonable to expect of α that α not-φ (because it is reasonable for γ to expect of α that α not-φ). It may be that one of the claims (β's or γ's) is actually stronger, but it may also be that there is no way of deciding on their strength. Such a possibility ought not to be excluded by fiat. It would be decided by fiat if the phrase were to be used in its universal sense.

• • •

Before going on to discuss the difficult topic of what makes expectations reasonable, it may be worthwhile assembling all the preliminary ingredients into an account of practical task-responsibility, and summarizing how some of the objections discussed in Chapter II are met.

> α is practically (task-)responsible for φing if and only if it would be reasonable to hold α to the practical normative expectation that α φ.

Three points ought to be emphasized. First, the concept at stake is one of *practical* (rather than legal or moral) responsibility. As explained in section 4, all non-empty expectations are considered to be practical. Second, it is a *forward-looking*, not a backward-looking, concept of responsibility. To expect something of a person is to hold her responsible (in a forward-looking sense) for the carrying out of a certain task, i.e. to hold her task-responsible.

These two features are important in allaying the fundamental problem. We will remember that the doubts arise in view of the fact that it is natural to think that responsibility ascriptions presuppose the knowledge whether an action has been

performed. It thus seems impossible to try to construe a concept of action in terms of the (apparently later, in the logical order of things) concept of responsibility. The developed concept of practical task-responsibility resolves the problem in two ways. First, the concept of practical responsibility is broader than the concepts of moral or legal responsibility. It is quite intelligible to claim that the concept of action logically precedes the concepts of moral and legal responsibility, while it depends on the concept of practical responsibility. Second, the concept of practical task-responsibility, unlike the concepts of moral or legal responsibility that give rise to the objection, is a forward-looking concept. We seem to be compelled to think that backward-looking concepts of responsibility presuppose the concept of action, for it is most natural for us to think that we are morally and legally responsible for our actions. It is not equally compelling to think that the forward-looking concept of task-responsibility presupposes the concept of action. As we saw (section 3), a case could be made that it does after all. One could argue that what fulfills normative expectations are actions, in which case the fundamental problem reappears. I have, however, argued that we can also understand normative expectations as being fulfilled by performances (comprising actions as well as non-actions, mere happenings). We will see in the next chapters that more conceptual work will need to be done before the fundamental problem is held at bay.

The third, and final, point about the characterization of practical task-responsibility is that it involves an appeal to the standard of *reasonableness*. This is intended to eliminate arbitrary or otherwise inappropriate expectations of a person as counting toward her being practically responsible for something. Chapters IV-V will be devoted to clarifying what it means to say that normative expectations are reasonable.

CHAPTER IV.

PRACTICAL RESPONSIBILITY II: TWO CONCEPTS OF REASONABLENESS

In Chapter III, we have seen how to develop a concept of practical task-responsibility in terms of normative expectations. It is now necessary to take up the last most difficult task of developing the concept of reasonableness of normative expectations.

At first sight, the notion may appear to be hopelessly riddled with difficulties. For one, it seems to be thoroughly expector-relative. What may be reasonable to you may not be reasonable to me. What is reasonable to me once I have corrected my false beliefs would not have been reasonable to me before. Although I have insisted in the last chapter that the concept that will matter for us is not what it is reasonable *for a particular person* to expect of another but what *it is* reasonable to expect of another, one might object that this move merely covers up a deep problem.

In section 1, I begin to address the problem by distinguishing two concepts of reasonableness: agent reasonableness (reasonableness$_A$) and normative reasonableness (reasonableness$_N$). In section 2, I show that the concept of reasonableness$_A$ can be construed in such a way as to avoid the difficulty. (This is an important result in view of the fact that only reasonableness$_A$ will be fundamentally relevant to the account of action to be given in Chapter VI.) Section 3.A answers the question whether reasonable expectations can stand in conflict: could it be that it is both reasonable and unreasonable to expect of an agent that she perform an action? Section 3.B considers whether contrary expectations can be both reasonable: could it be that it is reasonable to expect of an agent that she φ and to expect of her that she not-φ?

1. Two Concepts of Reasonableness

Normative expectations involve making demands, in the paradigmatic cases, on others. As such, the immediate concern that arises is that such demands be legitimate, appropriate or reasonable. There are at least two kinds of ways in which normative expectations may be inappropriate or unreasonable. In fact, we may speak of two senses of reasonableness.[1]

One reason why an expectation of a person may be unreasonable is, as we intuitively say, that it is not "within her power"[2] to do what she is expected to do. For instance, it would be unreasonable to expect of an athlete who broke a leg that she take part in a race, of a blind person that he drive a car, or of a newly arrived foreigner that he speak like a native. In all such cases, we think that the agent "lacks the basic ability to do what we are demanding,"[3] and thus we believe that it would be unreasonable to hold the agent to the expectation in such conditions.

Another reason why an expectation may be unreasonable is of a different nature. It may be that the person has the general power to do what we expect of her, but it may be nonetheless inappropriate for us to expect it of her. Let us suppose that you have a relatively ordinary relationship with your neighbors. You are polite to one another, occasionally help one another out in neighborly matters. But there are (many) expectations that it is simply inappropriate for you to hold them to, and not because it is not "within their power" to fulfill them. For instance, it would be inappropriate for you to expect them to regularly mow your lawn, to do your shopping, etc.

These two kinds of cases exemplify two different, though equally fundamental, concerns with the reasonableness of normative expectations. For want of better terminology, I shall speak of *reasonableness*$_A$ (agent-reasonableness) to capture the first

---

[1] I do not have a conclusive way of showing that two distinct concepts are involved. So I do not want to deny that there may be a way of elucidating one unified concept of reasonableness. It is fruitful for my purposes to treat them as distinct concepts, and I produce some further evidence to this effect in the course of the section.

[2] It is not until Chapter V that we will have a better understanding of what it means to say that something is "within the agent's power" to do. In order to signal that this notion functions as a metaphor and a theoretical place-holder, I consistently embrace it in scare-quotes.

sort of case, and of *reasonableness*$_N$ (specifically normative-reasonableness) in the second kind of case.

We have already seen that it is possible for an expectation to be reasonable$_A$ but unreasonable$_N$. Your expectation of your neighbors that they do your shopping would be reasonable$_A$ (because it is "within their power" to do so) but it would be highly unreasonable$_N$ for you to expect it of them. It is also possible for an expectation to be reasonable$_N$ but unreasonable$_A$. A teacher may reasonably$_N$ expect of his student that she turn the assigned paper on time, but the expectation may be unreasonable$_A$ in view of the fact that the student has been taken to the hospital.

I will not offer any account of the concept of reasonableness$_N$. In section B, I will attempt to clarify this concept a little bit, but the remarks are far from being either complete or entirely satisfactory. In the end, I will simply have to appeal to the reader's better judgment concerning particular cases. This will not obfuscate the account of action to be given, for the concept of reasonableness$_N$, as we shall see, plays a more modest role than that played by the concept of reasonableness$_A$. I will argue in Chapter V that the concept of reasonable$_A$ normative expectations is sufficient to decide whether a performance is an action or not. Throughout the discussion, I shall emphasize certain reasons that give additional support to the supposition that reasonableness$_N$ and reasonableness$_A$ are distinct concepts.

A. Reasonableness$_A$

There are two kinds of conditions that comprise our understanding of reasonableness$_A$ of expectations. First, there are conditions that can be classified under general competence. Usually, an agent's competence increases with age until adulthood and then diminishes in old age. A generally competent agent is attentive, conscious, intelligent, motorically responsive, possesses certain general skills, etc. Other individuals may lack such basic skills and be considered more or less competent; accordingly certain normative expectations of them will be unreasonable$_A$. Such individuals will include

---

[3] R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994), p. 161.

babies, infants, people with some forms of handicap (mental handicap, blindness), etc. Second, there are conditions that occur against the background of general competence locally, as it were, making the performance of some type of action in those circumstances not within "the agent's power." These are defeating conditions. They include various kinds of physical injury (illness, breaking a leg), physical force to which the agent is subject (being pushed by the wind, being pushed by somebody else).

There is a range of performances considered part of everyone's general competence. Among them: walking, sweeping, throwing, catching, running, counting, remembering, etc. If an agent is not competent in some of these ways, he acquires a special treatment (is qualified as a minor or as incapacitated in various ways). But there are expectations which, while they may not be reasonable$_A$ generally, may be appropriate in view of a person's special ability. It may not be reasonable$_A$ to expect of everyone to do the books with the skill of an accountant, but it is reasonable$_A$ to expect it of accountants because of their special skills. It may not be reasonable$_A$ to expect of just everyone to do a pirouette, but it may be reasonable$_A$ to expect it of a skilled skater.

It is important to point out that all normative expectations, which include reflexive expectations (directed at oneself), are subject to such an appraisal. It is equally unreasonable$_A$ to expect of a person who suffers regular muscle spasms that he become a surgeon as it would be to expect this of oneself if one suffered from such a condition. The concept of reasonableness$_A$ is also indifferent with respect to who expects something of the agent. If it is unreasonable$_A$ for John expect of Mary that she jump to the moon then it is unreasonable$_A$ for Lori to expect it of Mary.

It should be pointed out that although the concept of reasonableness$_A$ is related to the metaphor of a performance being "within the agent's power," there are important cases, where it is reasonable$_A$ to expect something of an agent despite the fact that the agent cannot do what is expected of him. Save for very special circumstances (which include illness, e.g.), when a director of a firm is expected to be at a meeting at 9am (provided he knew about the meeting, etc.), this expectation is reasonable$_A$ and continues

to be reasonable$_A$ even if the agent is still asleep at 9am.[4] In the next Chapter, we shall see how one can accommodate both the intuition that such an expectation is reasonable$_A$ and the intuition that the case is a rather special one.

B. Reasonableness$_N$ (Legitimacy) of Expectations

Normative expectations are subject to two kinds of appraisal. The need for one kind of appraisal (reasonableness$_A$) arises in view of a concern with "the agent's very power" to do what is expected of him. The need for the second kind of appraisal (reasonableness$_N$) arises in view of the interpersonal nature of many expectations, and hence the need to justify the expectations in terms of reasons.

This last point is best seen by contrasting self-directed normative expectations with expectations directed at other people. It seems intuitive to think that as long as what I expect of myself is "within my power" to do, i.e. as long as what I expect of myself is reasonable$_A$, there is no limit to what I can legitimately (reasonably$_N$) expect of myself.

> There are no practical expectations it would be unreasonable$_N$ for an agent to hold herself to.

I can expect whatever I want from myself. None of such reasonable$_A$ expectations will be unreasonable$_N$, though the expectations may vary in the degree to which they are reasonable$_N$. I can expect myself to fly to the Bahamas next month, to quit my job, to change my identity, to bake a cake for my neighbor, to write a novel. Were I to hold others to just such expectations, however, the matter would no longer be so clear. I can legitimately place demands on myself, any demands provided only they are not criticizable on the grounds of unreasonableness$_A$. But when it comes to my placing demands on others, or to others' placing demands on me, the situation changes dramatically.

The judgment whether it is reasonable$_N$ (legitimate) to expect something of another person will depend on achieving a delicate balance between the claims of the

---

[4] It is important to be careful here. The point holds for normative not predictive expectations. The predictive expectation that the director will come to the meeting at 9am given that he is asleep at that time is surely false ("unreasonable"); but this is not to imply that the normative expectation is unreasonable$_A$.

person who expects something, the person of whom something is expected, other people involved, as well as the weight of the expectation and the difficulty of fulfilling it.[5] It involves striking a balance between reasons. Let us consider some examples. When a person falls ill on a street, even among perfect strangers, it is reasonable$_N$ for her to expect of others that they come to her help. This is a case where the judgment of reasonableness$_N$ is dominated by the concern with the person who is in need of help and expects it from others, as well as by the weight of the expectation — it is possible that her well-being or even life is at stake. Suppose that an employee who is expected to deliver a presentation at the firm's annual meeting is taking his spouse to the hospital. Prima facie, we will judge the firm's expectation of the employee no longer reasonable$_N$ in view of the circumstances. Here too the weight of the expectation balances the employee's concern with his wife's health. Suppose that the person whose wife is taken ill is not an employee of a firm expected to deliver a presentation, but the president of a nation expected to make a decision on which the nation's survival may depend. In such a case, it seems that even an extreme state of his wife's health would not defeat the reasonableness$_N$ of the expectation to keep the professional appointment.[6] In general, the greater the importance of the object of an expectation, the more justified we think ourselves in placing greater demands on others, the more reasonable$_N$ the expectation. On the other hand, the greater the difficulty of fulfilling an expectation, the less justified do we think ourselves in placing a demand on another, although we might feel the more justified in holding ourselves to such an expectation.

---

[5] "An agent's freedom, and his responsibility 'before-the-fact' will ... depend on overlapping but nonidentical normative considerations. Both will vary with 'the stakes', conceived as the importance of an object of 'reasonable expectation', weighted against the difficulty of fulfilling it. However, the notion of responsibility apparently takes awareness of the reasons for action as a further object of reasonable expectation, with a further weighting — of the importance and the difficulty of *discerning* the reasons — imposed only hypothetically on freedom." (Patricia Greenspan, "Unfreedom and Responsibility," in (ed.) Ferdinand Schoeman, *Responsibility, Character, and the Emotions* [Cambridge: Cambridge University Press, 1987], p. 76.) Greenspan's aim is to capture the notion of unfreedom and so I believe that she focuses primarily on the notion of reasonableness$_A$, though many of her comments speak to the notion of reasonableness$_N$.

[6] Note that this does not necessarily contradict the suggestion that it is also reasonable$_N$ to expect him to be at the hospital. His self-expectation to be with his wife might still be reasonable$_N$. This would be a case (in this instance) of moral conflict.

The concept of reasonableness$_N$ (unlike that of reasonableness$_A$) is related to reasons. We could perhaps also draw a distinction similar to the distinction between prima facie and all-out reasons. We might say that it is prima facie reasonable$_N$ for Jenny to expect of herself that she go to the movie, as long as she has some (prima facie) reasons to go to the movie. It is all-out reasonable$_N$ for Jenny to expect of herself that she go to the movie if the balance of all considerations suggests that she should go to the movie.

Unlike the concept of reasonableness$_A$, reasonableness$_N$ does admit of an intermediate category. There may be performances that it is neither reasonable$_N$ nor unreasonable$_N$ to expect of the agent. When the agent actually acts in this way, we say that the agent acts spontaneously for no reason. For example, it is reasonable$_N$ for me to expect of my mailman that he deliver the post every day; it is unreasonable$_N$ for me to expect of my neighbor that she do my shopping; but it is neither reasonable$_N$ nor unreasonable$_N$ for me to expect of myself that I walk to and fro (when I have no reason for it).

As suggested earlier, it seems in general true that no expectations of oneself are unreasonable$_N$, so that any expectation of oneself may be either reasonable$_N$ or neither reasonable$_N$ nor unreasonable$_N$. Some of my expectations may be "unreasonable" in the sense that I may expect of myself what is beyond my power to do. But such expectations are unreasonable$_A$ not unreasonable$_N$ (illegitimate). In general, we leave it to the agent's discretion to expect of herself whatever her fantasy dictates. Not so for expectations directed at others. Because an expectation involves placing a demand on another person, such a demand must be justified and weighed against various kinds of considerations. Expectations toward others may be reasonable$_N$ and unreasonable$_N$. Can they be neither reasonable$_N$ nor unreasonable$_N$? Perhaps this would be true for a case where I expect of you what you can easily do (perhaps more easily than I), where I have no particular reason for expecting it of you and you have no particular reason either to do it or not to do it. Let us suppose that we sit together in a garden under a tree on a hot day, conversing amicably, and then I notice a daisy growing next to your foot. "Give it to me," I say, expressing my expectation of you that you pick it and forward it to me. Is my

expectation of you reasonable$_N$? Not in any clear sense, it is not really justified by any reasons. But there are no particular reasons to suppose that it is unreasonable$_N$ either.

It follows from the above characterizations that

> *it* is *never* unreasonable$_N$ to expect of α that she φ as long as it is reasonable$_A$ to expect of her that she φ (i.e. as long as it is within "her power" to φ)

This claim follows from two claims made above. First, we have suggested that the phrase 'it is reasonable to expect of α that α φ' be understood in terms of there being someone such that if she expected of α that α φ her expectation would be reasonable. In view of the fact that it is never unreasonable$_N$ for α to expect of herself that she φ, there will always be someone (viz. α herself) whose expectation of herself (provided that it is reasonable$_A$) will not be unreasonable$_N$. This means that *it* is never unreasonable$_N$ to expect of α that she φ, although it may well be unreasonable$_N$ for *somebody else* to expect of her that she φ. At the same time, in view of the fact that not all of the (reasonable$_A$) expectations that the agent has of herself are guaranteed to be reasonable$_N$ (only those that the agent has reasons for):

> It is not always reasonable$_N$ to expect of α that she φ even if it is reasonable$_A$ to expect of her that she φ.

The fact that these two platitudes follow from our considerations constitutes additional support for our analytic decisions and intuitions.

2. Reasonableness as an External Standard

The standard of reasonableness could be construed in external or internal terms. The distinction can be modeled on the distinction between an external and an internal reading of the notion of a reason.[7] Consider an example. An agent wants some gin and tonic. What is in her glass is in fact petrol but she believes it is gin. Does she have a

---

[7] Bernard Williams, "Internal and External Reasons," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), pp. 101-113. Williams argues that only internal reasons can motivate the agent to act. This is not an issue I am concerned with here.

reason to add tonic to her glass and drink it? The answer depends on whether we give an internal or an external reading to the concept of reason. On the external reading, she does not have a reason to drink what is in her glass — after all it is petrol. On the internal reading, she does have a reason to drink what is in her glass — she does not know it is petrol, she thinks it is gin.

For us the central question is whether it is reasonable for her to expect of herself that she pour tonic into the glass and drink it. To answer in the positive is to take it that the concept of reasonableness is internal, that it is responsive to internal reasons accessible to the agent. To answer in the negative is to take it that the concept of reasonableness is external, it is responsive to normative reasons not necessarily accessible to the agent at the time.

I will understand the concept of reasonableness in the *external* sense. If there is a disparity between the internal and the external concept, I will say that a person *believes that an expectation is reasonable* while *in fact it is unreasonable*.[8]

The choice to use the external reading is dictated by the purpose for which the concept is employed.[9] The notion of reasonableness (in particular reasonableness_A) is to be used in elucidating the nature of action. The adoption of an internal reading of the concept of reasonableness_A would lead to a subjective (expector-relative) reading of the concept of action. Whether an agent has performed an action in this sense would depend on whether somebody else (β) had internal reasons to hold the agent practically

---

[8] Unlike Williams, I am not concerned to investigate the question whether we can act on external reasons. And it is there that the question becomes controversial. See for example: Rachel Cohon, "Are External Reasons Impossible?," *Ethics* 96 (1986), 545-556; "Hume and Humeanism in Ethics," *Pacific Philosophical Quarterly* 69 (1988), 99-116; Brad Hooker, "Williams' Argument Against External Reasons," *Analysis* 47 (1987), 42-44; John McDowell, "Might There Be External Reasons?," in (eds.) J.E.J. Altham, Ross Harrison, *World, Mind, and Ethics* (Cambridge: Cambridge University Press, 1995), pp. 68-85; Alfred Mele, "Motivational Internalism: The Powers and Limits of Practical Reasoning," *Philosophia* 19 (1989), 417-436; Michael Smith, "The Humean Theory of Motivation," *Mind* 96 (1987), 36-61; "Internal Reasons," *Philosophy and Phenomenological Research* 55 (1995), 109-131.

[9] I should point out that the use of the idiom 'it is reasonable to expect of α that α φ' is justified only on the external reading of reasonableness. I have declared that the idiom is a shorthand for 'it is reasonable for some ξ to expect of α that α φ'. It follows that if it is reasonable for β to expect of α that α φ then it is reasonable to expect of α that α φ. If reasonableness were understood as an internal standard, this inference would be faulty. From the fact that β has internal reasons to expect of α that α φ, it does not follow that it is reasonable to hold α to such an expectation; perhaps β's reasons are completely esoteric.

responsible (whether β had internal reasons to believe that it was "within the agent's power" to fulfill an expectation). It is not immediately clear that the employment of such a concept would yield our concept of action (understood as: the agent actually doing something). It is more clear (providing our arguments in Chapter VI are sound) that the employment of the internal standard of reasonableness$_A$ would yield a concept of it being appropriate for β to take the agent to have acted. Whether these concepts are identical, whether there is any priority in the order of their explanations is subject to debate, which is orthogonal to the task before us. What is clear is that if one adopted the internal standard of reasonableness$_A$, one would have to argue that one has thereby captured our concept of action.[10] In deciding to use the external reading of reasonableness$_A$ we make a jump over a big metaphysical issue of how norms are instituted, how they relate to the participants' attitudes toward norms.[11] We will simply assume that these issues have been resolved.

The construal of reasonableness as an external standard should also answer the initial misgivings one may have had about the employment of the concept of reasonableness (see the introduction to the chapter, p. 79). Let us consider the suspicion that what may be reasonable for one person to expect of someone may not be reasonable for another.[12] Take the concept of reasonableness$_A$.

The objection is that it is possible that the following situation occur: it is reasonable$_A$ for β to expect of α that α φ, but it is unreasonable$_A$ for γ to expect of α that α φ. Here is an alleged example of such a situation. Suppose that β and γ are to judge whether Smith should take part in a car race. According to β's sources, Smith is in an excellent form. So, one might want to conclude that it will be reasonable$_A$ for β to expect of Smith that he take part in the race. According to γ's reconnaissance, Smith suffers

---

[10] The shape for such an argument is given by Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994). Brandom's concern is much more general, concerning the very nature of norms as such. He argues that we should understand the nature of norms in terms of the normative attitudes of participants in normative practices. At the same time, he shows that such an understanding does not obliterate the objectivity of norms, leaving space for the possibility that everyone is wrong.

[11] This is the central problem tackled in Brandom's *Making It Explicit, op. cit.*

[12] The second suspicion (what may be reasonable for a person to expect of another at one time may change when the person changes her false beliefs) can be treated in an identical fashion.

from a rare pulmonary disease which would cause him to lose consciousness in situations he is likely to encounter during the race. So, it will be unreasonable$_A$ for γ to expect of Smith that he take part in the race. However, my insistence that reasonableness$_A$ is to be used as an external standard prohibits the application of the concept in this way. Instead, we should say that β *believes* that it is reasonable$_A$ to expect of Smith that he take part in the race, and that γ *believes* that is unreasonable$_A$ to expect of Smith that he take part in the race.[13]

3. Reasonableness, Conflict and Contrary Expectations

Could it be that it is both reasonable and unreasonable to expect of an agent that she perform an action? Could it be that it is reasonable to expect of an agent that she φ and to expect of her that she not-φ? The answers to these questions depend on what concept of reasonableness is at stake.

A. Reasonableness$_A$, Reasonableness$_N$ and Conflict

Assuming that we interpret the concepts of reasonableness in external terms, the question might arise whether there is a possibility of conflict. We may formulate the question more precisely as follows. Is it possible for the following situations to occur:

(a) It is *reasonable$_A$* to expect of α that α φ and it is *unreasonable$_A$* to expect of α that α φ (reas$_A$[α φ] & unreas$_A$[α φ])?

(b) It is *reasonable$_N$* to expect of α that α φ and it is *unreasonable$_N$* to expect of α that α φ (reas$_N$[α φ] & unreas$_N$[α φ])?

Given our understanding of what it means to say that it is reasonable to expect something of a person, we are not committed to saying that it must be reasonable for the same person to expect contrary things of another. Let us make this explicit:

---

[13] The same will apply to the concept of reasonableness$_N$ except where we are dealing with a possible conflict of values. In view of the less fundamental importance of the concept of reasonableness$_N$ there is no reason to preclude the possibility that "what may be reasonable for you may not be reasonable for me" or to think it dangerous.

(a') It is reasonable$_A$ for β to expect of α that α φ but it is unreasonable$_A$ for γ to expect of α that α φ.

(b') It is reasonable$_N$ for β to expect of α that α φ but it is unreasonable$_N$ for γ to expect of α that α φ.

If one were to give an internal reading of reasonableness (a') and (b') would be satisfied trivially. It would be sufficient that β and γ held different (at least one of them erroneous) beliefs pertaining to the matter at hand. On the external reading of reasonableness (a') and (b') are not trivially satisfied.

Given the suggested intuitive meaning we have assigned to the concept of reasonableness$_A$, it is impossible for (a') to occur. Intuitively, it will be reasonable$_A$ for β to expect of α that α φ only if it is "within α's power" to φ. It will be unreasonable$_A$ for γ to expect of α that α φ only if it is not "within α's power" to φ.[14] It is not possible that φing both be "within α's power" and not be "within α's power." Hence, the expectation is either reasonable$_A$ or unreasonable$_A$ but not both. Indeed, in view of the role that we give to the concept of reasonableness$_A$, this guarantees the objectivity of our concept of action. In view of the fact that the concept of reasonableness$_A$ will play a fundamental role in determining whether an action has been performed, if (a') were possible, it would be also possible for an agent's performance to be both an action and a mere happening (a non-action). This would violate a fundamental truth, which is a prerequisite of any theory of action:

> No performance is both an action and a nonaction (a mere happening).

It is never the case that an agent's raising his arm (intentionally, say) is also a case of the agent's arm rising (uncontrollably, involuntarily). It is never the case that an agent's

---

[14] On an internal reading of reasonableness$_A$: it is reasonable$_A$ for β to expect of α that α φ if and only if β believes that it is "within α's power" to φ; it is unreasonable$_A$ for γ to expect of α that α φ if and only if γ believes that it is "within α's power" to φ. It is certainly possible for β to believe that it is "within α's power" to φ and for γ to believe that it is not.

bending his knee is also a case of the agent's knee curving in a spasm. A performance is either one or the other but never both.

But it is not clear that the concept of reasonableness$_N$ is similarly restricted. The possibility of (b') would imply that there is an irresolvable conflict of values in support of and against the expectation. Since what makes normative expectations reasonable$_N$ are not only moral values[15] but also cultural ones, the possibility of such a conflict is quite plausible. At the same time, it is clear that this is a proper subject for axiology or ethics, not specifically for action theory. In fact, the concept of reasonableness$_N$ will play a relatively minor role in the account of action we will develop. Its role will be limited to the interpretation we give of what an agent has done, once it is settled (by appeal to reasonableness$_A$) that the agent has done something. Given this role of the concept of reasonableness$_N$, conflict (b') (if possible) would amount to a dispute as to whether it is appropriate to interpret what the agent has done in a certain way or not. And that is a conflict the possibility of which would not undermine the very possibility of an account of action (in sharp contrast to possibility of the conflict generated by (a')).

In conclusion, it is impossible for an expectation to be both reasonable$_A$ and unreasonable$_A$. The possibility of such a conflict would undermine the very viability of an account of action that appeals to reasonableness$_A$. It is not as clearly impossible for an expectation to be both reasonable$_N$ and unreasonable$_N$. The possibility of such conflict depends on one's position on the possibility of conflicts of value more generally. I will remain uncommitted on this point.

B. Reasonableness$_A$, Reasonableness$_N$ and Contrary Expectations

Abstracting from possible conflicts of value, if it is (all-out) reasonable$_N$ (for β) to expect of an agent that she φ, then it is *not* (all-out) reasonable$_N$ (for β) to expect of her

---

[15] Though the topic is hotly disputed, there are views according to which even moral values are not absolute. See, e.g. Gilbert Harman, "Moral Relativism Defended," *Philosophical Review* 84 (1975), 3-22; "Relativistic Ethics: Morality as Politics," in (eds.) Peter A. French, Theodore E. Uehling, Jr., Howard K. Wettstein, *Studies in Ethical Theory* (Minneapolis: University of Minnesota Press, 1980), pp. 109-121; J.L. Mackie, *Ethics. Inventing Right and Wrong* (New York: Penguin Books, 1977); Bernard Williams, "Conflicts of Values," in *Moral Luck, op. cit.*, pp. 71-82. For a nice survey, see Robert M. Stewart, Lynn L. Thomas, "Recent Work on Ethical Relativism," *American Philosophical Quarterly* 28 (1991), 85-100.

that she not-φ, and vice versa. If, all things considered, it is reasonable$_N$ for me to expect of you that keep your side of the desk tidy, then it is unreasonable$_N$ for me to expect of you that you keep your side of the desk messy.

This is frequently not the case for reasonableness$_A$. Suppose that someone taking some (medical) drugs suffers from a temporary loss of control in his arms. Such a condition of his makes it unreasonable$_A$ to expect of him both that he perform certain tasks involving his arms as well as that he not perform them. To clarify, let us take the example of pushing a ball off a table. His condition makes it unreasonable$_A$ to expect of him that he push the ball off the table. It would be quite inappropriate for someone to complain that he failed to do so despite being asked, for instance. But it also makes it unreasonable$_A$ to expect of him that he not push the ball off the table. It would equally inappropriate for someone to complain that he did push the ball of the table despite being asked not to. In this case his condition renders two contrary expectations unreasonable$_A$. Frequently, when it is not within "the agent's power" to fulfill an expectation, it is not in his power to fulfill the contrary expectation.

• • • •

I have distinguished two senses in which normative expectations can be reasonable or unreasonable. Intuitively, a normative expectation is unreasonable$_A$ if it is not "within an agent's power" to fulfill it. It is reasonable$_A$ otherwise. A normative expectation is unreasonable$_N$ if it would be illegitimate for one person to hold another to the expectation (e.g. it is unreasonable$_N$ for you to expect your neighbor to do your laundry on a regular basis in normal circumstances). A normative expectation is reasonable$_N$ if there are reasons that justify or support the expectation (e.g. it may be reasonable$_N$ for you to expect your neighbors to collect your mail while you are gone in view of the fact that you will not be able to do it yourself, that you have asked them politely, that you have collected their mail for them in the past). A normative expectation can also be neither reasonable$_N$ nor unreasonable$_N$ if there are no reasons that justify the expectation and no reasons that make the expectation illegitimate (e.g. it may be neither reasonable$_N$ nor unreasonable$_N$ for you to expect yourself to gently touch the leaves of the trees you pass by).

As we have seen, the concepts of reasonableness$_A$ and reasonableness$_N$ are independent of one another. It is possible for an expectation to be reasonable$_A$ but not reasonable$_N$ (e.g. your expectation of your neighbor to do your shopping may be illegitimate but what is expected would be within your neighbor's power to do), and it is possible for an expectation to be unreasonable$_A$ but reasonable$_N$ (e.g. an expectation of a student to turn in his paper may be legitimate but unreasonable$_A$ in view of the fact that he lies incapacitated in the hospital).

It is the concept of reasonableness$_A$ that will matter in the account of action offered in Chapter VI. I will give an account of reasonableness$_A$ in Chapter V. In this chapter, we have seen that some of the initial worries about the concept of reasonableness can be allayed by appealing to the metaphor of a performance being "within the agent's power," which I proposed as an approximation of the meaning of reasonableness$_A$. In particular, in section 2, I have suggested that reasonableness$_A$ ought to be construed as an external rather than an internal standard. Accordingly, the epistemic position of a particular person does not affect whether it is reasonable$_A$ for her to hold another person to an expectation. She might have good reasons to falsely believe that it is reasonable$_A$ to hold a person to an expectation, but her belief in no way affects the judgment that it is unreasonable$_A$ to hold that person to the expectation.

With these preliminary issues settled, let us proceed to the account of reasonableness$_A$.

CHAPTER V.

PRACTICAL RESPONSIBILITY III:

REASONABLE$_A$ NORMATIVE EXPECTATIONS

In Chapter IV, I have distinguished two senses in which normative expectations can be reasonable and answered some preliminary questions about the concept. In view of the fact that the concept of reasonable$_A$ normative expectations will be crucial to the account in Chapter VI, we need to dispense with the guiding metaphor of what is "within the agent's power" and offer a systematic account of reasonableness$_A$. Section 1 proposes an account of prima facie reasonableness. Sections 2 and 4 develop the concept of a defeating condition, taking care to avoid the fundamental problem.

1. When Are Normative Expectations Prima Facie Reasonable$_A$?

Thus far, the only restriction I have placed on normative expectations is that they be practical. This is to say, when we expect of someone that he bring it about that *p*, '*p*' must be logically and physically contingent. However, there are examples of expectations whose results are contingent and which we would judge intuitively unreasonable$_A$. For example, we would think that it is unreasonable$_A$ to expect of a person that he speak all the known languages fluently, but his speaking so many languages is not impossible. We would also think that it is unreasonable$_A$ to expect of a person that she win a lottery, but it is not impossible for her to win it. By a similar token, we would think it unreasonable$_A$ to expect of a person that she breathe, but it is certainly not necessary that she does. It would be unreasonable$_A$ to expect of a person that she bring it about that the seasons change, yet it is not necessary that they do. These are all

examples of performances that we would think are not "within the agent's power." We now need to dispense with the metaphor.

In this section, I want to begin by characterizing the notion of prima facie reasonableness$_A$. Prima facie reasonableness$_A$ is meant to capture the idea of what is "within our power" (as humans, say) to do. In particular, it abstracts from any special considerations the agent deserves in virtue of her particular circumstances. We will take the special circumstances into account when discussing the defeating conditions in sections 2 and 4. For example, the expectation to tell colors apart is prima facie reasonable$_A$, for it is something that is in general "in our power" to do. However, the prima facie reasonableness$_A$ of such an expectation is defeated, if the agent whom we hold to the expectation is color-blind.

I suggest that we ought to understand prima facie reasonableness$_A$ negatively, viz. in terms of what is *not* prima facie *un*reasonable$_A$ (section A). There are two kinds of situations in which an expectation is prima facie unreasonable$_A$: first, when it would be systematically frustrated by most agents in most circumstances; second, when it would be systematically fulfilled while its contrary is systematically frustrated by most agents in most circumstances. A concept that is crucial in this characterization is that of a systematic correlation. I will treat it as a theoretical place-holder and not give an account of it, but I will say a few words about it in section B.

I will speak of an expectation *to* φ (rather than an expectation *of* α *that* α φ) as being systematically fulfilled or frustrated or neither. Similarly, I will speak of certain conditions (defeating conditions) being systematically correlated with the fulfillment or frustration of an expectation *to* φ (rather than an expectation *of* α *that* α φ). I will use this manner of speaking in order to emphasize that the systematic correlations at stake hold irrespectively of the particular agent who is held to the expectation on a particular occasion.

A. Prima Facie Reasonableness$_A$

There are at least two ingredients in the metaphor of a performance being "within an agent's power." First, there is a sense in which the agent must be able to perform the action in question. If the agent could not succeed in performing the action, we would

intuitively think that the action was not "within the agent's power" at the time. An expectation of a two-year old child to win an Olympic swimming competition would surely be unreasonable$_A$. Second, there is a sense in which the agent must be able to make a difference. If what the agent is about to do would happen whether or not the agent did anything, we would be inclined not to think that what happened was in the agent's power.

Consider three types of cases where it would be intuitive to say that it would be unreasonable$_A$ to expect an agent to perform an action. It would be unreasonable$_A$ to expect of someone that he win an (unrigged) lottery. Winning the lottery is not something that is "up to him," that is "within his power" — it is almost certain that he will lose. It would also be unreasonable$_A$ to expect of a person that she breathe,[1] or that she make her heart beat. Breathing and having one's heart beat are not "within the agent's power" — it is something that happens no matter what the agent does. Finally, it would be unreasonable$_A$ to expect of the agent that he throw a coin so it comes up heads. Unlike the first case, the coin will not almost certainly come up tails; unlike the second case, the coin will not almost certainly come up heads. Yet, the coin's coming up heads is not something the agent controls. We can capture these three kinds of cases using the following test.

Let us begin with a deceptively simple scenario, which will suggest the gist of the test. Let us imagine that we want to test whether an agent can perform a certain type of action. To do so, we will give him a series of tasks, to which he will respond in the best possible way: we are assuming, in other words, that he is cooperative, that there are no other designs, intentions, expectations in play, the agent is at ease, under no pressure, etc.[2] The tasks are of two kinds, to φ and not to φ, and they are interspersed randomly in a series.

Four situations are of special interest. Suppose that an agent systematically frustrates the expectation to φ (situations (iii) and (iv) in Table 1). When he is expected

---

[1] It might be reasonable$_A$ to expect of a person to take a breath at a particular moment, or stop breathing for a couple of seconds, but not to stop breathing altogether or breathe at all.

to φ, he does not. In such a case, it would be unreasonable$_A$ to expect of the agent that he φ. The agent cannot succeed in fulfilling the expectation. Suppose that the agent regularly fulfills the expectation to φ but frustrates the expectation not to φ (ii). What this will mean is that the agent φs indiscriminately. In such a case, we would tend to think that the agent's φing is not up to him, that the agent cannot make a difference, and hence that it would be unreasonable$_A$ to expect of him that he φ. This configuration would obtain if we expected the agent to breathe, for example. Finally (i), when the agent fulfills all the expectations (when expected to φ, the agent responds by φing, when expected not to φ, the agent responds by not φing), we would tend to think that φing and not φing are "within the agent's power," that it is not unreasonable$_A$ to expect of the agent that he φ.

	Task: φ	Task: not-φ
(i)	fulfilled (φ)	fulfilled (not-φ)
(ii)	fulfilled (φ)	frustrated (φ)
(iii)	frustrated (not-φ)	fulfilled (not-φ)
(iv)	frustrated (not-φ)	frustrated (φ)

Table 1. Possible result patterns of a simplified test sequence.

It may be worthwhile noting that there is an interesting difference between situations (iii) and (iv). Situation (iii) is analogical to situation (ii). When the agent systematically frustrates the expectation to φ but fulfills the expectation not to φ, the agent simply does not φ. Once again, it would be unreasonable$_A$ to expect him to φ (or not to φ). We would judge that his not-φing was not up to him. This case corresponds to what would happen were the agent expected to win the lottery, for example. That expectations would be systematically frustrated, while its contrary would be systematically fulfilled. Situation (iv) is different, however. Here the agent is counter-

---

[2] This is an unrealistic assumption. I am making it in order to sharpen the intuitions at stake. The account I am proposing does not depend on it, however.

competent, as it were. When expected to φ, he does not φ. When expected not to φ, he does φ. The case is somewhat curious. The most natural way of thinking about it is that the agent acts contrary to the expectations. But this contradicts our simplifying assumption that the agent is cooperative, under no other pressures, etc. At the same time, it would seem an altogether implausible accident of nature that despite being in the best possible conditions, the agent always frustrates the expectations to which he is held. However curious the case is in its pure form (given the simplifying assumptions), there are cases that appear to approximate it. Take the pair of expectations to throw a coin so that it comes up heads and to throw it so that it comes up tails. It is certainly not true that anyone of us is counter-competent with respect to throwing a fair coin. It is true, however, that most of the expectations in a random series would be frustrated in the long run.

This simplified test scenario allows to make a little clearer some of our intuitions concerning especially the situations in which it would be unreasonable$_A$ to hold an agent to an expectation. It will be immediately objected, however, that the test scenario is unrealistic. It presupposes that the agent responds to the expectation in the very best conditions. But such conditions are almost never present. And even if they were, it is not clear that we could count on them in giving an account of unreasonableness$_A$. I do not believe that we have to rely on such strict test conditions. Rather, the way in which we gather the knowledge concerning unreasonable$_A$ expectations and conditions that make expectations unreasonable$_A$ (defeating conditions) takes account of our general knowledge (cutting across times, places, particular agents) concerning the way in which most agents behave. Rather than requiring that the agent fulfill or frustrate an expectation, we might require that most agents across a wide range of circumstances that approximate the ideal test conditions systematically fulfill or frustrate the expectation. (I will say a little more about the concept of a systematic correlation in the next section.) In this way, the special circumstances will tend to be evened out, as it were. For instance, when subjecting a particular agent to such a test, we might worry about what social scientists worry about when testing humans, viz. that the individual's responses will be changed by the very fact that they are taking part in an artificial test situation. In the simplest case, the individual might not be responding in the best possible way to the task,

but might be doing the opposite on purpose, say. These kinds of peculiarities will tend to disappear if we collect a large number of data cutting across a variety of settings.

We can dress these intuitions thus:

(Success Condition):

> It is prima facie unreasonable$_A$ to hold α to an expectation to φ if the expectation to φ is systematically pf-frustrated.[3]

(Difference Condition):

> It is prima facie unreasonable$_A$ to hold α to an expectation to φ if (a) the expectation to φ is systematically pf-fulfilled while the expectation not to φ is systematically pf-frustrated.[4]

The conditions allow us to understand why non-practical expectations are prima facie unreasonable$_A$. The expectation to see to it that 2+2=3 would be systematically pf-frustrated and so unreasonable$_A$ in virtue of the success condition. Likewise, the expectation to see to it that 2+2=4 would be unreasonable$_A$ in virtue of the difference condition: it would be systematically pf-fulfilled while the contrary expectation (to see to it that $2+2\neq4$) would be systematically pf-frustrated. The expectation to see to it that the e-mail goes through faster than light would be systematically pf-frustrated, while the expectation that the Earth move around its axis would be systematically pf-fulfilled, while the contrary expectation systematically pf-frustrated.

But the concept of prima facie reasonableness$_A$ can discriminate further than cases of non-practical expectations. An expectation to win a fair lottery would be prima facie unreasonable$_A$. Such an expectation would surely be systematically pf-frustrated. An

---

[3] In Chapter III, I have distinguished between prima facie and agentive fulfillment (frustration, respectively) of expectations. An expectation to φ is agentively fulfilled only by *actions* of φing (raising the arm), while it is prima facie fulfilled by performances, whether they be actions or nonactions (raisings and risings of the arm). I use 'pf-' to mark the prima facie sense of fulfillment (frustration, respectively). I will discuss the importance of this constraint in section 2.

[4] The success and difference conditions roughly correspond to what Belnap calls the positive and the negative condition of agency ("Before Refraining: Concepts for Agency," *Erkenntnis* 34, 1991, 137-169; see also Belnap and Perloff, "Seeing to It that," *op.cit.*). Unlike the positive condition, the success condition does not require that the success be guaranteed. The difference condition excludes the situations where the agent cannot make a difference but without committing us to incompatibilism.

expectation to speak all the known languages fluently would be systematically pf-frustrated, and so is prima facie unreasonable$_A$. By contrast, an expectation to breathe would be systematically pf-fulfilled while its contrary would be systematically pf-frustrated, and thus prima facie unreasonable$_A$. It would also be prima facie unreasonable$_A$ to expect of a person that she bring it about that the seasons change, for such an expectation would be systematically pf-fulfilled, while its contrary would be systematically pf-fulfilled.

We will work under the hypothesis that no other conditions characterize prima facie unreasonableness$_A$. We can then define reasonableness$_A$ negatively:

(R) an expectation is *prima facie reasonable*$_A$ iff it is not prima facie unreasonable$_A$.

As agents, we are guilty until proven innocent.[5] It is reasonable$_A$ to expect of us any performance unless there are special conditions that would make such an expectation unreasonable$_A$. The concept of reasonableness$_A$ is thus characterized negatively in terms of what it is not unreasonable$_A$ to expect of an agent.[6]

We should be clear that the concept of prima facie reasonableness$_A$ is not yet a concept that would be sufficient to capture our intuitions concerning what it would be (all-out) reasonable$_A$ to expect of a particular agent. It is surely not reasonable$_A$ to expect of a student who has been taken seriously ill that he turn in homework on time, yet such an expectation would be prima facie reasonable$_A$. It is unreasonable$_A$ to expect a blind person to read aloud, but such an expectation is prima facie reasonable$_A$. It would be unreasonable$_A$ to expect of a newly arrived foreigner that he speak like a native but such an expectation is prima facie reasonable$_A$. It is clear that we need to take the special circumstances in which the agent finds herself into account. This is the role of defeating conditions.

---

[5] A similar principle concerning moral responsibility is defended in Keith D. Wyma, "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34 (1997), 57-70.

[6] This is also the deep reason why the concept of reasonableness$_A$ does not admit an intermediate category of performances that are neither reasonable$_A$ nor unreasonable$_A$. We will remember that the concept of reasonableness$_N$ does admit of such performances (see Chapter IV).

B. Systematic Correlations

As we saw in the last section, one of the concerns is that the agent might be uncooperative, intent on acting contrary to the expectations, etc. It is for this reason that the simplified test discussed there relied on assuming that the agent finds himself in ideal conditions (that he is cooperative, under no pressures, that no other intentions or expectations are in play, etc.). These conditions are never or rarely actually satisfied but they can be approximated.

We might imagine an expector, who chooses agents, times, occasions at random, and subjects the agents to the expectation that they φ. She will exclude those who are clearly uncooperative, who are under special pressures, or where she suspects other expectations to be involved. Given this large set of data, she can then decide that an expectation to φ is systematically frustrated if most agents, most of the time, have frustrated the expectation in question; or that the expectation to φ is systematically fulfilled if most agents, most of the time, have fulfilled the expectation in question.[7] Similarly, given large amounts of data, such an expector can tell whether a particular type of event, *C*, is systematically correlated with the fulfillment or frustration of an expectation to φ. A *C*-type event will be systematically correlated with the fulfillment (frustration) of an expectation to φ if, given the occurrence of events of type *C*, the expectation to φ is systematically fulfilled (frustrated) ceteris paribus.

Although we do not have access to such a large set of data, we do have access to hypotheses and theories concerning the mechanisms involved in the fulfillment and frustration of expectations. Thus, we would consider it quite intuitive to think that being in a coma is systematically correlated with the frustration of the expectation to talk, to smile, etc., and with the fulfillment of the expectation to lie motionless. Our judgment is affirmed not only by the preponderance of data but also by our understanding of the causal processes involved in a coma.

---

[7] By not insisting that the quantifiers be universal, we can further take into account cases where the ideal conditions have been violated.

Many of the systematic correlations (especially those involving defeating conditions) will be causal in nature. In fact, when such correlations are causal, and when we understand the causal mechanisms behind them, we are most confident of the correlation. However, it would not be wise to exclude the possibility of "merely statistical" correlations. This aggravates the issue of the vagueness of the concept of systematic correlation. The notion appeals to a vague quantifier 'most'. It is not clear furthermore that it could be made any more precise without introducing ad hoc arbitrariness. While this is certainly the case, it is not clear that we should seek any more precision for the purposes at hand. For one thing, this issue is not peculiar to the domain of agency. It is a more general problem confronted on a daily basis in scientific testing, for example. For another, it would be unwise to exclude the possibility that the vagueness is a part of the concept itself and that there will be gray cases where it will be simply unclear whether a systematic correlation is in place or not.

Another problem with cashing out the concept of a "systematic correlation" consists in the fact that any such attempt will involve an appeal to ceteris paribus conditions. Drinking great quantities of coffee may be systematically correlated with extreme agitation. But not if one is a "caffeine addict." If one consumes large amounts of caffeine in the first place, one's reaction will be very different. The original correlation holds only ceteris paribus. Once again, however, this fact does not present any special problem for action theory. It is a general problem not only for scientific[8] but also for most ordinary claims we make.[9]

2. Defeating Conditions

Defeating conditions make an otherwise reasonable$_A$ expectation unreasonable$_A$, or an otherwise unreasonable$_A$ expectation reasonable$_A$. We can speak of *defeating*

---

[8] Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983); Carl G. Hempel, "Provisos," in (eds.) Adolf Grunbaum, Wesley C. Salmon, *The Limitations of Deductivism* (Berkeley: University of California Press, 1988), pp. 3-22; Marc B. Lange, *The Design of Scientific Practice. A Study of Physical Laws and Inductive Reasoning* (Ph.D. Dissertation: University of Pittsburgh, 1990); Leszek Nowak, *The Structure of Idealization* (Dordrecht/Boston: Reidel, 1980).

[9] Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994). Nicholas Rescher, *Standardism*, forthcoming.

conditions proper in the former case (section A) and of *counterdefeating conditions* in the latter (section B). In section C, I consider the way in which the fundamental problem bears on the account of reasonableness$_A$. Finally, in section D I ask whether we should think of defeating conditions as causes.

A. Defeating Conditions Proper

The expectation that a student turn in homework on time is prima facie reasonable$_A$. The expectation is not systematically frustrated, nor is its contrary. But when the student falls seriously ill, its reasonableness$_A$ is defeated. The expectation that a person run in a race is prima facie reasonable$_A$. But it is no longer reasonable$_A$ if she has broken a leg. The expectation that a person walk straight is reasonable$_A$ but not when he has been pushed by another. These are examples of what I will call defeating conditions of the first kind, or hindering conditions.[10] They can be understood on the lines suggested above:

(1) An event of type *C* is a *defeating condition of the first kind (hindering condition)* with respect to an expectation to φ iff the occurrence of an event of type *C* is systematically correlated with the pf-frustration of the expectation to φ and with the pf-fulfillment of the expectation not to φ.

Breaking a leg is systematically correlated with the pf-frustration to run a race and with the pf-fulfillment of the expectation not to run a race. It is a defeating condition with respect to the expectation to run the race. Thus, while it may be prima facie reasonable$_A$ to expect of a person that she run the race, in view of the fact that she has broken a leg, it would be unreasonable$_A$ to hold her to the expectation. Being seriously ill is systematically correlated with the pf-frustration of the expectation to turn in homework on time. In view of the fact that a student has fallen seriously ill, it would be unreasonable$_A$ to expect him to turn in the homework. Being pushed is systematically correlated with the pf-frustration of an expectation to walk straight, hence it would be

unreasonable$_A$ to expect of an agent who has been just pushed by another that he walk in a balanced way. These and others are examples of defeating conditions of the first kind. Suffering a spasm in one's arm is systematically correlated with the frustration of various kinds of expectations having to do with the control over one's arm. Not knowing that one is to be present at a certain meeting is systematically correlated with the frustration of the expectation to be at the meeting. Being in a coma is systematically correlated with the frustration of a great many expectations. Not having access to the right equipment is systematically correlated with the pf-frustration of the expectation to build a bridge. And so on.

It needs to be emphasized that the concept of a defeating condition is relativized to an expectation. An event-type that may be a defeating condition with respect to one expectation need not be a defeating condition with respect to another. Breaking a leg makes the expectation to run a race unreasonable$_A$, but it does not defeat the reasonableness$_A$ of the expectation to remember your friend's birthday. When an agent suffers from a tic it is unreasonable$_A$ to expect of him that he wink three times, but it may still be reasonable$_A$ to expect him to do the fox-trott.

The second kind of defeating condition corresponds to the difference condition rather than the success condition.

(2) An event of type *C* is a *defeating condition of the second kind (compelling or forcing condition)* with respect to an expectation to φ iff the occurrence of an event of type *C* is systematically correlated with the pf-fulfillment of the expectation to φ and with the pf-frustration of the expectation not to φ.

It is prima facie reasonable$_A$ to expect of a person that he walk. But this expectation ceases to be reasonable$_A$ if the person is in fact physically forced to walk by another. The application of appropriate physical force is systematically correlated with the pf-fulfillment of the expectation to walk and with the pf-frustration of the expectation not to

---

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

walk. Breaking a leg is systematically correlated with the pf-fulfillment of an expectation not to take part in a race, and with the pf-frustration of the expectation to take part in a race, so breaking a leg counts as a defeating condition (of the second kind) for the expectation not to take part in race.

Finally, I want to mention defeating conditions of a third kind, to which I have already alluded in discussing the concept of systematic correlation. Suppose that an agent's hands tremble erratically, severely impairing his job manufacturing electronic chips, say. However, despite the tremble his chances of succeeding are about 50%. In other words, given an intuitive grasp of 'systematic correlation', it would be wrong to say that the condition is systematically correlated either with the frustration or with the fulfillment of the expectation to connect the chip. Yet, it is true that neither the manufacturer of the chips (expecting the agent to connect the chips correctly) nor the rival manufacturer (expecting the agent to sabotage the chip production) can count on the agent. In such a case, our intuitive judgment that what is within the agent's power is limited can be manifested if we subjected the agent to a test. Suppose that the agent was to fulfill a series of expectations to connect the chips correctly and to connect the chips incorrectly, in a random order. In the long run, it would become evident that although the agent does occasionally fulfill the expectations, he systematically frustrates most of them.

(3) An event of type *C* is a *defeating condition of the third kind* with respect to a pair of expectations to φ and not to φ iff the occurrence of an event of type *C* is systematically correlated with the pf-frustration of expectations to φ and not to φ (in a random series).

We can offer a preliminary characterization of reasonableness$_A$.

It is reasonable$_A$ to expect of α that α φ if no defeating condition with respect to the expectation to φ occurred.[11]

---

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

We should note that given the characterization of defeating conditions of the first and the second kind, if *d* is systematically correlated with pf-frustration of the expectation to φ and with the pf-fulfillment of the expectation not to φ then *d* is systematically correlated with pf-fulfillment of the expectation not to φ and with the pf-frustration of the expectation to φ. This is to say that a defeating condition of the first kind is a defeating condition of the second kind for the contrary expectation. For example, lack of shooting skills is systematically correlated with the pf-frustration of the expectation to shoot the bulls-eye (and with the pf-fulfillment of the expectation not to shoot the bulls-eye). It is eo ipso systematically correlated with the pf-fulfillment of the expectation not to shoot the bulls-eye and with the pf-frustration of the expectation to shoot the bulls-eye. Believing that one's meeting is at 9am is systematically correlated with the pf-frustration of the expectation to be at the meeting at 8am. It is eo ipso systematically correlated with the fulfillment of the expectation not to be at the 8am meeting.

It should be emphasized that reasonableness$_A$ is relative to the way in which the performances are described. Consider an example. Suppose that Tamara has lost control over some of her fingers. She can move her index finger at will. She can also move her middle finger without problems. But she cannot move the remaining fingers at all. Given her condition, it would be, among other things: unreasonable$_A$ to expect of her that she move her thumb, and reasonable$_A$ to expect of her that she move her index finger. Would it be reasonable$_A$ to expect of her that she *not move* her thumb? It seems clear that the answer must be that it would be unreasonable$_A$ to expect of her that she not move her thumb. Her condition is systematically correlated with the pf-fulfillment of that expectation (and with the pf-frustration of the expectation to move her thumb). However, her moving her index finger is a way of not moving her thumb. So, it might appear problematic that though the expectation to move her index finger is reasonable$_A$, the expectation not to move her thumb is not. This impression disperses, however, in view of the fact that reasonableness$_A$ of expectations is sensitive to description. One and the same performance may be reasonable$_A$ under some descriptions but not under others.

B. Counterdefeating Conditions

So far we have considered conditions that render prima facie reasonable$_A$ expectations unreasonable$_A$. I would like to briefly mention a class of counterdefeating conditions which render prima facie unreasonable$_A$ expectations reasonable$_A$. Once again the class includes varied conditions. A large portion of it is occupied by special abilities possibly due to special equipment. It is prima facie unreasonable$_A$ to expect of a person that she perform a pirouette. But it would be reasonable$_A$ to hold a skilled skater to the expectation.[12] The reason why the expectation to perform a pirouette is prima facie unreasonable$_A$ is that such an expectation would be systematically pf-frustrated by most people. However, the expectation would not be systematically pf-frustrated by skilled skaters. Having a leg amputated will usually make the expectation to walk without support unreasonable$_A$. It will count as a defeating condition of the first kind: it will lead to the systematic frustration of the expectation. However, when the agent is equipped with a prosthesis the expectation would no longer be systematically frustrated.

We can amend our preliminary characterization of reasonableness$_A$.

> It is reasonable$_A$ to expect of α that α φ if either (a) no defeating condition (with respect to the expectation to φ) occurred, or (b) such a defeating condition did occur but it was countered by an appropriate counterdefeating condition.

C. The Fundamental Problem and the Evolution of Defeating Conditions

Before going on, we should consider the fundamental problem once again. The concern with the sort of account I am proposing is that it reverses the natural order of the concepts of action and responsibility. This is also evident here.

The concept of reasonable$_A$ expectations, or more precisely the notion of what it is reasonable$_A$ to expect of an agent, is to give us a way of understanding the concept of

---

[12] The example, together with the observation, is borrowed from Annette Baier ("The Search for Basic Actions," *American Philosophical Quarterly* 8, 1971, p. 164).

action. In order for thus developed account not to be circular, the concept of reasonable$_A$ expectations needs to be construed without presupposing the concept of action. I have already explained how the concept of a normative expectation can be construed without presupposing the concept of action. Indeed, even though I allowed that we speak of holding the agent to an expectation to *perform an action*, I have shown that we can interpret this phrase in an innocent way (allowing that expectations are prima facie fulfilled by performances in general: actions and nonactions alike). I now have to demonstrate that the notion of reasonableness$_A$ can acquire an equally innocent interpretation. I have already hinted at how to do so in section 1.A above, where the concept of prima facie unreasonableness$_A$ is understood in terms of systematic prima facie frustration of normative expectations, and in section 2.A, where the concept of defeating conditions is understood in terms of prima facie fulfillment and frustration of expectations. Let me say a little bit to motivate this construal. It will be best to use to the notion of a defeating condition as an example.

Take the concept of a hindering condition with respect to the expectation to φ, i.e. a defeating condition that is systematically correlated with the frustration of an expectation to φ. So, one might say, breaking a leg is systematically correlated with the frustration of the expectation to run the race. The crucial question that we must ask is what sort of concept of frustration is at stake.

We can distinguish at least three concepts of frustration. An expectation to φ is *agentively* frustrated by actions that can be described as not-φings. An expectation to φ is *non-agentively* frustrated by nonactions (mere happenings) that can be described as not-φings. Finally, an expectation to φ is *prima facie* frustrated (pf-frustrated) by any performances (actions and nonactions alike) that can be described as not-φings. Take the expectation to run a race as an example. It will be agentively frustrated when an agent decides not to run just because he does not feel like it and intentionally fails to run. It will be also agentively frustrated if the agent is called out on an emergency, and so fails to run the race without intending to do so but foreseeing that he will do so. The expectation will be non-agentively frustrated when the agent does not run the race but when his not running is not an action of his, as when he is lying comatose in the hospital

or when his leg is broken. The expectation will be prima facie frustrated in all these cases.

Ignoring the fundamental problem for a moment, let us ask what concept of frustration would fit the notion of a defeating condition. Take agentive frustration first. It is relatively clear that this is not the notion that is at stake. Breaking a leg is not systematically correlated with the agentive frustration of the expectation to run the race. (Someone who breaks a leg *might* also intend or have intended not to run the race, but breaking a leg seems to break the pattern rather than be a part of it.) What about non-agentive frustration? Here the intuitions seem to be quite clear: it fits like a glove. Breaking a leg is systematically correlated with the non-agentive frustration of the expectation to run the race. Someone with a broken leg is not going to run the race, and her not running the race will not be an action of hers. Her not running the race (because of a broken leg) is something that happens to the agent. We would be thus led to conclude that the concept of non-agentive frustration (frustration by mere happenings not actions) should be involved in the notion of a defeating condition.

And it is here, once again, that the fundamental problem arises. For the notion of non-agentive frustration just like the notion of agentive frustration presupposes the very distinction between actions and mere happenings that we want to explicate in terms of (among others) defeating conditions. In order to make sense of the notion of non-agentive frustration we need the concept of a mere happening (and so the distinction between actions and mere happenings). It would be circular to proclaim that we can understand the distinction between action and mere happening in terms of such a notion of defeating conditions.

In other words, we cannot characterize defeating conditions in terms of the concept of agentive frustration, for it misses the target. But we also cannot analyze defeating conditions in terms of non-agentive frustration — although we would be right on the target, the account would be circular. The only option that remains is to choose the concept of prima facie frustration. Only the concept of prima facie frustration would not render the account circular, for only that concept does not presuppose the distinction between actions and mere happenings.

But our characterization of defeating conditions is not quite sufficient. Here are two objections which rely on the simple fact that the set of agentive fulfillments and the set of non-agentive fulfillments are included in the set of prima facie fulfillments of an expectation. Consider the first fact. When an expectation is agentively fulfilled it is also prima facie fulfilled. So suppose that at a certain stage of the development of the concept of agency, it is discovered that there is a condition that is systematically correlated with the frustration of an expectation. As it turns out, however, the performances that it is correlated with are exclusively agentive (relative to the understanding of 'agentive' at that stage). In this case, it is still true that the condition is systematically correlated with prima facie frustration of the expectation but this is only because the set of agentive frustrations is by definition included in the set of prima facie frustrations. Were such a situation to occur, we would not have a reason to speak of a defeating condition. There would be little reason to speak of a condition that takes a certain kind of performance out of the agent's power. After all, when the condition occurs, the agent systematically performs only *actions*. Were the correlation to extend to cover not only what is recognized as actions but also what is recognized as mere happenings, there would be reason to suppose that a new defeating condition is at work, which would require us to change our conception of what is an action and what is not.

We can summarize this in the form of an informal principle. Let d be a potentially new defeating condition that is systematically correlated with the frustration of an expectation to φ. Let D be the set of existing defeating conditions, which determine whether an expectation is fulfilled and frustrated agentively (relative to D): when some condition of the set occurs, the expectation is fulfilled or frustrated non-agentively (relative to D).

Principle I :

> If d is systematically correlated with agentive fulfillment/frustration (relative to D)[13] of the expectation to φ but not with the non-agentive

[13] Note that the concept of agentive frustration and fulfillment are relativized to an existing set of defeating conditions. The use of such a concept does not lead to circularity.

> fulfillment/frustration (relative to *D*) of that expectation, then *d* is not a new defeating condition (relative to *D*).

For a different reason, the opposite situation does not engender new defeating conditions. Suppose that at a certain stage of the development of the concept of agency, it is discovered that there is a condition that is systematically correlated with the frustration of an expectation. As it turns out, however, the performances that it is correlated with are exclusively non-agentive. In this case, it is still true that the condition is systematically correlated with prima facie frustration of the expectations but this is only because the set of non-agentive frustrations is by definition included in the set of prima facie frustrations. Were such a situation to occur, we would not have a reason to speak of a new defeating condition either. Here, however, the reason is different. Given that the frustrations in question are all non-agentive, this means that in all these cases, some defeating conditions are already in play. The alleged new condition covers a terrain that is already covered by the existing set of defeating conditions.[14] It is not a new defeating condition.[15]

Principle II (Economy of Defeating Conditions):

> If *d* is systematically correlated with non-agentive (relative to *D*) fulfillment/frustration of the expectation to φ but not with agentive (relative to *D*) fulfillment/frustration of that expectation, then *d* is not a new defeating condition (relative to *D*).

The spirit behind the second principle could be in fact generalized. As long as the systematic correlation of a condition *d* can be fully understood in terms of the existing set of defeating conditions *D*, *d* is not a new defeating condition. The principle of economy of defeating conditions is intended to bar the introduction of "funny" agglomerative conditions.

---

[14] I can see one exception here. It may be that a bunch of defeating conditions correlated with the frustration of a variety of expectations is replaced with one defeating condition ('a syndrome'). Such a unification, however, would need to be supplemented with some theoretical benefits.

[15] This seems to be a general practice in law making. New laws are only introduced if the cases they are intended to cover are not already covered by any combination of already established laws.

Principle III (of Non-Composition of Defeating Conditions):

> If d_1 is systematically correlated with pf-frustration of the expectation to φ, d_2 is systematically correlated with pf-fulfillment of the expectation not to φ, then d_1-or-d_2 is not a new defeating condition with respect to the expectation to φ.
>
> If d is systematically correlated with pf-frustration of the expectation to φ, then not-d is not a new defeating condition with respect to the expectation to φ.

D. Are Defeating Conditions Causes?

So far, I have been speaking of defeating conditions occurring rather than causing the events that would otherwise be actions. There is no problem in supposing that defeating conditions sometimes do cause the events that would otherwise be actions. Sometimes it is in fact plain that they do. For example, when a spasm causes a hand to tremble and the spoon to fall out, the spasm causes a performance (the falling out of the spoon) that might appear as if it fulfills the expectation that the agent drop the spoon, and yet, the performance is something that happens to the agent in virtue of the fact that it has been caused by the spasm. Similarly, drugs may cause memory problems causing one to forget to pick up a child from school. A sudden wind gust may throw one forward causing one to fall onto somebody else. And so on.

Indeed, it might not be perhaps too outrageous to suggest that it is because defeating conditions frequently cause mere happenings (nonactions) that the language of causality suggests itself with respect to reasons causing actions. The mere happenings are events that might look like actions but are (typically) caused by defeating conditions. This might lead one to search for the corresponding "agentive" causes of actions.

Be this as it may, it is not clear what is gained by speaking of the defeating conditions as causing mere happenings. The problem is not only that the idea of causality is notoriously difficult to understand but rather that it is notoriously difficult to apply in at least certain kinds of cases. For while we have little problem understanding how the wind causes one to fall onto a crowd, we have progressively more problems in grasping

the sense in which one's oversleeping caused one not to go to the meeting, one's forgetting caused one not to do the homework, or the absence of power tools "caused" one not to build the bridge.

In the last case, the problem in construing defeating conditions as always causing mere happenings is perhaps most vivid. The fact that one does not have power tools suffices to make it unreasonable$_A$ to expect of one that one build a bridge. Suppose that John does not have the required power tools (through no fault of his own), so despite the fact that he intends to build a bridge and is contracted to do so, it would be unreasonable$_A$ to expect of him that he build the bridge. In such a case, we might be tempted to say that John's not building the bridge was caused by his not having the required power tools. But suppose that while it is true that John does not have the power tools, he does not build the bridge not because he does not have the required equipment but because he does not intend to do so in the first place. In fact the idea might appear ridiculous to him if he were confronted with it. In such a case, we might think that it is John's indisposition (lack of intention, preparation or what not) that caused him not to build the bridge rather than the lack of relevant tools. And yet, I think that both situations are exactly alike with respect to the defeating condition: it is because John lacks the power tools that it would be unreasonable$_A$ to expect him to build the bridge.[16] This may be independent of what actually *explains* his not building the bridge.[17]

I prefer therefore not to require that defeating conditions must cause mere happenings even though many of them do.

3. Some Objections

A. Are Desires Defeating Conditions?

When I want chocolate, I eat it. When I want to go for a walk, I usually go for a walk. Desires appear to be the paradigmatic examples of conditions that are systematically correlated with the fulfillment of the expectations they justify and with the

---

[16] In the second case, we might also have additional defeating conditions: his lack of skills, for instance.

frustration of the contrary expectations. Are they defeating conditions? Would they make the expectations unreasonable$_A$? I will answer this question more systematically in section 4. For now let me make three points.

Not all desires make expectations unreasonable$_A$, but some do. Compulsive desires do indeed make it unreasonable$_A$ to expect of the agent that she perform the action justified by the desire. A person's desire to wash her hands every five minutes makes it unreasonable$_A$ to expect her not to wash her hands, or to wash her hands. The washing of the hands is in this case beyond the compulsive-obsessive's control. But it seems clear that a person's non-compulsive desire for a walk does not make it unreasonable$_A$ in any way to expect of the person that she does or that she does not take the walk. In section 4, I will explain how to understand the difference between compulsive and non-compulsive desires.

One may develop some degree of skepticism with respect to the alleged systematicity of the correlation. Recall that in order for a condition to count as being systematically correlated with the frustration of an expectation say, it must be the case that most agents *under favorable conditions* would frustrate the expectation. The "favorable conditions" comprise the agent's cooperativeness, lack of extraneous pressures, etc. In order for a desire to φ to count as being systematically correlated with the frustration of an expectation not to φ, say, it would have to be the case that in situations where agents are cooperative, under no pressures, etc., they would systematically satisfy the desire rather than the expectation. It is plausible to think that this would be satisfied for some very strong desires, paradigmatically for visceral desires like thirst or hunger. It is less vivid with respect to other desires to walk on the beach, to climb Mount Everest, to do the most outrageous thing one can think of in a public place, to vote against one's convictions, etc. Such desires frequently do not lead to their fulfillment.

Finally, one could try to employ Principle I (see p. 110, above) to argue that desires ought not to qualify as defeating conditions. This is because to the extent that

---

[17] This underscores the point that the question of the nature of action and nature of action explanation are different issues.

desires tend to be correlated with the fulfillment of the expectations they justify, the fulfillment in question tends to be agentive.

For now, I will simply assume that non-compulsive desires should not be counted as defeating conditions.

B. Defeating Conditions and Frankfurt-Type Cases

Frankfurt-type cases purport to illustrate that there are situations where we would hold an agent responsible despite the fact that he could not have *done* otherwise.[18] Consider the following case: Jones decides to kill the mayor of the town. He carries out his plan to the letter, shoots the mayor who dies as a result. Unbeknownst to Jones, evil scientists have implanted a device into Jones' brain which, were Jones to decide not to kill the mayor (or waver after his decision), would have swayed Jones to kill the mayor anyway. The intuitions about cases of this sort have been almost uniform. Jones is responsible for killing the mayor. At the same time, it has been claimed, Jones could not have *done* otherwise: he could not have not killed the mayor (see Figure 2). The question for us is first of all whether the presence of the counterfactual intervener functions as a defeating condition in this case. I will argue that it does not.

The structure of these cases can be captured as follows. In the ordinary case (without the counterfactual intervener), we can suppose that it would be both reasonable$_A$ to expect of Jones that he kill the mayor and reasonable$_A$ to expect of Jones that he not kill the mayor. Does the presence of the counterfactual intervener render it unreasonable$_A$ to expect of Jones that he kill the mayor? One might think that it does. After all, given the presence of the counterfactual intervener it is determined that the mayor will die at Jones' hands. It would thus seem that the presence of the counterfactual intervener is systematically correlated with the pf-fulfillment of the expectation that Jones kill the mayor. However, there are good reasons not to treat the

---

[18] Belnap and Perloff ("Seeing to It that: A Canonical Form for Agentives," in (eds.) H.E. Kyburg, Jr., R.P. Loui, G.N. Carlson, *Knowledge Representation and Defeasible Reasoning* [Dordrecht: Kluwer, 1990], pp. 175-199.) point out that there are two interpretations of the phrase "could have done otherwise." On the stronger, to say that α, who φed, could have done otherwise is to say that it was possible that α see to it that α not φ. On the weaker: it is to say that it was possible that it was not the case that α see to it that α φ. Frankfurt-type cases are directed against the stronger interpretation of the phrase.

presence of the counterfactual intervener as a defeating condition. There are at least two ways to argue for this conclusion.

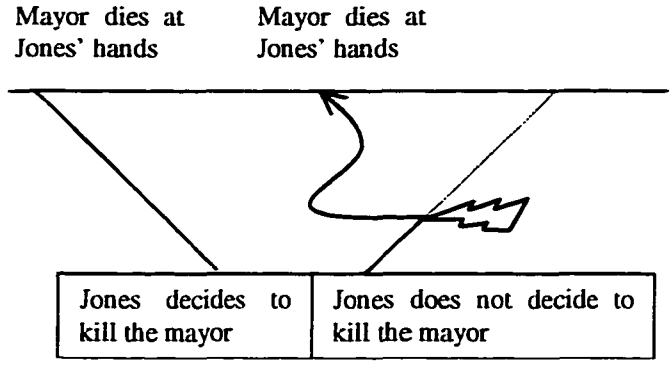


Figure 2. The structure of Frankfurt-type cases.

The first way to argue that the presence of the counterfactual intervener does not defeat the reasonableness$_A$ of holding Jones to the expectation that he kill the mayor is similar to the impact the case has had on the literature of the subject. The lesson that is sometimes drawn from Frankfurt's cases is that they show that our conception of conditions of responsibility is based on what *actually* happens rather than on what might happen.[19] Indeed as the case is described the counterfactual intervener does not affect the course of events in the actual sequence. This is different for the possible sequence. Were his intervention to occur, it would make it unreasonable$_A$ to expect of the agent that he kill the mayor. Insofar as it is the supposition of the examples that the counterfactual intervener will not intervene, we appear to have no reason for thinking that it would be unreasonable$_A$ to expect of the agent that he kill the mayor.

---

[19] The most prominent representatives of this actual-sequence approach to responsibility are Harry G. Frankfurt, *The Importance of What We Care About. Philosophical Essays* (Cambridge: Cambridge University Press, 1988); John Martin Fischer, "Responsiveness and Moral Responsibility," in (ed.) Ferdinand Schoeman, *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987), pp. 81-106 and John Martin Fischer, *The Metaphysics of Free Will. An Essay on Control* [Oxford: Basil Blackwell, 1994].

One might object to this response that it presupposes that we think of defeating conditions as causes of mere happenings, which is what I decided not to require (section 2.D). Moreover, it might be argued that what it shows is that the interference by the counterfactual intervener does not defeat the reasonableness$_A$ of holding Jones to the expectation to kill the mayor in the actual case, but it would in the possible case. But the original question was not whether the interference (*C*) is a defeating condition but rather whether the presence of the counterfactual intervener is (*K*-or-*C*). After all, given the fact that the counterfactual intervener is present, it is settled that the mayor will die at Jones' hands. The only way to dismiss this alleged defeating condition as bogus, the objection continues, would be to suggest that it is not a causal condition, while the interference by the counterfactual intervener is. As I explained in section 2.D, the issue of deciding what is and what is not a causal condition is rather delicate. Rather than offering an account of the matter, I propose an alternative (though ultimately not unrelated) explanation why the presence of the counterfactual intervener is not a defeating condition.

There are exactly two avenues to the mayor's death at Jones' hands envisaged in the example. First, Jones might decide to kill the mayor (*K*) and so kill him. Second, Jones might not decide to kill the mayor, in which case the counterfactual intervener will take over (*C*), leading Jones to kill the mayor. In the first case, the expectation is agentively fulfilled, in the second case it is also fulfilled but non-agentively. The case is constructed so that either *K* or *C* occurs (this is what the presence of the counterfactual intervener amounts to). Given the presence of the counterfactual intervener (*K*-or-*C*), the expectation that Jones kill the mayor will be prima facie fulfilled. Since the correlation is not with exclusively agentive fulfillment, *K*-or-*C* does not violate Principle I. Principle II would exclude *K*-or-*C* if *K* and *C* were themselves defeating conditions. While *C* is a defeating condition, *K* is not (see section A, above, and section 4, below).

However, the case does violate Principle III. What is special about the example is the fact that two conditions are identified, either one or the other occurs, and both of them are systematically correlated with the pf-fulfillment of the expectation to kill the mayor. It follows from Principle III that the condition *K*-or-*C* is not a defeating condition with respect to the expectation that Jones kill the mayor. The condition is not a new defeating

condition, for it relies on the disjoining of systematic correlations we have a good understanding of. Hence it is reasonable$_A$ to expect of Jones that he kill the mayor.

In this case, the presence of the counterfactual intervener does not make unreasonable$_A$ either the expectation that Jones kill the mayor or the expectation that Jones not kill the mayor. The presence of the counterfactual intervener is not properly construed as a defeating condition. However, the actual intervention by the scientist would be construed as a defeating condition, were it to occur.

In Appendix A, I shall discuss how this approach can be used to shed some light on the debate concerning the so-called asymmetry thesis.

C. Unintentional Omissions

I want to close this section by noting that the account thus far is too poor to capture the concept of reasonableness$_A$. This is best illustrated with respect to unintentional omissions.

Here is a familiar scenario. An employee is expected to be at a meeting at 9am, but he oversleeps. As indicated, I want to insist that despite the fact that the agent is sleeping at 9am, it would still be reasonable$_A$ to expect of him that he be at the meeting. Yet this is not the result that the concept of reasonableness$_A$ thus far developed yields. Surely being asleep at the time one is expected to be at the meeting is systematically correlated with the frustration of the expectation to be at the meeting. It would thus appear that being asleep is a defeating condition with respect to the expectation to be at the meeting and so renders the expectation unreasonable$_A$. We will see in the next section how to avoid this conclusion.

4. Defeating Defeating Conditions

So far I have adopted a relatively straightforward characterization of defeating conditions as those conditions that are systematically correlated with the frustration or fulfillment of an expectation. I have suggested that this idea of defeating conditions constitutes a way of delimiting our understanding of what it means to say that it is within the agent's power to do something. But there is a complication.

Let us take an expectation of α that α φ. Let us suppose that *C* is systematically correlated with the frustration of the expectation to φ. Intuitively, when *C* occurs it is not "within the agent's power" to φ. A question that might be reasonably raised is: Is it "within the agent's power" to see to it that *C* does not occur?

It seems clear that there are such cases. It is a well-known fact that drinking an immoderate amount of alcohol will reliably result in a loss of much control: it is systematically correlated with the frustration of a range of expectations (to drive safely, to behave responsibly, etc.). According to the account so far, given that a person has ingested an immoderate amount of alcohol, it will be unreasonable$_A$ to expect her to drive safely, or to behave responsibly. So, looking forward a little, if she drives unsafely it will not be a breach of expectation, it will not be something she did. Nor will it count as her doing it if she abuses someone while drunk. But the account is too simple-minded. We need to add a normative condition on what counts as a defeating condition, by allowing for the possibility of there being circumstances where the defeating character of a defeating condition is itself defeated.

Let *C* be systematically correlated with the frustration of the expectation to φ. The defeating character of *C* will be itself defeated if it is reasonable$_A$ to expect of the agent that she bring it about that *C* does not occur. We can thus enrich our characterization of reasonableness$_A$.

> It is reasonable$_A$ to expect of α that α φ if and only if either (a) no defeating condition (with respect to the expectation to φ) occurred, or (b) such a defeating condition did occur but it was countered by an appropriate counterdefeating condition, or (c) such a defeating condition did occur and it was unreasonable$_A$ to expect of α that α bring it about that it not occur.

Let us note that according to this characterization, conditions that are systematically correlated with the frustration of an expectation to φ will not defeat that expectation's reasonableness$_A$ *unless* it is also unreasonable$_A$ to expect the agent to prevent them from occurring. So, suppose that α is reasonably$_A$ expected to φ and a condition *C* occurs which is systematically correlated with the frustration of the expectation to φ. Suppose

further that it is reasonable$_A$ to expect of the agent that *C* not occur. It follows that the expectation of α to φ is reasonable$_A$; its reasonableness$_A$ has not been defeated by *C*.

Let us consider two examples to illustrate the point. Let us first take the example mentioned above. A person is reasonably$_A$ (and surely reasonably$_N$) expected to drive safely (φ). She is at a party and has far too much to drink (*C*). It is well known that a state of drunkenness systematically interferes with agents' fulfillment of the expectation to drive safely. However, it is also reasonable$_A$ to expect the agent not to drink too much. If this is so then it is also reasonable$_A$ to expect of her that she drive safely despite the fact that she is no longer in a state to fulfill the expectation.

An employee is reasonably$_A$ (and reasonably$_N$) expected to be at an important meeting (φ). He oversleeps (*C*). His oversleeping makes it impossible for him to be at the meeting, and would thus appear to make the expectation that he be there unreasonable$_A$. However, it is reasonable$_A$ to expect of him that he not oversleep. (And in support of our so thinking we cite the fact that the agent can put alarm-clocks all around him, alert his neighbors, go to bed early, etc.) If so then (in absence of further conditions, to be explained below) it straightforwardly follows on our account that despite the fact that while the agent is asleep it is not "within his power" to be at the meeting, it is still reasonable$_A$ to expect of him that he be there. The fact that he oversleeps does not defeat the reasonableness$_A$ of the expectation that he be at the meeting because it is reasonable$_A$ to expect of the agent that he not oversleep.

As Figure 3 shows, the structure is indeed complex. We can see a further complication arising if we consider the case of the employee who oversleeps and so fails to come to the meeting but who oversleeps because he has been drugged. In such a case, our intuitive attitude toward such a person changes. We are right in believing that it is not within the agent's power to come: it is unreasonable$_A$ to expect the agent to be at the meeting.

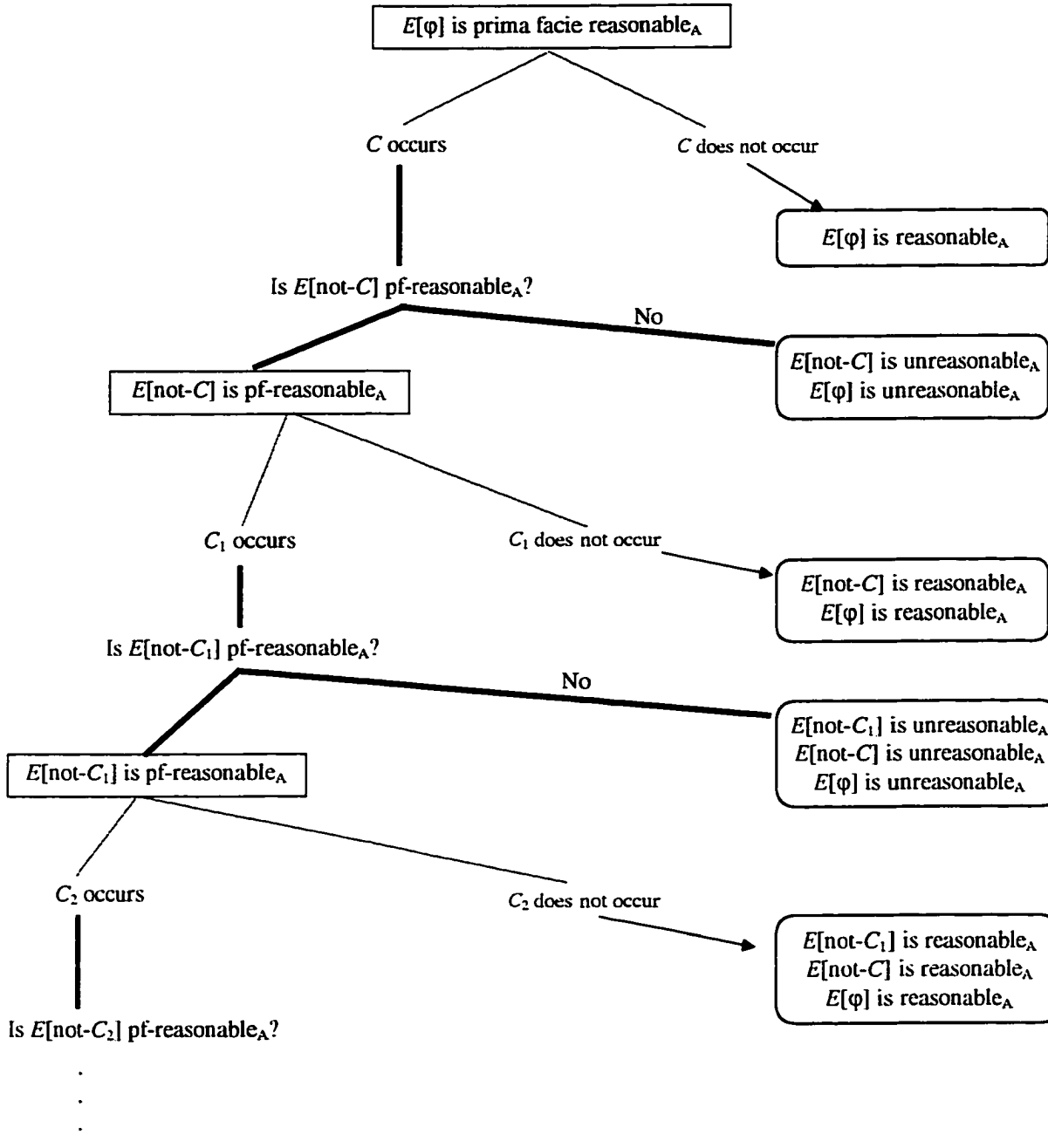


Figure 3. The structure of dependence of reasonableness$_A$ of expectations on further defeating conditions. Condition C is a defeating condition with respect to the expectation to φ (E[φ]); C_1 is a defeating condition with respect to the expectation to prevent C from occurring (E[not-C]); C_2 is a defeating condition with respect to the expectation to prevent C_1 from occurring (E[not-C_1]).

The reason why such a complication is possible is this. For any potential defeating condition (with respect to φing), a condition that is systematically correlated with the frustration of an expectation to φ, it must be ascertained whether or not it is reasonable$_A$ to hold the agent to the expectation that he prevent the condition from occurring. Assuming that it is reasonable$_A$ to expect of the agent that he prevent the condition from occurring, we have not one but two expectations in play. And just as there are possible defeating conditions to the first expectation, so there are defeating conditions to the latter.

Abstractly, let us assume that it is reasonable$_A$ to expect of α that α φ. *C* occurs. *C* is systematically correlated with the frustration of the expectation to φ. However, *C* does not defeat the reasonableness$_A$ of the expectation to φ because it is reasonable$_A$ to expect of the agent that he prevent *C* from occurring. This second expectation (that the agent prevent *C* from occurring) also has potential defeating conditions, however. Conditions C_1 and C_2 are systematically correlated with the frustration of the expectation that he prevent *C* from occurring. However, it is unreasonable$_A$ to expect of the agent that he prevent C_1 from occurring, but it is reasonable$_A$ to expect of the agent that he prevent C_2 from occurring.

Suppose first that C_1 occurs. Since it is unreasonable$_A$ to expect of the agent that C_1 not occur, C_1 defeats the reasonableness$_A$ of the expectation that *C* not occur. We will remember, however, that the reason why *C* did not defeat the reasonableness$_A$ of the original expectation that α φ was that it was reasonable$_A$ to expect of α that *C* not occur. Now, however, condition C_1 has defeated the reasonableness$_A$ of the expectation that *C* not occur. As a result, the occurrence of *C* defeats the reasonableness$_A$ of the expectation of α that α φ.

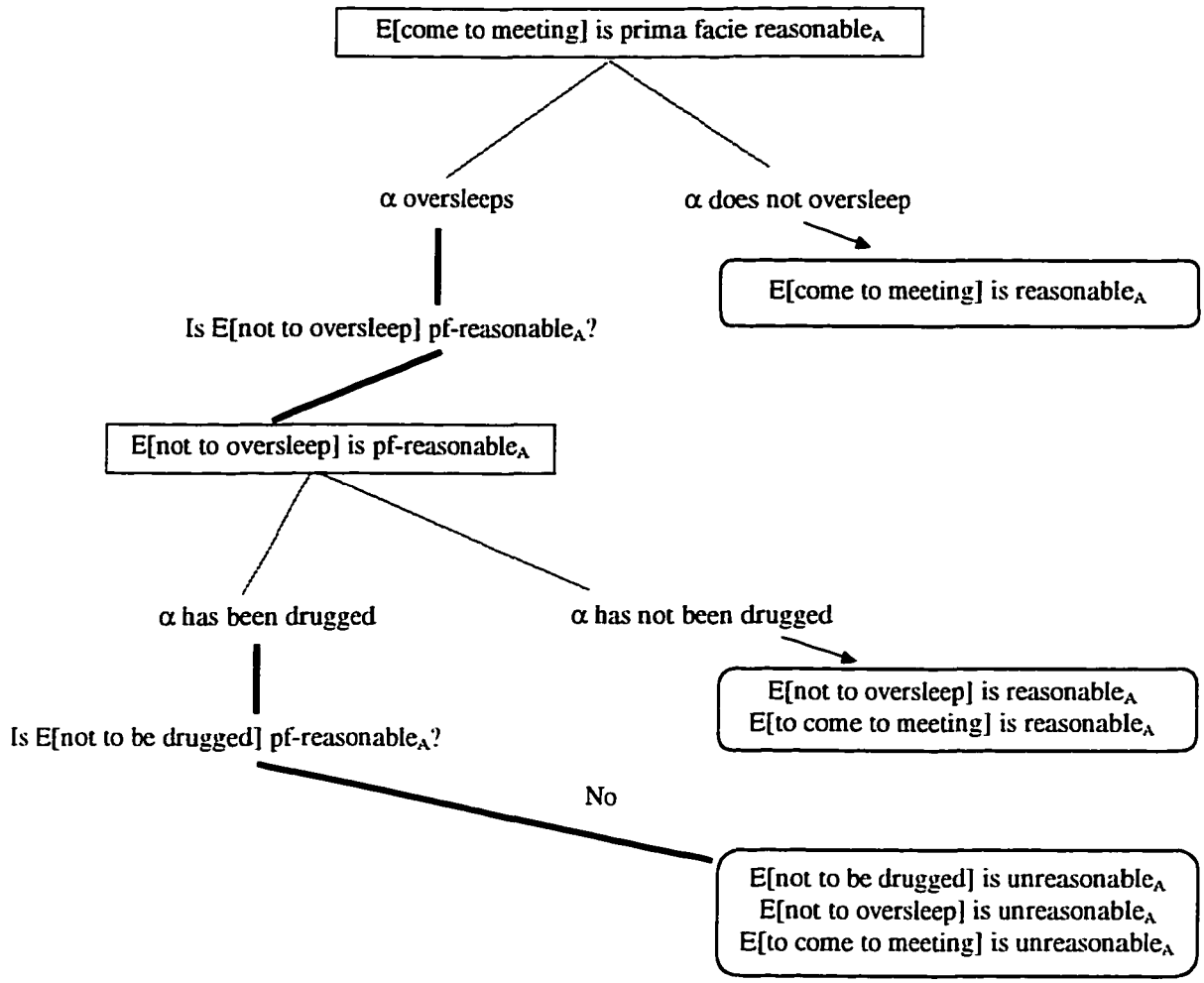


Figure 4. The structure of dependence of reasonableness$_A$ of the expectation to come to the meeting on the conditions that the agent overslept and that he overslept because he has been drugged.

Although complicated, this is the structure exhibited by the case of the employee not coming to the meeting (not φing) because he has overslept (C) as a result of being drugged (C_1) (see Figure 4). It is prima facie reasonable$_A$ to expect the employee to come to the meeting (φ). He oversleeps (C occurs). Oversleeping is systematically correlated with the frustration of the expectation to come to meetings. However, it is prima facie reasonable$_A$ to expect of the agent that C does not occur (not to oversleep). The agent has been drugged, as a result of which he oversleeps (C_1 occurs). Being drugged is systematically correlated with the frustration of the expectation not to oversleep. And, it is not prima facie reasonable$_A$ to expect of the agent that he not be drugged. So, the agent's being drugged defeats the reasonableness$_A$ of the expectation that he not oversleep. Since now it is no longer reasonable$_A$ to expect of the agent that he not oversleep, his oversleeping does defeat the reasonableness$_A$ of the expectation to come to the meeting. In other words, the expectation to come to the meeting is no longer reasonable$_A$ in view of the fact that the agent has overslept as a result of being drugged.

The well-like character of defeating conditions (captured in Figure 4), which may be subject to defeat by further conditions, is responsible for much of the open-ended nature of our attribution of actions. One way in which this feature has been manifest in the literature is by the necessity of introducing the open-ended qualifier "in the right way."[20] It is also sometimes captured by the introduction of the standard of *due care*. In the above terms, the standard of due care (relative to a certain expectation) comprises all those conditions (systematically correlated with the frustration of the expectation) where it is prima facie reasonable$_A$ to expect of the agent that they not occur (and so not interfere with the fulfillment of the expectation).

Some Desires Render Expectations Unreasonable$_A$, Others do Not. I have already answered one of the objections set out at the end of the last section. I have shown that sometimes when a person unintentionally omits to do something it would be still reasonable$_A$ to expect of him that he do it. Let me answer the second one. The objection

---

[20] Donald Davidson, "Freedom to Act," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 63-81. The context of Davidson's discussion may appear different because it is introduced with respect to the causal theory of action. But that it is not as different as it might appear will become clear.

was that desires, which are systematically correlated with the fulfillment of the expectations they justify, would count as defeating conditions and hence render the expectations unreasonable$_A$.

We can now see the limitations of that objection. It will be indeed the case that a desire to φ, if it is systematically correlated with the fulfillment of the expectation to φ, renders the expectation unreasonable$_A$ if it is unreasonable$_A$ to expect of the agent that she prevent her desire from affecting her action. It is arguable that there are some desires like that. Compulsive desires, for instance, are desires that we would intuitively think beyond the agent's power to control. The compulsive-obsessive's desire to wash his hands every five minutes is not something that he can control. But this is not so for most other desires. Even if someone's desire for chocolate is very strong, so whenever he has it he submits to it and eats chocolate, it would (or at least might) be reasonable$_A$ to expect him not to eat the chocolate. It might be reasonable$_A$ to expect him to control the desire. Let me illustrate with a somewhat too graphic example.

Let us stipulate that there is a pretty much stable pattern for a particular type of chocoholic. He occasionally gets an urge to eat chocolate, which stimulates his thinking about it, which further strengthens his desire for chocolate, and so on. Finally, the desire becomes strong enough to move him to search for chocolate. Once the chocolate is in his sight, only force could prevent him from eating it. The piece of chocolate is doomed, he can do nothing about it.

Most desires do not function like that. The connection between the desire and the action is usually not so strong. In fact, it is intuitively implausible to think that there is any single type of mental state that is so strongly tied to action. But even in this scenario, where it is clear that not so much the desire on its own but the desire together with the sight of chocolate is systematically (inescapably) correlated with the fulfillment of the expectation to eat chocolate, it would be reasonable$_A$ to expect the agent not to eat the chocolate. Why? Because it would be reasonable$_A$ to expect the agent to prevent himself from ever getting to the stage where he sees the chocolate, which overwhelms him. It would be reasonable$_A$ to expect him to counter the thoughts about chocolate with thoughts about an experiment he should conduct instead, for example.

It is, of course, possible that there are desires that the agent cannot control in such a manner. A compulsive desire to wash one's hands every five minutes may be systematically correlated with the fulfillment of the expectation to wash one's hands. Yet, it may be that the agent can do nothing to prevent the desire from leading to action, i.e. that all attempts to counter the desire (whether by thinking other thoughts or by engaging in other activities) may be systematically frustrated.

• • • • •

I have suggested in Chapter II that one of the basic commitments of a responsibility-based approach to action is to develop an account of practical responsibility that would be significantly different from moral and legal responsibility. It is in part because H.L.A. Hart did not offer such an account that his theory has been subjected to sharp criticism, which pertained not only to the details but to the very core of his account. One of the fundamental charges that responsibility-based accounts of action face is the fundamental problem: the objection that such accounts rely on an initial mistake — they take the concept of responsibility to characterize the logically prior concept of action. In Chapter II, I have promised to develop a concept of practical responsibility that would be immune to the fundamental objection. The account is now complete.

I have argued that

> an agent α is practically (task-)responsible for φing if and only if it would be reasonable$_A$ to expect of α that α φ.

In Chapter III, three tasks have been accomplished. First, I have distinguished between normative and predictive expectations (between what it means to expect of α that α φ and to expect that α will φ). Second, I have indicated that despite the fact that what complements the normative expectation appears to be an agentive statement ('α φs'), this does not necessarily mean that the account that appeals to normative expectations falls prey to the fundamental problem. For I have declared that in the first instance, we shall assume that an expectation of α that α φ will be fulfilled not only by α's actions but, more generally, by α's performances (which include actions and mere happenings). Third, I have argued that we ought to focus on practical normative expectations, not on

specifically moral (with a moral justification) or legal (with a legal justification) ones. In this way, the concept of responsibility is broader than its specifically moral or legal counterparts.

In Chapter IV, I have then undertaken the task of characterizing the sense in which normative expectations must be reasonable. I have distinguished two senses of reasonableness: agent-centered reasonableness$_A$ and specifically normative reasonableness$_N$. Roughly, expectations are reasonable$_N$ if there are good reasons for holding the agent to them; expectations are reasonable$_A$ if it is "within the agent's power" to do what is expected of him. I have not attempted to analyze the concept of reasonableness$_N$, for as we will see in the next chapter, this concept is less important to the distinction between actions and mere happenings.

In the present chapter, I have proposed an account of reasonableness$_A$, which is meant to elucidate the meaning of the metaphor of what is "within the agent's power." I have argued that those normative expectations that are not unreasonable$_A$ are reasonable$_A$. Normative expectations are made unreasonable$_A$ by the occurrence of defeating conditions, conditions that are systematically correlated either with the fulfillment or the frustration of a given expectation (sections 1-2). I have also pointed out that a defeating condition can be itself defeated if it is reasonable$_A$ to expect of the agent that she prevent the defeating condition from occurring (section 4).

CHAPTER VI.

ACTIONS, OMISSIONS, AND MERE HAPPENINGS

Chapters III-V have delineated the concept of practical task-responsibility. The present chapter will discharge the main task of the dissertation and show how to distinguish between actions and non-actions (mere happenings). My main aim, in other words, is to answer Wittgenstein's question: What is the difference between my raising my arm and my arm rising?

I begin with a preview of the answer (section 1). In particular, I shall contrast my approach to the most popular, intentionalist, approach to the question. (In Appendix B, I explain why someone committed to giving an account of action as conduct should reject the intentionalist view.) Section 2 distinguishes two senses of the question "What has an agent done?". Section 3 gives an account of the answers to the first sense of the question. Section 4 gives an account of the answers to the second sense of the question, thereby grounding the distinction between actions and mere happenings. In section 5, I show how the account handles one type of wayward causal chains problems.

1. A Preview

It may help the reader to be given an introduction to the purpose of the present chapter in relation to one aspect of the Anscombe-Davidson intentionalist account.[1] I will focus on Davidson's view. One of the virtues of his account is that it sharply

---

[1] Though many philosophers hold the view, two deserve special mention: G.E.M. Anscombe, *Intention* (Ithaca: Cornell University Press, 1957); Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980). I should point out that what I mean by 'intentionalism' is captured by (I) below.

distinguishes between actions, understood as particular events, and action descriptions under which actions may be intentional or not. On Davidson's view, an action can be intentional only relative to a description. He follows Anscombe in thinking that actions are intentional under at least one description.

(I) An event is an action if and only if it is intentional under some description.

But an action (intentional under some description) can be described in a multitude of other ways. As long as the descriptions are true of the action, they specify something the agent did. In a slogan, an action is intentional under some description and an action under all of them.

(D) For any description *d*, if it is a description of an event that is an action, it specifies something the agent did (intentionally or unintentionally).

On this account, the notion of "doing something" is purely extensional, in contrast to the concepts of doing something intentionally or of doing something unintentionally, which are sensitive to descriptions.

(I) allows us to capture the distinction between actions and mere happenings. Mere happenings are events that are not intentional under any description. When a spasm makes an arm twist as a result of which a cup of tea falls, the event is not intentional under any description. There is nothing the agent was doing intentionally. By contrast, when the agent reaches for some sugar and knocks a cup of tea on the way, the agent did something. He did something because he did something intentionally: he intentionally reached for some sugar. But he also *did* knock a cup of tea, though not intentionally. And indeed there are countlessly many things the agent did. As Davidson notes, we frequently describe actions in terms of their consequences. If the falling cup of tea caused the carpet to be ruined, which caused the hostess to be upset, then we can describe

---

There are of course enormous differences in how the notion of being intentional under a description is understood. There are causal and non-causal interpretations of the concept.

the action as upsetting the hostess. Upsetting the hostess is also something the agent did. Indeed, any description of the action is a specification of something the agent did.

I will argue that there is a group of intuitions according to which we do not uniformly allow just any description of an action to be a specification of something the agent *did*. This is not tantamount to suggesting that actions are not particulars. Rather it can be taken to show that besides the intensional notions of "doing something intentionally" and "doing something unintentionally," there is also the notion of "doing something" that is sensitive to description. I shall argue for the intuitive plausibility of this claim in section 2, where we will see, for instance, that when actions are described in terms of very long-term or accidental consequences, there is a sense of the judgment that the agent *did* it, that we are prepared to withhold. (I will not argue that it is illegitimate to say that the agent did it, but only that it is illegitimate in one sense of the notion.) I will distinguish what an agent *did* (the narrower sense of 'do') from what the agent *happened to do* (corresponding to the remainder of the wider sense of 'do').

In other words, I shall reject (D). Some descriptions of an action specify what the agent did, others specify what the agent merely happened to do. The object of section 3 is to argue that this distinction can be captured in terms of the concept of reasonableness$_A$.

(a) If performance *p* pf-fulfills the expectation of α that α φ, then α φed (i.e. φing is something he did, not something he happened to do) just in case it was reasonable$_A$ to expect of α that α φ.

(b) If performance *p* pf-fulfills the expectation of α that α φ, then α happened to φ (i.e. φing is something he happened to do, but not something he did) just in case it was unreasonable$_A$ to expect of α that α φ.

The distinction between what the agent did and what the agent happened to do is a distinction at the level of action descriptions, it is not a distinction between events that are actions and events that are mere happenings. In section 4, I will show how to use the narrower notion of doing something to give an account of the distinction. I shall argue that:

(H) A performance *p* is a mere happening if and only if for every φ such that *p* pf-fulfills the expectation of α that α φ, it was unreasonable$_A$ to expect of the agent that she φ.

(A) A performance *p* is an action if and only if for some φ such that *p* pf-fulfills the expectation of α that α φ, it was reasonable$_A$ to expect of the agent that she φ.

(A) bears a striking resemblance to (I), except that the notion of being intentional under a description is replaced by the notion of doing something (not merely happening to do something) under a description.

I should note that I do not undertake the task of explaining the notion of being intentional under a description. The concept has turned out to be very hard to capture. An approach to answering the Wittgensteinian challenge (what is the distinction between an action and a mere happening) that does not appeal to the idea of being intentional under a description might, for that reason, be welcome.

2. What Has Been Done: Two Senses of the Question

The question "What has been done?", in contrast to the question "What has been done intentionally?", has not been given too much attention in the literature. It is generally, though not universally, assumed that the question admits of a rather straightforward answer. The answer to the question is given by giving a true description of the agent's action.[2] To give a true description of the event that is the agent's action, is indeed to answer the question "What has the agent done?" in one sense. But there is a

---

[2] The adherents to this view include: Donald Davidson, "Agency," in *Essays on Actions and Events, op. cit.*, pp. 43-61; Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980). Dissenters fall into two categories: those who think that either there are limits on the descriptions that could be legitimately given in answer to the question or who think that there are two senses of the question. See e.g. John R. Searle, *Intentionality. An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983); Gilbert Harman, "Practical Reasoning," *The Review of Metaphysics* 29 (1976), 431-463. And there are those who think that the only answer to the question what has been done is given by the stricter interpretation of the question. Nuel Belnap, Michael Perloff, "Seeing to It that: A Canonical Form for Agentives," in (eds.) H.E. Kyburg, Jr., R.P. Loui, G.N. Carlson, *Knowledge Representation and Defeasible Reasoning* (Dordrecht: Kluwer, 1990), pp. 175-199. Nuel Belnap, "Before Refraining Concepts for

narrower sense of the question as well. We shall see that our intuitions indeed pull in two different directions using three kinds of examples: descriptions in terms of consequences, microscopic descriptions and negative descriptions of actions.

Suppose that a person has *won* a lottery. Should we say that this is something he *did*? On some of our intuitions winning a lottery is *not* the sort of thing we can do, it is more or less an accidental event, over which we have little control. Of course, if our control were to be increased (for instance, by rigging the lottery mechanism or buying all tickets), we could describe winning the lottery as an action of the agent, as something the agent has done in a stronger sense.[3]

And yet, there is a weaker sense in which it would be unobjectionable to say that what the agent did was win the lottery. One might want to argue that the agent did after all buy a ticket and that led to the unlikely consequence of his winning the lottery. But he bought the winning ticket; and to buy a winning ticket in this case is to win the lottery. To say that he won the lottery is to describe something he did, viz. his action of buying the ticket.

But the same bifurcation of intuitions arises in general for descriptions in terms of consequences provided they are long-term or accidental enough. While we may describe actions in terms of their effects (where the idea of an effect is rather liberal as it is in the case of winning a lottery[4]), we are rather picky about the sorts of effects in terms of which we choose to describe actions. There is a sense in which any one of our actions causally contributes to some distant or unlikely event. Suppose that you have swapped theater seats with some man. Unbeknownst to you the man was sitting just behind his wife. The seats turned out to be rather uncomfortable, you have very long legs and there is not much space for them anyway. As a result of your almost constant rearranging, you kick your neighbor's wife a lot. She happened to be on the verge of a breakdown, and

---

Agency," *Erkenntnis* 34 (1991), 137-169. Annette Baier, "The Search for Basic Actions," *American Philosophical Quarterly* 8 (1971), 161-170.

[3] This example is Harman's, though he describes it in terms of the agent not having won the lottery intentionally (despite trying and succeeding). See G. Harman, "Practical Reasoning," *op. cit.*, p. 433.

[4] To say that an event *c* is a cause of some event *e* is to say that it contributed to the occurrence of *e*, but it is not to say anything about how many other events must have contributed for *e* to occur or even whether *e* would have occurred without *c* (because of overdetermination).

your pounding of her seat was the last straw. When she and her husband go home she erupts. As a result of the fight, her husband does not get a wink of sleep, despite the fact that he has an important business meeting early in the morning. Tired and exhausted, he does a terrible job at the meeting and loses his job. If we ask for a list of things you did that evening, should we include among the things you did the fact that you cost him his job? (If the reader's intuitions are still positive, one can obviously go further in the chain of effects.)

Once again, it seems, we could be convinced that such a description indeed specifies something you have done. But this involves weakening our intuitive standards significantly. To say this is not to say that answering the question in this weaker way is illegitimate. Such a weakening of our standards does get a grip on some of our intuitions. But it is also important to recognize that it violates some others. One explanation of what happens is that there are simply two senses of question "What has the agent *done*?".

Consider yet another group of examples which support this conclusion. Ordinarily, we would not include among the things you have done when you raised your arm the fact that you have thrashed some air molecules about. In fact, it may be that you thrashed some particular air molecule off its course by a certain distance. Is this something that we would include in our answer to the question what you have done? Again, the answer depends on how we interpret the question. There is a weaker sense of the question, in which it is something you have done. After all, you raised your arm and your raising your arm was, on this occasion, identical to your moving the air molecule off its course. But there is a stronger sense of the question, in which this cannot be said to be something *you* have *done*. And not just because the description was not available to you. If you were given sophisticated equipment to mark out the path of that particular molecule, still we would not count you as having *done* this. Our disposition would change, however, if somehow you acquired a skill of changing the path of the molecule, if changing the path of that molecule in a particular sort of way was "within your control." So, once again, it is reasonable to suppose that the question has two senses.

Another class of cases involves negative descriptions. We select the things the agent has not done (in the narrower sense) from among the things the agent has not done (in the wider sense). As Vermazen has warned:

> Certainly we don't want to say that a person is not-ψ-ing[5] just in case he is not ψ-ing. ...It won't help much to add the rider "if the agent is doing something" to this last, since the agent will then be doing far too many negative acts: Andy, as he sits twisting his buttons, would also be not-sweeping the table clear of canapés, not-preparing for a Channel swim, not-attempting to cross the Sino-Soviet border, and so on.[6]

With respect to negative descriptions, then, once again, it seems that we do exercise some discretion with respect to those descriptions that we would give in answer to the stricter sense of the question "What did the agent do?" and those that we would not.

Speaking of the narrower and the wider sense of 'do' is awkward. Let me stipulate a better way of putting the distinction. When an agent did something in the narrower sense of 'do' I shall simply say that that the agent did it. Occasionally, I will add disambiguating clauses like 'rather than happened to do it' or '(in the narrower sense)'. When an agent did something in the wider sense of 'do' but not in the narrower sense of 'do', I shall say that the agent happened to do it. Henceforth:

> 'The agent φed' stands for 'The agent φed (in the narrower sense)'.
>
> 'The agent happened to φ' stands for 'The agent φed (in the wider sense) but did not φ (in the narrower sense)'.

Although I do believe that the terminology is suggestive and has some intuitive base, it is quite sufficient if it is simply understood by the reader as a terminological convention. Thus, I will say that winning the lottery is something the agent happened to do rather than something he did; that dislocating some particular air molecule off its course by a certain distance is something the agent happened to do; that not preparing for a Channel swim is something Andy, twisting his buttons, only happened to do.

In sum, these examples suggest that two kinds of intuitions are available to us. First, in each case, it seems intuitively plausible to give a narrower list of actions in answer to the question, "What has the agent done?" Those intuitions support a narrower interpretation of the question. Second, in each case, however, the appeal to the simplicity

---

[5] Vermazen adopts the convention of inserting a hyphen between 'not' and the action-verb in order to mark that the description specifies a negative action.

of something like a Davidsonian picture does seem to release another group of intuitions that support the wider interpretation of the question. I do not claim that the cases cited are the only kinds of cases where the divide becomes apparent. Nor do I believe that it is evident that our intuitions split uniformly. All that can be claimed at this point is that the examples and the intuitions they spark do not render implausible the hypothesis that there are two senses of 'do' or of 'the agent did something'.

Before closing and going on to attempt to capture the narrower sense of 'doing', let us consider another possible understanding of the situation. One might claim that the suggestion that there are two senses of 'do' is mistaken. The only sense of 'do' we have is the wider one, the alleged narrower sense of 'do' is merely a result of various pragmatic restrictions to which we succumb. One argument that might support such a position is that we do not have a clear understanding of the alleged narrower sense of 'do' while we do have an understanding of the wider sense. In what follows, I shall show that we can obtain an understanding of what it is to do something in the narrower sense by appealing to what it would be reasonable$_A$ to expect of an agent.

3. What Has Been Done?

One philosopher who has advanced a view based on similar intuitions is Annette Baier. I briefly discuss her view in section A. Section B shows how to cast these thoughts in terms of the concept of reasonable$_A$ expectations. Section C considers an objection to the proposal, while D applies the apparatus to a problem.

A. Doings and Tasks

Annette Baier is concerned with specifying the things we do when we perform an action. She agrees with Anscombe and Davidson that whenever there is an action there is some description of it under which it is intentional. Given that this is so,

---

[6] Bruce Vermazen, "Negative Acts," in (eds.) Bruce Vermazen, Merrill B. Hintikka, *Essays on Davidson* (Oxford: Clarendon, 1985), p. 96.

> the next problem comes when we attempt to decide which of the many things the agent does during the intentional action are also to count as actions, intentional or unintentional.[7]

It should be clear that Baier thereby attempts to give an answer to the narrower question. After all, the answer to the wider question is immediate: any description of the event that is the agent's action is something the agent does.

Baier proceeds to suggest that what actions the agent performs (what the agent does in the narrower sense) ought to be construed in terms of the tasks the agent could be thought to accomplish:

> we [can] define action as anything we do ... and can be known to have done, which might be the correct response to an order, instruction, or task-specification, usually a self-imposed one.[8]

The crucial feature of the performance of a task in contrast to merely purposive behavior is the fact that in performing a task, one's performance is subject to "public standards of success or correctness."[9]

Tasks are subject to what Baier calls a double monitorability requirement. We do not set tasks whose performance we could not check. Nor do we set tasks whose performance cannot be checked by the agent. She argues that "what counts as an action will ...be relative to the normal capability for monitoring."[10] If an agent's capacity for monitoring is larger than normal, she will be able to perform more tasks. For instance, if the agent is able to tell when she fired a particular neuron, then "it will be for [her] an action, but since this capability is not generally shared, it will not... be an action whoever does it."[11] However, if the agent's capacity for monitoring is smaller than normal, this will not mean that she performs only those tasks that she is able to monitor, though it may mean that it will narrow the tasks that the agent is able to perform intentionally.

While monitorability is indeed built into the concept of a task, more seems to be at work than just monitorability. It seems that something like the idea of it being "within

---

[7] A. Baier, "The Search for Basic Actions," *op. cit.*, p. 163.
[8] Ibid., p. 163.
[9] Ibid., p. 163.
[10] Ibid., p. 164.
[11] Ibid., p. 164.

the agent's power" to do something, or of it being reasonable$_A$ to expect the agent to carry out the task, is at least as close to the idea of a task. After all, we would generally consider it to be inappropriate to assign someone a task that we know he cannot accomplish, the accomplishment of which he could nonetheless monitor.[12]

Baier's account already allows us to exclude the microscopic descriptions of an agent's action from specifying things the agent did. They will be excluded because we usually do not have ways of monitoring our activity at a microscopic level. But descriptions in terms of long-range and accidental consequences are no longer as easily excluded. The result of a lottery can surely be monitored by others and by oneself. The same applies to other consequences of one's actions.

This is not to suggest that the concept of a task as understood by Baier does not narrow down the truly wide concept of "doing" or that it does not capture some distinct intuitions pertaining to such a narrowing. All that I mean to suggest is that it does not quite suffice for the task at hand, to narrow the concept of action so as to at any rate exclude the cases mentioned in section 2.

We can extend Annette Baier's suggestion to understand the narrower sense of 'doing' in terms of the tasks that the agent accomplishes. We can do so by exploiting the connection between the idea of a task and the idea of normative expectations, which as I argued in Chapter III ground the notion of practical task-responsibility. More precisely, I will suggest that the narrower sense of 'doing' applies to those performances of an agent that it would be (at least) reasonable$_A$ to expect of him.

B. What Would Be Reasonably$_A$ Expected of the Agent?

The narrower sense of the concept of doing can be fruitfully captured in terms of what it would be reasonable$_A$ to expect of the agent. If an expectation of an agent to perform an action under a description would be unreasonable$_A$ (if it is not "within the

---

[12] The reason why for A. Baier the idea of monitorability stands out in the way in which the idea of it being within the agent's power does not is not only the fact that the latter is rather vague and ambiguous but also the fact that she asks the question "What has the agent done?" only after it is settled that there is something the agent has done intentionally. So, in general, it may seem that since what the agent has done intentionally will have been in the agent's power to do, any task that has been thus accomplished will have been in the agent's power to do also.

agent's power" to perform the expected action), the agent cannot be said to have performed the action under that description. If the expectation is nonetheless fulfilled, it is fulfilled but accidentally: the agent can only be said to have happened to do it. But as long as it is reasonable$_A$ to expect of the agent that he perform the action under a description (as long as it is "within his power") then it is appropriate to say that he performed (rather than happened to perform) the action under that description.

(a) If performance *p* pf-fulfills the expectation of α that α φ, then α φed (i.e. φing is something he did not just something he happened to do) just in case it was reasonable$_A$ to expect of α that α φ.

(h) If performance *p* pf-fulfills the expectation of α that α φ, then α happened to φ (i.e. φing is something he happened to do, but not something he did) just in case it was unreasonable$_A$ to expect of α that α φ.

Let me remind the reader of two points. First, the characterizations use the construct of *it* being reasonable$_A$ to expect something of the agent. To say that it would be reasonable$_A$ to expect something of an agent α is not to imply that some person actually does hold α to this expectation. Rather it is to say that there is some person (possibly α herself) who is such that if she held α to the expectation, the expectation would be reasonable$_A$. In this way, our judgments as to what has been done are not contingent on others' (or the agent's) holding the agent to an expectation. This coincides with the intuition that what the agent does is not determined by a contingent fact about another person's (or the agent's own) attitude toward the agent. Second, to say that it is reasonable$_A$ to expect something of an agent is to say either that no defeating condition occurred or that a defeating condition did occur but was either countered by a counterdefeating condition (special skill) or defeated because it was reasonable$_A$ to expect of the agent that he prevent the defeating condition from occurring.

Let me now demonstrate that (a)-(h) do indeed narrow down the concept of doing so as to exclude the unwelcome descriptions: microscopic descriptions, descriptions in terms of long-term accidental consequences, and some negative descriptions.

An agent raises his arm to vote for a motion thereby changing the path of a particular water molecule. Is his changing the path of the water molecule something he did or something that he happened to do? In order to see that it is something that he happened to do, we need to ask the question whether it would be reasonable$_A$ to expect of the agent that he change the path of the water molecule. It seems intuitively clear that such an expectation would be prima facie unreasonable$_A$. We generally lack the competence to change the paths of water molecules. Agents held to the expectation to change the path of the water molecule in a particular way would systematically frustrate it. Such an expectation is prima facie unreasonable$_A$ and it is not countered by a counterdefeating condition.[13] It would be countered by such a condition if the agent acquired a skill to reliably change the path of the water molecule. Then it would be reasonable$_A$ to expect of him that he change the path of the water molecule in a particular way, and his doing so would be something he did rather than something he happened to do.

The same would apply if one expected of the agent that he perform the voting motion under its complete microscopic description. For since a voting motion can be realized under a multitude of different microscopic descriptions (as well as macroscopic ones), the expectation that the agent move his arm in such a way as to satisfy a particular microscopic description would be systematically frustrated. Suppose, however, that instead of the complete microscopic description of a particular voting motion we consider a disjunction of complete microscopic descriptions of all possible voting motions. It is arguable that such a description is more than science fiction, but let us grant that it is possible. In this case, it would be reasonable$_A$ to expect of the agent that he perform the action under this extraordinary microscopic description. But this is not in any way objectionable. There is nothing about microscopic descriptions per se that makes them the unlikely candidates for the lists of things we do (in the narrower sense of 'do'). I have already noted that if the agent possessed a special (albeit peculiar) skill to change

---

[13] For simplicity of the overall characterization of reasonableness$_A$, I have suggested that prima facie unreasonableness$_A$ be construed in terms of the presence of a tautological defeating condition (see Chapter V, footnote 11, p. 105). In such a case, the expectation can also not be defeated: it is unreasonable$_A$ to expect of the agent that he prevent *p* or not-*p* from occurring.

the path of the water molecule reliably, there should be no quarrel with our thinking that it is reasonable$_A$ to expect this (under the microscopic description) of that particular agent. His special skill would counterdefeat the prima facie unreasonableness$_A$ of such an expectation. What explains our intuitive unwillingness to include such descriptions among things done is our general unreliability with respect to most of them. It would be inappropriate, however, to insist that the scientist who by means of his sophisticated (and reliable) apparatus arranges a particular molecule in a particular way only happened to arrange it in this way.

In a similar fashion, we can exclude unwelcome descriptions in terms of long-term or accidental consequences from counting as things the agent did (in the narrower sense). The expectation to win the lottery is prima facie unreasonable$_A$. Agents held to such an expectation would systematically frustrate it. The same conclusion can be reached if we think of winning a lottery explicitly as a consequence of the action of buying a ticket. For our limited purposes, let me propose a plausible principle concerning the conditions under which it would be reasonable$_A$ to expect of the agent that he ψ, where ψing is a consequence of his φing.[14] Roughly, it is reasonable$_A$ to expect of α that α ψ, if it is reasonable$_A$ to expect of α that α φ and it is reasonable* to believe[15] that α will ψ if α φs.[16] So in the case of the lottery, given that it is reasonable$_A$ to expect of the

---

[14] I do not offer an account of consequences of action in the dissertation. This is not a trivial enterprise and I only want to signal one of many difficulties here. It may very well be that though by all counts it is within an agent's power to bring about a consequence, as a matter of fact she brings it about by an extremely unlikely chain of events. Suppose that John is a master golf-player. He is almost 100% reliable in striking the hole from 50 feet distance. (In view of his ability, it would be reasonable$_A$ to expect of him that he strike the hole from the distance of 50 feet.) As it happens, he hits the ball, which bounces off three trees before it luckily lands in the hole. In view of the circumstances (in particular, the ball's hitting the trees), it would seem quite unreasonable$_A$ to expect of John that he strike the hole. I think we are inclined to think that it is the second judgment that we should go by. And I think that there are good reasons for this, for the ball's hitting the trees functions here in a way that is similar to a defeating condition. However, it would take a more systematic consideration of consequences to assert the claim with any justification. I leave this as a post-dissertation endeavor.

[15] What is involved here is a predictive expectation (I dispense with the terminology for simplicity). Note also that I have in no way characterized the sense of reasonableness* with respect to beliefs. I will simply rely on the reader's intuitions. It is clear, however, that it is a different notion than either reasonableness$_A$ or reasonableness$_N$. The asterisk is a reminder that a completely different concept is in play here.

[16] This reminds one of the appeal to the notion of foreseeability common in the literature. See, e.g. Gilbert Harman, *Change in View. Principles of Reasoning* (Cambridge, MA: The MIT Press, 1986); Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

agent that he buy a ticket, it will be reasonable$_A$ to expect of him that he win the lottery only if it is reasonable* to believe that when he buys the ticket he will win the lottery. Given the probabilities involved, it would be quite unreasonable* to believe that the agent will win the lottery.

The same reasoning can be repeated for accidental or very long-term consequences of the agent's actions, as long as it is unreasonable* to believe that they will occur given that the agent performs the action. Thus, when one causes a person to lose his job (because one has been kicking the seat of that person's wife in the cinema, which turned out to be the final straw that caused her to break down, thus preventing her husband from preparing for an important meeting, as a result of which he lost his job), it is not something one has done but only something one has happened to do. It would be unreasonable$_A$ to expect of one that one bring it about that he loses his job, for it would be unreasonable* to expect (believe) that one would cause him to lose a job if one kicks his wife's seat in the movies.

Finally, many unwelcome negative descriptions will be also excluded from counting as descriptions of what the agent did in the narrower sense. Consider Vermazen's example. Andy sits in the doctor's office twisting his buttons. He thereby does not sweep the table clear of canapés, does not prepare for a Channel swim, does not attempt to cross the Sino-Soviet border. Given that there are no canapés in the doctor's office, it would be unreasonable$_A$ to expect of Andy that he not clear the table of them (the expectation would be systematically frustrated). Given that Andy is in the doctor's office, it would be unreasonable$_A$ to expect of him that he prepare for a Channel swim (unless perhaps, the doctor's office was rather close to the Channel). For a similar reason, it would be unreasonable$_A$ to expect of him that he not attempt to cross the Sino-Soviet border.

However, this will leave a lot of negative descriptions still eligible as descriptions of what the agent is doing. For example, there are a lot of things that it may be reasonable$_A$ to expect of a president delivering a victory speech. It may be reasonable$_A$ to expect him to repeat his electoral promises. But it may be equally reasonable$_A$ to expect him to turn around every three minutes during the speech. To capture the sense in which more agency is involved in the president's *not* repeating his electoral promises in the

victory speech than in the president's not turning around every three minutes, we need to appeal to the concept of reasonableness$_N$. It is reasonable$_N$ (as well as reasonable$_A$) to expect of the president that he repeat the electoral promise while it is not reasonable$_N$ to expect of the president that he turn around every three minutes. In general, the concept of reasonableness$_N$ may be helpful in arranging the descriptions of what the agent did in the order of their significance.

Although the concept of reasonableness$_A$ of expectations does allow us to restrict many of the negative descriptions from counting as part of what the agent did, it still leaves a lot of negative descriptions. As I am typing these words, it would be reasonable$_A$ to expect of me that I drink my tea, that I make some coffee, that I have a banana, that I knock at the table a couple of times, that I walk about the room, etc. Since I am not doing any of those things, my not doing them will count as something I *did* rather than happened to do. Yet, on most of our intuitions, my not drinking tea, not making coffee, not having a banana, not knocking at the table, not walking about the room, not taking a walk are simply insignificant. On the view proposed here, we can understand our hesitation by employing something like the concept of reasonableness$_N$. The reason why these negative descriptions seem unfit to be listed as among the things the agent has done is the fact that it would not be particularly reasonable$_N$ to expect them of the agent on a particular occasion. If that were to change, however, our assessment concerning the value of their being included would change as well. Thus, if it was reasonable$_N$ to expect of me that I drink the tea, perhaps because my mother made it and she takes great pride in her tea and not drinking it would be an offense to her, then my not drinking the tea would be a description that would be worth listing among the descriptions of things I did. Or, if it were reasonable$_N$ to expect of me that I not walk about the room because it disturbs the neighbor downstairs who is very ill, then again my not walking about the room would be worth mentioning as among the things I did. The reason why we might shrink at the thought that I did (in the narrower sense) so many negative things does not have to do with the fact that I did *do* them, but rather with the

fact that although I did do them they were not significant enough to be listed in most circumstances.[17]

The fact that we impose additional criteria on descriptions of actions worth being listed is not limited to negative actions. As I am typing, I am trying to finish my dissertation, I am moving my fingers in a particular way, I am leaning forward, I am arching my back, I am looking at the computer screen, I am leaning my head a little, etc. These are some of the things that I am doing that it would be reasonable$_A$ to expect of me that I do, and it would seem a waste to include any or most of them in the journal of my daily accomplishments. Once again, we seem to subject the descriptions to another normative standard (such as reasonableness$_N$), which allows us to single out the most important among the descriptions.

One might object at this point that the fact that the narrower concept of "doing" does not sufficiently narrow the descriptions jeopardizes the argument for drawing the distinction between the narrower and the wider concept of "do". For our argument for drawing the distinction relied on the fact that we had two sets of intuitions. Some of our intuitions would dictate that we withdraw the judgment that the agent did something, while others would dictate that we uphold that judgment. I have then argued that we can capture those intuitions suggesting the narrower sense of 'do' by thinking about what it would be reasonable$_A$ to expect of the agent. Now, however, I am claiming that although there exists a further set of intuitions that in certain circumstances would lead us to narrow the application of the concept of "doing" even further (than rendered by the concept of what it would be reasonable$_A$ to expect), this does not speak against the construal of the narrower concept of "do" in terms of what it would be reasonable$_A$ to expect. One might worry that this is just an ad hoc maneuver.

I offer two responses. First, even if there were no reasons for thinking that there is indeed a distinction to be drawn, the concept of doing in the sense of what it would be

---

[17] I do not want to suggest that reasonableness$_N$ would be the only additional (beside reasonableness$_A$) consideration for us to include an action description as worth mentioning. In fact, it would be rather implausible to think so. A lot will depend on the pragmatic and contextual factors. Sometimes the sheer unlikelihood of a certain event will merit it special mention. (As I raised my arm to vote, I knocked down a fly I have been trying to get rid of all morning.)

reasonable$_A$ to expect of an agent would still be useful to the extent that it will allow us to capture the distinction between actions and mere happenings. Second, however, there are reasons for thinking that the distinction should be drawn in this way. I have suggested that there are additional reasons why we might refrain from listing certain descriptions of the action as worth mentioning (viz. when it is not reasonable$_N$ to hold the agent to the appropriate expectations). And indeed when we manipulated the cases so that it would become reasonable$_N$ to hold the agent to the expectation, it became appropriate to mention the relevant description in listing the things the agent did. Given the fact that the issue what the agent did is related to the issue whether the agent did anything (as we will see in the next section), it would be inadvisable to settle the question what the agent did in terms of reasonableness$_N$ of expectations, unless one was also prepared to let the question *whether* the agent did anything also be decided in terms of the standard of reasonableness$_N$. In view of the possibly perspectival nature of the latter concept, I think it would be highly inadvisable, unless there were additional reasons for choosing this option.

C. "He did it though it was unreasonable$_A$ to expect it of him"

One might object that it is possible for a person to do something despite the fact that it was unreasonable$_A$ to expect it of him. Consider an action such as breaking a world record in some sport, for instance. It seems indisputable that the expectation to break the world record is prima facie unreasonable$_A$. Yet, if the agent does break the world record, it would appear hard to deny him or her the credit of having broken the world record.

It will be true in general that the breaking of a swimming world record, say, is not something one can reasonably$_A$ expect of just anyone. Such an expectation would be systematically frustrated. But we have also suggested that on top of our understanding of general competence, we also allow for the agent's special abilities. In view of a swimmer's special talents, it may be reasonable$_A$ to expect of her that she break the world record.

To humor the objection, however, let us suppose that it is unreasonable$_A$ to expect it of her, that she just broke her ankle, or that her talent was not so great, and yet, in spite

of this she does break the world record. While these circumstances would make any reasonable person doubt the plausibility of such an event, let us grant it for the sake of the argument. Indeed, in such a case, (h) commits us to saying that the agent's breaking the world record was not a doing of hers in the narrower sense: it is something she happened to do, not something she did. But this is not an implausible result. In view of the *extreme* unlikeness of such an event, her breaking the world record begins to resemble her winning the lottery. It would be more appropriate to say that she only *happened* to break the world record, that this is not something she has done (in the narrower sense). This is not to say that there is no description under which it was a doing of hers in the narrower sense, e.g.: 'swimming' or 'taking part in a race'.

Perhaps a better case than breaking the world record in swimming would be shooting the bull's eye. The expectation to shoot a bull's eye is prima facie unreasonable$_A$. But if the person does hit the bull's eye, is it not something he did? That will depend on among other things, whether the person is a reliable shooter.

If the person is a good reliable shooter, it may be reasonable$_A$ to expect of him that he shoot the bull's eye (in the absence of special circumstances: something happened to his eye, he broke his fingers, etc.). In such circumstances, when he shoots the bull's eye, it is something he did. However, if a person is a bad unreliable shooter, then it will be unreasonable$_A$ to expect of him that he shoot the bull's eye. What if he does shoot one? Then it is reasonable to conclude that his successful shot was a matter of chance, an accident, it was something that *happened* as a result of what he did (aimed and fired toward the target) rather than something that he did (in the narrower sense of 'do').

This is an intuitive result except that it is possible that this successful shot was really a beginning of a series of successful shots. It may have been that he has been training hard, and with this first successful shot he became skilled in shooting bull's eyes. From then on it would be reasonable$_A$ to expect of him that he shoot bull's eyes. So, is it right not to count this first successful shot as something he did?

There is no reason to suppose that we need to have an answer to this question. Our sense of the concept of agency is geared toward circumstances where people are by and large reliable in fulfilling the expectations to which they are held, and by and large it does not apply in circumstances where we do not exhibit such competence. It is not clear

that we need to have any clear intuitions on what happens when we undergo a transition from one phase to the other.[18]

D. Butler's Problem

Let us close the discussion by showing how the apparatus developed so far helps us to make sense of the following puzzle.

> If Brown in an ordinary game of dice hopes to throw a six and does so, we do not say that he threw the six intentionally. On the other hand, if Brown puts one cartridge into a six-chambered revolver, spins the chamber as he aims it at Smith and pulls the trigger hoping to kill Smith, we would say if he succeeded that he had killed Smith intentionally. How can this be so, since the probability of the desired result is the same?[19]

The puzzle concerns the notion of a performance being intentional under a description rather than being a doing under a description, but I shall simply assume, as is plausible, that when someone does something intentionally, he does it in the narrower sense.

The reason why we would not say that Brown threw a six is that it would be unreasonable$_A$ to expect of Brown that he throw a six. The expectation to throw a six is prima facie unreasonable$_A$ (if the die is fair): it would be systematically frustrated.

Would it be reasonable$_A$ to expect of Brown that he kill Smith? The answer here depends on what exactly we take to be the content of this expectation. The expectation may be understood (a) widely, as suggested by the description of the action as a killing, and (b) narrowly, as suggested by the description of the method of killing. In case (a), the expectation will be fulfilled just in case Brown kills Smith (by any method). In case (b), the expectation will be fulfilled just in case Brown kills Smith from a randomly spun revolver with only one bullet in it with only one chance of a shot. Properly speaking, in

---

[18] This is essentially Wittgenstein's sense of our intuitions on this matter: "Take the case of a pupil...: if he is shewn a written word, he will sometimes produce some sort of sound, and here and there it happens 'accidentally' to be roughly right. A third person hears this pupil on such an occasion and says: "He is reading." But the teacher says: "No, he isn't reading; that was just an accident." — But let us suppose that this pupil continues to react correctly to further words that are put before him. After a while the teacher says: "Now he can read!" — But what of the first word? Is the teacher to say: "I was wrong, and he *did* read it" — or: "He only began really to read later on"? — When did he begin to read? Which was the first word that he *read?* This question makes no sense here" (Ludwig Wittgenstein, *Philosophical Investigations* (New York: Macmillan, 1958), §157).

case (b), the expectation is no longer the expectation of Brown that he kill Smith, but rather an expectation that he kill Smith in a particular way (from a particular revolver, etc.).

In asking whether it would be reasonable$_A$ to expect of Brown that he kill Smith, we must really ask two questions. Would it be reasonable$_A$ to hold Brown to the expectation to kill Smith by any method (case (a))? Would it be reasonable$_A$ to hold Brown to the expectation to kill Smith by shooting the revolver with only one bullet in it, etc. (case (b))? The answer to the first question is positive (case (a)). Even if there is a low chance for Brown to kill Smith when he fires the gun once, there are a lot of other ways in which Brown can kill Smith. The answer to the second question, on the other hand, is negative. It would be unreasonable$_A$ to expect of Brown that he kill Smith with a bullet from the gun in view of the fact that there is a one in six chance that he will do so, etc. We seem to reach the following conclusion: Brown killed Smith but he only happened to kill Smith with the revolver with one randomly located bullet with only one chance at a shot.[20]

To the extent that the puzzle is generated in the first place, it would seem that we tend to interpret the case as case (a) rather than (b) as is suggested by the context. One reason for this might be the fact that in view of moral and legal considerations it would be inappropriate to restrict the description of the action to (b). After all, what matters for our moral and legal practices is not so much a particular esoteric way in which a person gets killed but rather the fact that a person is killed by another person. This is what distinguishes the structure of the killing from the structure of the throwing of a six. To see this, consider a similar move on the side of throwing a six. Just as there are many

---

[19] Ronald Butler, "Report on Analysis Problem No. 16," *Analysis* 38 (1978), p. 113. The puzzle, together with a solution, first appeared in G. Harman, "Practical Reasoning," *op. cit.*

[20] The possibility of such bifurcation is envisaged by Harman who suggests that our judgment depends on the context in which we consider the action. "The reason why we say that the sniper intentionally kills the soldier but do not say that he intentionally shoots a bulls-eye is that we think that there is something wrong with killing and nothing wrong with shooting a bulls-eye. If the sniper is part of a group of snipers engaged in a sniping contest, they will look at things differently. From their point of view, the sniper simply makes a lucky shot when he kills the soldier and cannot be said to kill him intentionally" (Ibid., p. 434). I develop essentially this insight as a solution to the puzzle. For his own part, Harman suggests that what explains the puzzle is the moral value attached to the action of killing but not to the action of throwing dice.

more ways of killing Smith, so there are many more ways of throwing a six which would increase the chance of getting a six beyond the 1/6 chance. For instance, one can throw the die and then help it roll until it comes up six. In such a case, it would be reasonable$_A$ to expect of a person that he throw-roll the die so that it comes up six. The problem is that such an action is not recognized as throwing the die in the game of dice. Only one kind of way of throwing dice is legal in the game of dice, viz. throwing a fair die without any help.

In other words, the puzzle arises because the game of dice and our moral-legal practice recognize the respective actions in different ways. While what counts as throwing a die in the game of dice is restricted in a way that fixes the low probability, what counts as a killing in our moral-legal practice is not restricted to the case that fixes the low probability. Accordingly, we judge that the die was thrown by chance (taking account of the low probability), but that the person was not killed by chance (not taking account of the low probability).

4. Actions and Mere Happenings

Thus far, I have suggested that there are reasons for developing a concept of doing that would be sensitive to the way an action is described. I have argued that we can find some support for postulating a distinction between what an agent happened to do and what she did. I have further argued that many of these intuitions are explained if we understand what the agent did in terms of what it would have been reasonable$_A$ to expect of her. I will now suggest that we use the concept of doing something under a description in the way in which Anscombe and Davidson have used the concept of being intentional under a description, viz. to delimit the category of performances that are to count as actions. We can follow their recipe: an agent's performance is an action just in case there is a description under which it counts as the agent doing something (in the narrower sense). In a slogan, the agent did something (in the wider sense) if and only if he did something (in the narrower sense).

(A) A performance p is an action if and only if for some φ such that p pf-fulfills the expectation of α that α φ, it was reasonable$_A$ to expect of the agent that she φ.

Correlatively:

(H) A performance *p* is a mere happening if and only if for every φ such that *p* pf-fulfills the expectation of α that α φ, it was unreasonable$_A$ to expect of the agent that she φ.

I want now to show that (A)-(H) capture both unproblematic and more problematic cases of actions.

In what follows, I will assume, following Davidson, that the performances that qualify for the status of actions are bodily movements. This is a simplifying assumption on this account. I have already noted in Chapter III that a proper treatment of the concept of performance would require an account of the ontology of actions, as well as an account of the consequences of actions. Since I cannot undertake either of the tasks here, I will simply follow the trodden path.[21]

I demonstrate now that (A)-(H) cover all that is covered by the intentionalist criterion (in particular section A and part of section B). Section B then proceeds to discuss which omissions qualify as among the things the agent does. Finally, in section C, I consider defeating conditions that render performances non-agentive and those that do not. In the next section 5, I will show that the account straightforwardly excludes the cases of basic wayward causal chains from qualifying as actions.

A. Positive Actions

On the intentionalist criterion, a performance is an action just in case there is a description under which it was intentional. I show that the non-intentionalist criterion (A)-(H), allows us to capture all that is captured by the intentionalist criterion. (I consider cases of intentional omissions in section B).

---

[21] The fact that the path is trodden does not mean that there are no disagreements in the vicinity. One particular debate concerns the question what exactly should count as a bodily movement. Davidson appears to think that ordinary bodily movements qualify as actions ("Agency," *op. cit.*). His main challenger is Hornsby, who has denied that we should identify actions with bodily movements in the way we would be tempted to conceive of them (*Actions, op. cit.*). John McDowell develops a conception of agency under which Davidson's intuitions can be defended from Hornsby's arguments (presented during a seminar on Philosophy of Action, University of Pittsburgh , Fall 1994).

When the Agent Acts Intentionally: Intended and Unintended (but Foreseen) Intentional Doings.[22] It is appropriate to begin with intentional performances where the agent acts on some prior intention since they have been paradigmatic to the intentionalist view. I will distinguish between intended and unintended (but merely foreseen) intentional actions.[23]

Consider an example of Harman's. Albert intends to improve the appearance of his lawn by cutting the grass with a power lawn mower, realizing (though not intending) that he will thereby release some fumes into the air and irritate his neighbor who wants her lawn to look the best. Harman believes that when Albert performs the action it is appropriate to say that he intentionally improves his lawn, intentionally cuts the grass. just as he intends to do. Albert also intentionally releases fumes into the air and intentionally irritates his neighbor, albeit he does not intend to do either (he merely foresees that he will do so when he does what he intends to do).

I am not concerned to see whether the actions are indeed best construed as being intentional under all the descriptions. All I am concerned to show is that they indeed describe an action according to (A)-(H). The performance in question is Albert's walking to and fro cutting grass with the mower. Is it an action on our account? It will be as long as there is a description of the performance under which it would be reasonable$_A$ to expect the agent to perform the action. Let us take the description 'cutting the grass with the lawn mower'. Is it reasonable$_A$ to expect of Albert that he cut grass with the lawn mower? The answer is positive. It would be negative if Albert suffered from a temporary disability, if the lawn mower was damaged, etc. As things stand, it is reasonable$_A$ to expect of Albert that he cut the grass. This is sufficient to show that the performance (however described) is an action of Albert's. Thus if we chose to describe

---

[22] It should be clear that the category of intentional actions is a misnomer. On the most popular Anscombe-Davidson view, it does not single out a class of actions but rather a class of action descriptions. I employ the terms 'intentional action' and 'unintentional action' without implying thereby that we are dealing with separate classes of actions. My point is only to consider how examples of such actions would be categorized in terms of (A)-(H).

[23] There is considerable debate whether the bringing about of what the agent merely foresees but does not intend ought to be considered as something he does intentionally. For an affirmative answer, see G. Harman, *Change in View, op. cit.* M.E. Bratman, *Intention, Plans, and Practical Reason, op. cit.* For challenge, see e.g. Carlos J. Moya, *The Philosophy of Action* (Cambridge: Polity Press, 1990).

the performance in terms of one of the consequences, as "improving the lawn," we could say that Albert's improving the lawn is in this case also an action of his. For the same reason, Albert's releasing the fumes into the air, his irritating his neighbor, but also his moving about a particular water molecule in a certain fashion, his not flying to the moon on this occasion, all describe Albert's action.[24]

In other words, the descriptions under which the agent intended to perform the intentional action he did perform settle it that it is reasonable$_A$ to expect of the agent that he perform the action under those descriptions, thereby settling it that the intentional doing is indeed an action according to (A)-(H). What if we considered some of the descriptions under which the agent did not intend but merely foresaw that he will act? In other words, would it be reasonable$_A$ to expect of Albert that he release the fumes into the air or that he irritate his neighbor? We might think that there is a defeating condition rendering such expectation unreasonable$_A$. Given that Albert started the (reliable) motor, the expectation to release the fumes into the air would be systematically fulfilled (and its contrary systematically frustrated). Indeed, intuitively we would think that once he started the motor, there is nothing he can do about the release of the fumes. But in this case, it is of course reasonable$_A$ to expect of Albert that he not start the motor, so the defeating condition is defeated. A similar reasoning applies to the description 'irritating the neighbor'. This means that in the case as it is described, even the descriptions under which the intentional action was not intended by the agent would suffice to render the performance an action according to (A)-(H).

When the Agent Acts Unintentionally. Alongside things we do intentionally, there are many things we do unintentionally. When Oedipus married Jocasta, he did not know she was his mother: he unintentionally married his mother. These cases can be handled in a

---

[24] I am simply following Davidson here in thinking that the action is a particular event, and it can be described in many however irrelevant ways. I should note that not all of the descriptions of the action count as action descriptions in the narrower sense, i.e. as specifying something Albert did rather than happened to do. It is clear that Albert's cutting the grass and his improving the lawn is something he did (in the narrower sense) on this occasion. It is also clear given the arguments in section 3 that Albert's moving about a particular water molecule or his not flying to the moon on this occasion do not qualify as things Albert did (in the narrower sense) even though they qualify as Albert's doings (in the wider sense), and even if they could be foreseen.

way suggested by Davidson. Davidson's account of what it means to φ unintentionally is simple. It means that the agent performed an action, which can be described as φing, but that it is not intentional under that description. Since Davidson believes that an event is an action if and only if it is intentional under some description, it follows that when an agent φs unintentionally, her φing is intentional under some description different from 'φing'.

In our terms, when an agent φs unintentionally, there is a description of the action under which it would be reasonable$_A$ to expect the action of the agent. In Oedipus' case, it would be reasonable$_A$ to expect of Oedipus that he marry Jocasta, even though given Oedipus' ignorance of the identity of his mother, it would be unreasonable$_A$ to expect him to marry his mother.[25] Because there is a description of Oedipus' performance under which it would be reasonable$_A$ to expect of him that he do it (viz. that he marry Jocasta), the performance described as his unintentionally marrying his mother is an action.

Consider another example of Davidson's. He imagines someone entering a room, switching on the light, thereby unintentionally frightening a burglar who, unbeknownst to the agent, is plundering one of his rooms. Here once again, given the agent's ignorance, it would be unreasonable$_A$ to expect of him that he frighten the burglar. However, it is still reasonable$_A$ to expect of him that he switch on the light. Hence, the performance described as his unintentionally frightening the burglar is an action.

Spontaneous Actions. One of the virtues of the account proposed thus far is that it makes a clear division between two questions, the question of what actions are and the question of how they are explained (I address it in Chapter VII). While similar categories (normative expectations) are employed in both accounts, the account of the nature of action does not require that the agent act because of any particular normative expectation to which she holds herself or to which she is held by another. Rather our criterion is

---

[25] Ordinarily, it would be reasonable$_A$ to expect of Oedipus that he marry his mother (we should remember that we are talking about reasonableness$_A$, what is within the agent's power, not about reasonableness$_N$, what is appropriate). Prima facie, it would be reasonable$_A$ (though not reasonable$_N$) to expect of any man whose mother was alive that he marry his mother. If, however, Oedipus does not know who his mother is, it would no longer be reasonable$_A$ to expect of him that he marry his mother. Given such an ignorance, the expectation to marry one's mother would be systematically frustrated.

counterfactual: were someone to hold the agent to an expectation, it would be reasonable$_A$. It is this separation that allows us to capture another category of actions, spontaneous actions, that have caused some tensions on the intentionalist accounts.

G.E.M. Anscombe has suggested that we delimit the sphere of our agency to those events to which a special sense of the question "Why?" applies. The special sense of the question is understood by the special answer that is appropriate to it, viz. an answer that gives the reason for the action. There is a class of actions to which the question applies but a special case of the typical answer is appropriate. Rather than giving a reason for an action, the answer can be "For no reason."[26] For want of better terminology, let us call the actions done with no reason "spontaneous" actions. They include walking down the meadow and picking up daisies for no apparent reason, pacing the room to and fro, and so on.

Such actions are included in our characterization. When I am walking down the meadow picking up daisies for no apparent reason, it would surely be reasonable$_A$ to expect of me that I do so. What would make it unreasonable$_A$ to expect of me that I pick up daisies, for example, is the fact that I have a bad back-ache and cannot bend down to pick them up. But in absence of such and other debilitating circumstances, it would be reasonable$_A$ to expect of me that I do as I do in this case.

I have already emphasized that what allows us to capture spontaneous actions is the fact that for it to be reasonable$_A$ to expect something of an agent, the agent need not be actually held to a reasonable$_A$ normative expectation by any one. As we said, it is quite sufficient to require only that were the agent held to the expectation, it would be reasonable$_A$ to hold her to it. In this way, the agent can act quite spontaneously, not responding to any expectations, and her performance will count as her action. In order to capture such actions, the intentionalist needs to appeal to the notion of intention-in-

---

[26] Anscombe originally characterizes such actions as done "for no reason." One has to be careful, however, to distinguish the force that the reason occupies. An action may be done for no reason while the agent has some reason. In such a case to say that it is done for no reason is to imply that the reason is not efficacious. On the other hand, an action may not only be done for no reason, but the agent not even have a reason to do it. And many of the cases of spontaneous actions seem to belong to the latter category. In any case, we cannot account for actions done for no reasons strictly speaking until we know what the explanatory

action. I have argued in Appendix B that there are conditions under which such an appeal is questionable.[27]

B. Omissions

One of the virtues of the account is that it qualifies omissions, including some unintentional omissions, as actions. An omission involves a breach of an expectation to which the agent is or would be reasonably held.[28] The agent can breach the expectation intentionally or unintentionally, thus be committing either an intentional or an unintentional omission. I will argue that intentional omissions and many unintentional omissions qualify as actions on our account. First, however, we need to be clear about the special status of the expectation that is being breached.

Both standards of reasonableness will be involved in judging a performance to be an omission. In taking an omission to be something the agent has done rather than something that happened to her, we will take an omission to φ as a performance that it would have been reasonable$_A$ to expect of the agent under the description 'not φing'. So when Jane intentionally omits to pay the taxes, it is something she does, because it is reasonable$_A$ to expect of her that she not pay the taxes. When Tim intentionally omits to meet a friend in the library, it is reasonable$_A$ to expect of him that he not go to the library. In taking an omission to be an omission, in turn, we will take the performance (the not-φing) to frustrate a reasonable$_N$ expectation. Jane's action of not paying taxes breaches a reasonable$_N$ expectation to pay the taxes, to which she is held by the state. Tim's action of not going to meet his friend breaches a reasonable$_N$ expectation on his friend's part.

---

relation between actions and reasons amounts to. And this will be explained only in Chapter VII. For now, we will speak of actions done with no reasons.

[27] One might speculate that what distinguishes spontaneous actions from other actions is the fact that they are actions that it would be neither reasonable$_N$ nor unreasonable$_N$ to expect of the agent. For this is the most natural rendition in our terms of what it means to say that they are done with no apparent reason.

[28] This general account of omissions is presented in Steven Lee, "Omissions," *Southern Journal of Philosophy* 16 (1978), 339-354 and in a series of articles by Patricia Smith: "Allowing, Refraining, and Failing. The Structure of Omissions," *Philosophical Studies* 45 (1984), 57-67; "Ethics and Action Theory on Refraining: A Familiar Refrain in Two Parts," *The Journal of Value Inquiry* 20 (1986), 3-17; "Contemplating Failure: The Importance of Unconscious Omission," *Philosophical Studies* 59 (1990), 159-176.

Suppose John is committed to being at a meeting at 9am, but that he does not feel like going and does not go, merely watching the minutes slide by. In such a case, we would say that John intentionally omitted to go to the meeting. John's failure to go is an action. There are no special circumstances that would make it unreasonable$_A$ to expect of him that he not go to the meeting or that he go to the meeting. It is because it is reasonable$_A$ to expect of John that he not go to the meeting that his not going to the meeting is something he does, and hence that it is an action. It is because it is reasonable$_N$ to expect of John that he go to the meeting, since he is committed to being there, that his not going to the meeting is an omission.

Similar reasoning applies to cases of unintentional omissions, though not all unintentional omissions will qualify as actions. Suppose Jane is committed to being at that same meeting but that she simply oversleeps. Jane's failure to come to the meeting qualifies as an omission since her performance frustrates a reasonable$_N$ expectation that she be at the meeting. Whether Jane's not coming to the meeting will qualify as something she has done will depend on whether it was reasonable$_A$ to expect of her that she not come. I have already argued in Chapter V that despite the fact that Jane is asleep, it would be in this case reasonable$_A$ to expect of her that she not come to the meeting. This is because although being asleep is systematically correlated with the fulfillment of the expectation not to go the meeting and with the frustration of the expectation to go to the meeting, in a normal case it is also reasonable$_A$ to expect of the agent that she prevent herself from oversleeping. Thus Jane's failure to come to the meeting is something she has done rather than something that she happened to do and so it is an action.

There are circumstances, where it would be unreasonable$_A$ to expect of the agent that she prevent herself from oversleeping. This will be the case when she oversleeps as a result of being drugged or as a result of serious illness, for example. These conditions defeat the reasonableness$_A$ of the expectation that she not oversleep. Being drugged is systematically correlated with the frustration of the expectation that the agent prevent herself from oversleeping. At the same time, it will be normally unreasonable$_A$ to expect of the agent that she prevent herself from being drugged. Similarly, being seriously ill might be systematically correlated with the frustration of the expectation that the agent not oversleep, and it would be unreasonable$_A$ to expect of the agent that she not fall ill.

The fact that the apparatus does not allow all unintentional omissions to qualify as actions is a virtue of the account. One of the reasons one might be wary of admitting the most straightforward cases, which we recognize as actions in holding the agents responsible for them, is that once the door is open to them, nothing can stop the others. And indeed, the intentionalist criterion of agency does not allow for such a discrimination. But using the criterion of reasonableness$_A$ of expectations allows us to distinguish problematic from unproblematic cases.

C. Mere Happenings

Consider now examples of cases where defeating conditions occur. In some cases, the occurrence of the defeating condition results in the performance not counting as an action, in other cases it does not. Some defeating conditions render all expectations fulfilled by the performance unreasonable$_A$, in which case the performance is a mere happening (we may call them "global defeating conditions"). But other defeating conditions render only some expectations fulfilled by the performance unreasonable$_A$, in which case the performance thus described would not count as something the agent did (in the narrower sense), but it would nonetheless be an action. I briefly discuss cases where spasms and physical compulsion function as global defeating conditions. (I also show how physical compulsion differs from coercion: the former is a defeating condition, while the latter is not.) Other common global defeating conditions include: coma, various forms of handicap, physical force, hypnosis, etc. I also briefly discuss a case with a local defeating condition, which does not render the performance nonagentive.

Spasms. When a spasm causes my arm to rise, which hits the lamp causing it to break, it may appear as if I am raising my arm, as if I am breaking the lamp. My performance does not qualify as action, for none of these descriptions of the performance qualify as something I have done. In view of the fact that the spasm occurred, it was unreasonable$_A$ to expect me to raise my arm (or indeed, not to raise it). For the same reason, it would be unreasonable$_A$ to expect of me that I hit the lamp (or indeed, not hit it). In general, the occurrence of the spasm, makes expectations having to do with the temporary control over my arm unreasonable$_A$. This is why when my arm rises and breaks the lamp, the performance is not an action, but a mere happening.

Physical Compulsion. Suppose that a person is physically forced to sign a document by another. As long as the force applied is overwhelming, it would be unreasonable$_A$ to expect of the person that she sign the document (the expectation would be systematically fulfilled and its contrary systematically frustrated). With the force in play, it is no longer "within her power" to sign the document. For the same reason, it would be unreasonable$_A$ to expect of her that her arm move in the way involved in the signing, or that she put a dot over 'i'. Thus her signing the document (led by another's hand), her arm moving in a certain way, her putting a dot over 'i' are all specifications of what happened to her rather than of what she did. Her performance is accordingly a mere happening, not an action.

Coercion. Aside from physically forcing a person to sign a document, one might coerce her to do so. One might threaten her life if she does not sign the document. In fact, a superficial application of our account could present this as an objection. For when a coerced person does sign the document, her signing the document is an action (albeit coerced, it is something she does intentionally). And yet, it might be objected, in such a case, it would be unreasonable to expect of her that she not sign the document.[29]

That this is a superficial application of the account becomes clear when we ask what sense of reasonableness is at stake. It seems clear that in view of the extreme danger the agent finds herself in it would be inappropriate to hold her to the expectation that might endanger her life. In other words, it is unreasonable$_N$ (in the normative sense) to expect of her that she not comply. But the specifically normative standard of reasonableness$_N$ does not enter into the judgment that an action has been performed. And when we ask whether it is reasonable$_A$ to expect of her that she sign (or not sign) the document, the threat regarding the consequences of her actions does not make it

---

[29] Indeed, such an account of coercion is presented in Robert Nozick, "Coercion," in (eds.) Sidney Morgenbesser, Patrick Suppes, Morton White, *Philosophy, Science, and Method* (New York: St. Martin's Press, 1969), pp. 440-472. See also Bernard Gert, "Coercion and Freedom," *Nomos* 14 (1972), 30-48; Patricia Greenspan, "Behavior Control and Freedom of Action," *Philosophical Review* 87 (1978), 225-240; "Unfreedom and Responsibility," in (ed.) Ferdinand Schoeman, *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987), pp. 63-80.

unreasonable$_A$ to expect of her that she sign the document, or indeed that she not sign it. Her performance is thus an action.

When Nature Does Not Cooperate. Finally, let us consider some examples of defeating conditions that do make it unreasonable$_A$ to expect a performance of the agent under some description, but not under sufficiently many descriptions (i.e. all those true of the performance) to render the performance a mere happening.

Suppose that Jane is angry with Tamara and intends to finally tell her about it. When they meet, Jane begins her well-rehearsed sermon. Suddenly Tamara faints and is rushed to a doctor, making it impossible (not to mention inappropriate) for Jane to continue. As a result Jane does not tell Tamara off. Jane's failure to tell Tamara off is an action of hers but it is not something she has done under the description "not telling Tamara off." Given that Tamara faints in midway, it would be unreasonable$_A$ to expect of Jane that she tell Tamara off (for the expectation would be systematically frustrated) as it would be unreasonable$_A$ to expect of Jane that she not tell her off (for the expectation would be systematically fulfilled while its contrary systematically frustrated). Jane's failure to complete her sermon is something she happened to do, not something she did. But there are other descriptions of Jane's performance that qualify it as an action rather than a mere happening. The defeating condition, Tamara's fainting, does not render it unreasonable$_A$ to expect Jane to begin her telling Tamara off, for example. Nor does it render it unreasonable$_A$ to expect of Jane that she gesticulate as she is uttering the words. The defeating condition is in this case of a local rather than of a global nature.

5. Wayward Causal Chains

The causal theorist conjectures that only those events that are caused by mental states count as actions. What stands in the way of claiming that all events caused by mental states are actions are the cases involving so-called wayward causation. It is then incumbent upon the causalist to restrict the events caused by mental events to include only those that are actions. And various ways of doing so have been suggested.[30] By

[30] Davidson suggested that the causation has to be of the right sort and argued that we cannot explicate it more in view of the nature of the anomalous relation between the physical and the mental ("Freedom to

contrast, wayward causal chain cases do not present any additional problems for the responsibility-based accounts. Our requirements are quite sufficient to sort out the wayward cases from the normal ones. They are simply cases where otherwise reasonable$_A$ expectations cease to be reasonable$_A$ in view of the occurrence of a (global) defeating condition.

Before going on, let us make a distinction between two kinds of cases of wayward causal chains. First, there are cases of consequential waywardness,[31] where although the waywardness of the causal chain interferes with a given event's being an intentional doing, it does not interfere with its being an action. The classic example is due to R. Chisholm.[32] He imagines a nephew who plans to murder his uncle in order to inherit his fortune. His intention causes him to drive so recklessly on the way to carry out his plan that he runs over a pedestrian, who, unknown to him, is his uncle. This is a case where the nephew does perform an action of killing his uncle but unintentionally (the action is intentional under the description "driving the car as fast as possible" but not under the description "killing the uncle").

Act," in *Essays on Actions and Events, op. cit.*, pp. 63-81; C.J. Moya, *The Philosophy of Action, op. cit.*). Frankfurt suggested that it involves the notion of agent guidance ("The Problem of Action," in *The Importance of What We Care About* [Cambridge: Cambridge University Press, 1988], pp. 69-79.). Others claimed that it involves the idea that among the causal antecedents of action are intentions that represent themselves as causing the action in question and such self-referring intentions are then not realized in the wayward cases (G. Harman, "Practical Reasoning," *op. cit.*; J.R. Searle, *Intentionality, op. cit.*; J. David Velleman, *Practical Reflection* [Princeton, NJ: Princeton University Press, 1989]). On the other hand, there were attempts to give at least some account of the conditions under which the causation involved would be of the right sort. One of the most promising ways of handling deviance cases is due to Adam Morton ("Because He Thought He Had Insulted Him," *Journal of Philosophy* 72, 1975, 5-15). Morton observed that what is characteristic of intentional behavior is that it is sensitive to relevant information in appropriate ways. The deviant cases are deviant because the behavior involved in them is not appropriately responsive. So, for instance, had the nervous mountaineer realized that there was a high probability that if he loosened his hold on the rope he would be likely to fall (due to a complicated safety system), if his behavior were intentional he would not loosen his hold. However, this realization would not inhibit his losing control over his fingers. This sensitivity strategy seems to work rather well (though see John Bishop, *Natural Agency. An Essay on the Causal Theory of Action* [Cambridge: Cambridge University Press, 1989]; Christopher Peacocke, *Holistic Explanation. Action, Space, Interpretation* [Oxford: Clarendon Press, 1979]). However, its focus is on intentional action. Accordingly, the concept of action thus singled out is quite different from the one developed here. For instance, many omissions are actions that are hardly responsive to the relevant information in the required fashion. To the extent that the sensitivity strategy will work then it will work too well from our point of view.

[31] The terminology (consequential vs. antecedential waywardness) is due to Myles Brand, *Intending and Action. Toward a Naturalized Action Theory* (Cambridge, MA: The MIT Press, 1984).

Second, there are cases of basic (antecedential) waywardness where the waywardness of the chain interferes not only with the event's being intentional under an appropriate description but also with that event's being an action. Paradigmatic here is Davidson's example of a nervous mountaineer:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally.[33]

In such a case, where the nervousness is severe enough, the agent does not perform an action. The agent's intention causes him to lose control as a result of which the intended effect happens.

We are committed here to being able to accommodate the latter cases of basic waywardness, for it is only if we are able to accommodate them that I will be able to claim that the account adequately captures the distinction between mere happenings and actions.[34]

In general, what happens in the cases of wayward causal chains is that the intention to φ that normally causes φing causes some state ξ which in turn causes φing, except that because φing is mediated by ξ it is not an action. This would be easily understandable if the event ξ were a defeating condition. And indeed this is the case.

Let us recast a Davidson-like case in our terms. A mountaineer forms an expectation of himself that he rid himself of a piece of equipment. He then becomes very nervous at the thought that he might have trouble making a safe return without some of the equipment.[35] We know that when a person is really nervous, he might temporarily lose control over some bodily movements. His palms might sweat and objects might slide out of them. Such a person is not reliable in holding objects, making precise

---

[32] Roderick M. Chisholm, "Freedom and Action," in (ed.) Keith Lehrer, *Freedom and Determinism* (New York: Random House, 1966), pp. 11-44.

[33] D. Davidson, "Freedom to Act," *op. cit.*, p. 79, original emphasis.

[34] I do not show that the account can be extended to cover the cases of consequential waywardness. Such an extension would require a systematic account of the consequences of actions.

movements, holding onto ropes. A person in this state would systematically frustrate the pair of expectations to let go and to hold on to the rope. In other words, it would be unreasonable$_A$ to hold a very nervous person to an expectation to hold on (or let go of) the rope. At the same time, it does not seem particularly reasonable$_A$ to expect of the agent that he not become nervous. If so then given the state the mountaineer found himself in (or caused himself to be in), his dropping the rope would not count as an action of his.

It should be noted here that the state of nervousness must have been really severe. It could not have been ordinary stage-fright for it to count as a defeating condition. It must have been severe enough to reach a state which interferes with an ordinary reliability in responding to the expectation of dropping a rope with rope dropping.

It may be worthwhile to stress why it is so easy to accommodate wayward causal chain cases on our account. This can be best seen by considering why wayward causal chains constitute a problem for the causal theory of action. The simplest (and abandoned) version of a causal theory holds that a performance is an action just in case it has been caused by a preceding mental state that justifies the action under some description. One way of diagnosing the problem with this thought is that it leaves the causal process largely out of the agent's purview: once the agent's reason or intention sparks off the causal process, it is out of her hands.[36] It is as if once the agent puts the process in motion, the action just happens, but the agent does not "actively" perform it. A general way to aid the problem has been to stipulate that the agent guide the process[37] or that the process be sensitive to the agent's reasons or to relevant information.[38]

---

[35] Note that it is significant from the point of view of a causal theory of action that it is the very beliefs and desires that justify the mountaineer's forming the intention that cause the state of nervousness. It is but a curious feature from the present point of view.

[36] This is essentially Frankfurt's objection to causal theories of action developed in his "The Problem of Action," *op. cit.* Frankfurt's response is to require that the process remain under what he calls "agent guidance."

[37] Frankfurt, Ibid.

[38] A. Morton, "Because He Thought He Had Insulted Him," *op. cit.*; David Lewis, "Veridical Hallucination and Prosthetic Vision," in *Philosophical Papers, vol. II* (Oxford: Oxford University Press, 1986), pp. 273-290.

No such additional amendments are required on our account. The demand that it would be reasonable$_A$ to expect of the agent that he perform an action pertains to the time at which the agent performs the action. The judgment whether or not it would be reasonable$_A$ to hold the agent to a particular expectation is sensitive to any untoward circumstances that happen up until the time when the action takes place. And it is precisely this feature that allows us to disqualify cases of wayward causal chains from counting as actions. One might try to turn this virtue into a vice, however. Here is how.

When the Agent Can No Longer Stop the Course of the Action... Consider the action of taking a step down (but any action would do). No untoward circumstances make the expectation that the agent take a step down unreasonable$_A$. However, just before the agent completes the action and takes the step down, it will be true that there will be a point (a stage in the performance of the action, we might call it "the point of no return") which once it occurs makes it physiologically impossible for the agent not to take the step down. In the case of taking the step down, such a state may even be felt if one takes the step very slowly. One may feel in control of taking the step and then suddenly feel oneself leap forward. Given the occurrence of the point of no return, the expectation that the agent take a step down will be systematically fulfilled (and its contrary systematically frustrated). Hence, it would seem, the expectation would be rendered unreasonable$_A$. Insofar as such a point of no return will occur for all actions, the objection shows that nothing qualifies as an action on our account.

The objection fails, however. Even if in the case of every action, there is such a point of no return, which is systematically correlated with the fulfillment of a relevant expectation, it does not yet follow that it would be unreasonable$_A$ to expect the agent to perform the action. It would be unreasonable$_A$ to expect the agent to perform the action if it were also unreasonable$_A$ to expect of the agent that he bring it about that the point of no return occurs. In the example just given, it is not clear that it would be unreasonable$_A$ to expect of the agent that he bring it about that the point of no return occurs.

• • • • • •

So, what is the difference between my raising my arm and my arm rising? I have suggested that the difference amounts to it being reasonable$_A$ to expect of me that I raise

my arm under some description in the former case, and it being unreasonable$_A$ to expect of me that I raise my arm under any description in the latter case.

The structure of this answer resembles the structure of the intentionalist solution of the problem of action. On that view, a performance is an action just in case there is some description under which it is intentional. On the view defended in this chapter, a performance is an action just in case there is some description under which it is something the agent has done (in the narrower sense discussed in sections 2-3). Despite this resemblance and the fact that our account captures all the cases captured by the intentionalist view, the account allows us to understand the broader notion of conduct. I have shown how even some unintentional omissions, omissions that occur while the agent is sleeping e.g., can qualify as the agent's doings. But the account is sensitive enough not to qualify all unintentional omissions as actions (if an agent's oversleeping was caused by his being drugged, his omission would not count as his action). It is also one of the virtues of the account that it excludes cases of wayward causal chains without the need for amendments (section 5). The wayward causal chain cases are simply special cases where it would be unreasonable$_A$ to expect of an agent that she perform an action. Finally, our account appeals to normative expectations only counterfactually. It is only required that if the agent were held to a given normative expectation, such an expectation be reasonable$_A$. This allows us to divorce the notion of action from the arbitrary facts concerning whether someone is actually being held to the expectation by another (or indeed by himself). This is also ultimately responsible for the ease with which the account applies to spontaneous actions, actions done with no reason.

I have now completed all but one task. I have given a nonintentionalist answer to the problem of action by appealing to the concept of practical responsibility (Chapters III-V). I have shown the concept of practical responsibility to be immune to the fundamental problem (Chapters III-V). The proposed account shows how it is possible to develop a concept of action in terms of the concept of practical task-responsibility. In Chapter V, I have offered an unified account of defeating conditions, all the conditions in the presence of which we are inclined to withhold attributions of agency. One final task that remains is to try to understand the relation between reasons and actions. I proceed to do so in the final chapter.

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.

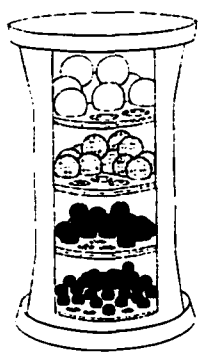


Figure 5. Sober's selection toy

(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 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 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^SE^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-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), [-E^SE^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), [-E^S-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^SE^G] and [E^S-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-E^G]), but would not have raised it if he were not nearing the turn ([-E^SE^G], [-E^S-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 [-E^SE^G]), but would not have raised it if he were not passing the friend ([E^S-E^G], [-E^S-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^DE^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-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*. [-$E^D E^T$] 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*. [-E^D-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_{ψ}[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.

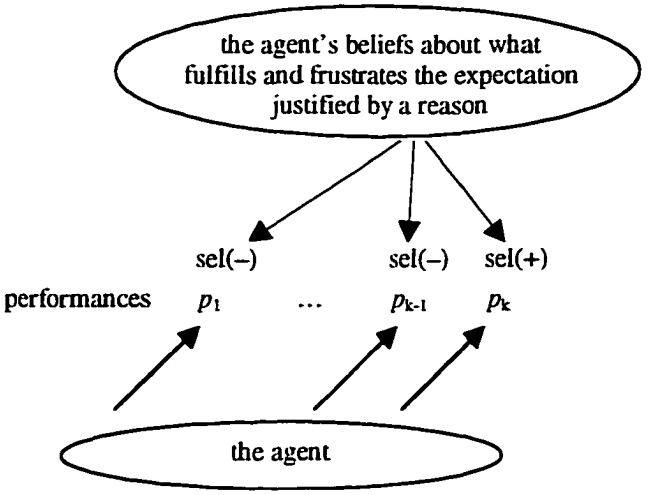


Figure 6. Causal relations involved in the agent's acting for a reason on the selectional model. 'Sel(+)' and 'Sel(–)' mark positive and negative selection, respectively.

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.

CONCLUSION

I have addressed two issues in the philosophy of action: Wittgenstein's question how to understand the distinction between actions and mere happenings, and Davidson's challenge to give an account of the explanatory relation between reasons and actions. I have followed an old-fashioned strategy in answering the first question, seeking clues to the answer not by searching for the conditions that render actions actions (thought of as reasons or intentions on the intentionalist approach) but rather by searching for the conditions that render mere happenings mere happenings (defeating conditions like spasms, coma, sleep, handicap, etc.). More systematically, I have argued that a performance is an action just in case there is some description under which it would have been reasonable$_A$ to expect of the agent that he perform it (Chapters III-V have explained the special sense assigned to the technical terms invoked). In Chapter VI, I have shown that the proposed account captures all the cases captured by the intentionalist view, and straightforwardly excludes cases of basic wayward causal chains from qualifying as actions. Moreover, it is able to qualify some unintentional omissions as actions. In this way, the nonintentionalist view I have sketched gives an account of our conduct, including our agentive voice and silence.

The concept that figures crucially in the answer to the first question is the concept of a normative expectation. In Chapter I, I have suggested that, contrary to first impressions, that concept is not unrelated to the way in which we explain actions. In fact, I have shown that the concept of normative expectation may be thought to play just the double role that the concept of intention has been thought to play on the intentionalist account. For the concept of intention is usually thought to be important in answering both of the above questions. Insofar as the idea of a performance being intentional under a description presupposes some concept of intention, it figures in the intentionalist answer to the question what actions are. It also plays a role in the causalist answer to the question how reasons, intentions, etc., relate to actions, viz. causally. The concept of

normative expectation figures not only in the answer to the first question, but also in the answer to the second question.

In Chapter VII, I have argued that we can think of reasons as justifying the normative expectations on which the agent acts. I have sketched a selectional model of what it means to say that an agent acts on one expectation rather than another (acts for one reason rather than another). The model relies on, among others, causal relations but not on the causal relation between the reason or the expectation and the action. I have suggested that it illuminates what we mean when we suppose that there is a "causal" relation between the reason and the action.

In Chapter VII, I have also argued that the selectional model allows us to understand how it is possible for an agent to act on another person's expectation of him. I have also shown that there is no reason, internal to that model, to suppose that the agent's acting on another person's desire must be mediated by her acting on her own desire that is suitably related to that person's desire. In Chapter I, I have demonstrated that many external arguments also fail to establish this conclusion. I have thus defended the position of explanatory nonindividualism, whose distinctive claim is that aside from being sometimes moved by our desire to satisfy another person's desire, we are also sometimes moved by that person's desire without thereby being moved by our desire to satisfy that person's desire.

I should emphasize, as I have been doing throughout, that though the answers to the two questions, the problem of action and the problem of the explanatory force of reasons, sound common notes (both employ the notion of normative expectation, for example), they are really different answers to different questions. The question of how to explain an action is a question about which of the normative expectations, to which the agent is *actually* held, has been operative in the agent's acting. The question of whether a performance is an action depends on whether or not it *would be* reasonable$_A$ to expect the performance of the agent under some description. The answer to the second question is independent of any actual expectations to which the agent is held.

The theme that reverberates in the answers to both questions, however, is the need to look to the social nature of our agency. The answer to Wittgenstein's question appeals to a social criterion of what it would be reasonable$_A$ for *us* to expect of the agent.

Whether or not the agent's performance lives up to the standard is largely a question of whether or not untoward circumstances (defeating conditions) have interfered. The focus is thus removed from the agent's inner life, from the practical reasoning in which she is sometimes involved, and shifts toward the way in which her performance affects the social fabric of normative expectations. An agent's habitual, unreflective, spontaneous actions as well as unintentional omissions intuitively require no mental involvement on the agent's part, and yet they do form a part of the agent's conduct, they can affect others in ways which would be agentively traceable to the agent. Likewise the selectional nonindividualist answer to Davidson's challenge opens a new way of looking at the interactions between others and the agent. In allowing for the possibility that the agent acts on another's desire directly as it were, without acting on her own desire (though, as we saw, still acting on some of her beliefs), the account shows vividly that our being embedded in the network of social expectations does not necessarily leave the agent cold, but can move her to action.

APPENDICES

APPENDIX A.

THE ASYMMETRY THESIS

J.M. Fischer has argued[1] that there is an important asymmetry in responsibility ascriptions between actions and omissions. Fischer offers two kinds of cases which are to show that we indeed do harbor intuitions supporting the thesis. We will see that if one casts Fischer's cases in our apparatus, one will be able to explain the intuitions without needing to postulate any asymmetry between action and omissions. Indeed, the apparatus handles objections put forward against the asymmetry thesis equally well.

1. The Asymmetry Thesis

Fischer agrees with Frankfurt that responsibility for actions does not depend on our ability to have done otherwise. He suggests, however, that responsibility for omissions does depend on our ability to have done otherwise. This is the asymmetry thesis: there is an asymmetry in responsibility ascriptions between actions and omissions.

To substantiate the thesis Fischer considers two kinds of cases: Frankfurt-type cases of actions for which we are responsible despite not having been able to do

---

[1] The original thesis appeared in John Martin Fischer's "Responsibility and Failure," *Proceedings of the Aristotelian Society* 86 (1985/86), 251-270. It has been elaborated in a paper with M. Ravizza ("Responsibility and Inevitability," *Ethics* 101, 1991, 258-278), and is upheld by Fischer in *The Metaphysics of Free Will. An Essay on Control* (Oxford: Basil Blackwell, 1994). In the meantime, the thesis has received some attention, see Randolph Clarke, "Ability and Responsibility for Omissions," *Philosophical Studies* 73 (1994), 195-208; Harry Frankfurt, "An Alleged Asymmetry between Actions and Omissions," *Ethics* 104 (1994), 620-623; Ishtiyaque Haji, "A Riddle Regarding Omissions," *Canadian Journal of Philosophy* 22 (1992), 485-502; Alison McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency," *Philosophical Review* 103 (1994), 453-488; David Zimmerman, "Acts, Omissions and 'Semi-Compatibilism'," *Philosophical Studies* 73 (1994), 209-223.

otherwise, and then cases of omissions for which we are not responsible *because* we have not been able to do otherwise. Let us consider them in turn.

Frankfurt-Type Cases purport to illustrate that there are situations where we would hold the agent responsible despite the fact that he could not have done otherwise. Consider the following case: Jones decides to kill the mayor of the town. He carries out his plan to the letter, shoots the mayor who subsequently dies. Unbeknownst to Jones, evil scientists have implanted a device into Jones' brain which, were Jones to decide not kill the mayor (or waver after his decision) would have swayed Jones to kill the mayor anyway. The intuitions about cases of this sort have been almost uniform. Jones is responsible for killing the mayor. At the same time, it has been claimed, Jones could not have done otherwise: he could not have not killed the mayor.

Fischer's Cases of Omissions. Here is a case of an omission for which, Fischer suggests, the agent is not responsible. Jones does not have any fancy mechanism in his brain. He is strolling along the beach when he sees a child struggling in the water. Though he believes he can rescue the child with little effort, he decides not to go to the trouble. The child drowns. Unbeknownst to Jones, had he jumped into the water, the sharks patrolling the beach would have attacked him, so Jones could not have saved the child after all.

Fischer believes that this is a case where Jones is not responsible for failing to rescue the child precisely because he could not have rescued her. Fischer does not deny that Jones is responsible for something. He is responsible for his "failure to *try* to save the child (and his failure to jump into the water, etc.)."[2] But he is not responsible for his failure to save the child.

Assuming that there are no qualms with respect to the intuitions themselves, there is indeed a striking difference between these cases. In the case of the action, we are inclined to think that the agent is responsible for it. In the case of the omission, we inclined to think that the agent is not responsible for it. Yet both cases are similar with respect to the fact that the agent could not have done otherwise, a fact that has traditionally been held to be of great significance in ascribing responsibility. It would

---

[2] J.M. Fischer, "Responsibility and Failure," *op. cit.*, p. 253.

indeed be plausible to agree that this indicates that there is an asymmetry in the responsibility conditions for actions and omissions if one agreed that the cases are similar with respect to the responsibility-engendering condition. Here in outline is the argument that they are not.

I will assume (but not defend the assumption) that the fact that a performance is something the agent has done (in the sense discussed in Chapter VI, section 3) is a necessary but not a sufficient condition for the agent's being held responsible for the action (under that description). In other words, if an agent is held responsible for an action under a description, it follows that it would have been reasonable$_A$ to expect of the agent that she perform the action under that description.[3] I will first argue (section 2) that the cases Fischer takes to support the asymmetry thesis are dissimilar in this respect. In the Frankfurt-type actions, we hold the agent responsible for φing and φing counts as something the agent has done: it is reasonable$_A$ to expect of her that she φ. In the Fischer-type omissions, we do not hold the agent responsible for φing and φing counts as something the agent happens to do: it is unreasonable$_A$ to expect of her that she φ. The fact that the cases are dissimilar in what I assume to be the responsibility-engendering condition (that what the agent is held responsible for is something she has done rather than happened to do) does not yet disprove Fischer's asymmetry thesis. It would if Fischer had to agree with my assumption but, as indicated, I do not offer a defense of it (see footnote 3). What this will demonstrate is the fact that there is an alternative explanation of our intuitions concerning the cases that appeals to the assumption and which does not demand that we postulate an asymmetry between actions and omissions. In support of this alternative explanation I then consider (section 3) an omission which

---

[3] This thought might indeed be the healthy core of what Mackie has called the "straight rule of responsibility" according to which we are responsible for all and only intentional actions (*Ethics. Inventing Right and Wrong* [New York: Penguin Books, 1977]). The rule is too restrictive. We are often held responsible for unintended consequences of our actions, for unintentional omissions, etc. It does not seem implausible to suggest that the notion of it being reasonable$_A$ to expect something of an agent, which as I argued in Chapter VI can replace the notion of a performance being intentional in the understanding of the nature of action, could also replace the latter in the understanding of the performance for which we are responsible. This is a conjecture that needs developing. In particular, I have not offered a systematic treatment of consequences of actions, which would be required before any such hypothesis can claim to be more than a conjecture.

has the structure of a Frankfurt-type action and a case of an action which has the structure of a Fischer-type omission. Indeed, we shall see that our intuitions coincide with the suggested explanation thus vindicating the assumption.

2. The Reconstruction of the Two Types of Cases

Let us compare the two kinds of cases in the apparatus developed.

Frankfurt's Cases. Does the presence of the counterfactual intervener render it unreasonable$_A$ to expect of Jones that he kill the mayor? One might think that it does. After all, given the presence of the counterfactual intervener it is determined that the mayor will die at Jones' hands. It would thus seem that the presence of the counterfactual intervener is systematically correlated with the pf-fulfillment of the expectation that Jones kill the mayor. However, I have argued that the presence of the counterfactual intervener violates Principle III (p. 112). There are exactly two avenues to the mayor's death at Jones' hands envisaged in the example. First, Jones might decide to kill the mayor and so kill him, in which case we are invited to suppose that his decision to kill the mayor (*K*) is systematically correlated with the (agentive) fulfillment of the expectation that he kill the mayor. Second, Jones might not decide to kill the mayor in which case the counterfactual intervener will take over leading Jones to kill the mayor. In this case, we are invited to suppose that the scientist's intervention (*C*) is systematically correlated with the (non-agentive) fulfillment of the expectation that he kill the mayor.

The reason why we are originally inclined to think that the presence of the counterfactual intervener would be systematically correlated with the fulfillment of the expectation that Jones kill the mayor relies solely on the fact that the case is constructed in such a way that either one or the other condition operates. In other words, given the details of the case, we might be tempted to construe condition *K*-or-*C* as a defeating condition. Principle III blocks this move. In view of the fact that we already understand the operation of the existing conditions (here *K* and *C*) the "new" condition *K*-or-*C* does not defeat the reasonableness$_A$ of the expectation that Jones kill the mayor. It is thus reasonable$_A$ to expect of Jones that he kill the mayor despite the presence of the counterfactual intervener.

Given the account of Chapter VI, Jones's following through his decision to kill the mayor is something he has done rather than something he happened to do. Given that it is reasonable$_A$ to expect of Jones that he kill the mayor, that the mayor does get killed at Jones' hands (and that the counterfactual intervener does not *actually* intervene), Jones performed an action of killing the mayor. (If the counterfactual intervener did actually intervene leading Jones to the killing, the intervention would render the expectation unreasonable$_A$, and so we could at most say that his killing the mayor is something he happened to do.[4])

Fischer's Cases of Omissions have a different structure. Here the potential defeating condition with respect to the expectation to save the child, the presence of the sharks, does not counterfactually depend on the agent's decision. It is a condition that operates up front as it were.

In view of the sharks patrolling the beach, it would be unreasonable$_A$ to expect of Jones that he save the child. This is because the presence of the sharks is assumed to be systematically correlated with the pf-frustration of the expectation to save the child. (We are asked to suppose that it is in fact impossible for Jones to do so — he would be attacked before he ever got to the child.)

Is it reasonable$_A$ to expect of Jones that he not save the child? Once again, the answer is negative. The presence of the sharks guarantees that the any expectation not to save the child will be pf-fulfilled. The presence of the sharks is a defeating condition of the second kind with respect to the expectation not to save the child (and of the first kind with respect to the expectation to save the child).

As long as it would be unreasonable$_A$ to expect of Jones that he prevent the sharks from attacking, it is unreasonable$_A$ to expect of Jones that he rescue the child as well as that he not rescue the child. Hence, in view of Chapter VI, Jones' failure to save the child is not something he does.[5]

---

[4] In fact, in this case the defeating condition is severe enough to make it unreasonable$_A$ to expect the performance of the agent under all descriptions.

[5] This echoes Frankfurt's response to the case: "The real reason [why Jones bears no moral responsibility] is that what he does has no bearing at all upon whether the child is saved. The sharks operate both in the

In both these cases, we have arrived at the right kind of judgment about them without presupposing that there is a deep asymmetry between actions and omissions. The difference concerns rather the structure of the cases. In Fischer's cases of omissions, there is a defeating condition which makes it unreasonable$_A$ to expect of the agent that he performs the action as well as that he does not perform it. In the Frankfurt-type cases, on the other hand, the presence of the counterfactual intervener does not count as a defeating condition: only the actual intervention by the scientist would count as such. Since in the actual situation, as it is construed in the Frankfurt-type case, the only defeating condition (the scientist's intervention) does not occur, it is reasonable$_A$ to expect of Jones that he kill the mayor.

3. Frankfurt-Type Omissions and Fischer-Type Actions

The case against the asymmetry between actions and omissions can be strengthened further if we could find examples of Frankfurt-type omissions, for which we would be responsible, and examples of Fischer-type actions, for which we would not. Such examples can indeed be found.

Frankfurt-Type Omissions. Let me begin by illustrating an example of omission that exactly parallels the structure of Frankfurt-type actions.[6]

Brown has an implant similar to Jones'. She is walking along the beach and sees a child struggling in the water. Though Brown cannot swim, she can throw a life jacket but decides not to. The child drowns. Unbeknownst to Brown, had she shown any inclination to try to save the child, the implant would have been activated as a result of which Brown could not attempt to rescue the child after all. As it happens, the implant did not need to be activated. In this case, the intuition seems to be that Brown is morally responsible for failing to throw the jacket to the child even though she could not have done otherwise.

---

actual and in the alternative sequences, and they see to it that the child drowns no mater what John does" ("An Alleged Asymmetry between Actions and Omissions," *op. cit.*, p. 623).

The case is exactly parallel to Frankfurt-type actions. It might appear that it is unreasonable$_A$ to expect of Brown that she not throw the life jacket, for given the arrangement of the case, the expectation not to throw the life jacket will be systematically pf-fulfilled. However, as before in the Frankfurt-type case, there is in fact no defeating condition at work. The alleged defeating condition is a composite of (a) the decision not to throw the life jacket *D* which is supposed to be systematically correlated with the (agentive) fulfillment of the expectation not to throw the jacket, and (b) the scientist's possible intervention *C* which is systematically correlated with the (non-agentive) fulfillment of that expectation. It is because the case is so constructed that either *D* or *C* will occur that we might think a defeating condition is in place. In virtue of Principle III, however, *D*-or-*C* does not qualify as a defeating condition. Hence, despite *D*-or-*C* it is reasonable$_A$ to expect of Brown that she not throw the life jacket. By similar reasoning (which exactly parallels the Frankfurt-type case), despite *D*-or-*C* it is reasonable$_A$ to expect of Brown that she throw the life jacket. The situation would change were the scientist to intervene. The scientist's actual intervention would qualify as a defeating condition. It would no longer be reasonable$_A$ to expect of Brown that she throw the life jacket if the scientist intervened.

Because in the actual situation, the scientist does not intervene, it is reasonable$_A$ to expect of Brown that she not throw the life jacket. So, when in the actual case, she does not throw the life jacket, while the counterfactual intervener does not intervene, her not throwing the jacket is something she does.

Fischer-Type Actions. It is in general more easy to describe a Fischer-type omission than action but perhaps the following example will bring the point home. It does not involve a counterfactual intervener. Smith wants to switch on the light. He presses the switch. The light comes on. It might appear that Smith switched on the light. However, unbeknownst to Smith, the light would have come on at exactly the moment that Smith actually pressed the switch, but independently of Smith's intervention. It seems intuitive to describe the case as that of Smith having nothing to do with the light going on. (Indeed Smith could

---

[6] The case is borrowed from I. Haji, "A Riddle Regarding Omissions," *op. cit.* A similar case is

not have done otherwise: he could not have not switched the light.) It would be inappropriate to hold Smith responsible for switching on the light. Smith might still be held responsible for flipping the switch, but not for actually switching the light on. This is indeed borne out if we ask whether it was reasonable$_A$ to expect of Smith that he switch on the light.

Given that the light will come on at *t*, is it reasonable$_A$ to expect of Smith that he not switch on the light at *t*? It seems clear that the expectation would be unreasonable$_A$. The fact that light will come on at *t* is systematically correlated with the pf-frustration of the expectation that Smith not switch on the light at *t* and with the pf-fulfillment of the expectation that Smith switch on the light. Hence, it was unreasonable$_A$ to expect of Smith that he switch on the light at *t*.

4. Final Remarks

The asymmetry thesis concerns the asymmetry in ascriptions of responsibility. I have not defended any view regarding the conditions of responsibility. I have rather followed a simpler strategy. I have assumed that it is necessary for an agent's being responsible for φing that φing count as something the agent has done (i.e. that it was reasonable$_A$ to expect of the agent that she φ). I have then shown that the assumption allows us to understand the difference in our dispositions to hold the agent responsible in the cases of Frankfurt-type actions and Fischer-type omissions (section 2). If the assumption is correct this has nothing to do with the fact that the former are actions and the latter are omissions, but rather with the fact that in the former cases we can say that the performance is something the agent does while in the latter it is only something the agent happens to do. I have then vindicated the suggestion by showing that we would be inclined to hold the agent responsible in the case of a Frankfurt-type omission, while we would not hold the agent responsible in the case of a Frankfurt-type action (section 3). This supports the view that there is no fundamental difference between actions and omissions with respect to responsibility ascriptions.

---

constructed by H. Frankfurt, "An Alleged Asymmetry between Actions and Omissions," *op. cit.*

APPENDIX B.

ACTION AS A PERFORMANCE INTENTIONAL UNDER A DESCRIPTION

The currently most popular answer to the question whether anything has been done appeals to what might be called the criterion of intentionalness. The criterion was first proposed by Anscombe[1] and later adopted by Davidson[2]:

(I) An event *e* is an action if and only if *e* is intentional under some description.

I will argue that while (I) might be useful for those theorists of action who aim to understand the category of intentional movements, it must be rejected by anyone aiming to understand the category of action as a unit of conduct.

I begin by sketching six distinct ways in which (I) may be understood (section 1). In section 2, I discuss the ramifications of the most plausible (non-reductive explicatory) reading of (I). In section 3, I argue that the considerations raised about the non-reductive explicatory reading of (I) actually constitute a reason to take the reductive explicatory readings of (I) as being either circular or faulty.

1. A Methodological Prelude

The first point that ought to be raised about (I) concerns its status. (I) might be taken to constitute an *analysis* either of the concept of action or of the concept of being intentional under a description. Alternatively, it could be thought of as not analyzing but rather as *reporting a conceptual connection*. It seems undeniable that Anscombe does

---

[1] *Intention* (Ithaca: Cornell University Press, 1957).
[2] "Agency," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 43-61.

not intend (I) as a reductive analysis. It is deniable, though there are good reasons not to deny, that Davidson likewise does not intend (I) to be reductive.[3]

However, both the reductive and the non-reductive interpretation of (I) can come in a number of flavors. On the reductive side, (I) functions as something of a definition and can correspondingly be understood in at least three ways: as a conventional definition (RC), as an analytic definition (RA) and as an explication (RE).

(RC) It may be understood as *stipulating* that one concept is to be understood in terms of the other. In this case, the definition is purely conventional, and it does not depend on any prior usage of the definiendum. Such a definition cannot be criticized on cognitive grounds, for it coins a new concept. It seems rather clear that (I) is neither meant nor taken as a conventional definition.

(RA) It may be understood as an *analytic definition.* In this case, the concepts are supposed to be well-established with clear areas of application. The purpose of the analytic definition is to analyze one concept in terms of the other; it is to suppose that the definiendum can be understood in terms of the definiens, where the definiens is treated as logically prior to the definiendum.[4] In general, an analytic definition is subject to criticism if the extensions (or the intensions) of the concepts are not identical, as well as if the areas of vagueness or imprecision do not overlap. An example of an analytic definition could be[5] "A bachelor is a married man," or "Mental states are (nothing but) physical states."

(RE) Finally, and most plausibly, (I) may be understood as an *explication.* In such a case, the explicandum is treated as a concept whose intension or extension is sharpened, clarified or illuminated by explicating it in terms of the explicans. Unlike an analytic definition, an explication requires neither that the areas of

---

[3] See Donald Davidson, "Freedom to Act," in *Essays on Actions and Events, op. cit.*, pp. 63-81.

[4] It is hard to escape the impression that analytic definitions are also in some sense conventional. After all, if the two concepts are well-established, then the claim that one ought to be understood in terms of the other rather than the other way around seems purely conventional or arbitrary. It is only if they are embedded in a larger explicatory project that they may be useful. In such a case, the ordering of concepts is set by the explicatory definitions.

[5] "Could be" because the statement is probably better construed as reporting a pre-existing conceptual connection, see (NA) below. It would be an analytic definition if one were to treat the concepts of being married and being a man as somehow more basic than the concept of a bachelor (see footnote 4).

application of both concepts overlap completely nor that their intensions match. In this respect, an explication is partly stipulatory. Unlike a conventional definition, however, an explication does rely on some overlap between extensions and intensions of the concepts. In this respect, an explication is partly responsive to existing conceptual connections. As such an explication is criticizable, although its value is supposed to lie in how well it functions within a system of explications, in the extent to which it introduces conceptual order.

A famous example of an explication is the statement "Water is H_2O." Unlike a conventional definition, it does rely on preexisting conceptual connections. Some of the conceptual connections involving "Water" are preserved when "H_2O" is substituted (providing other substitutions are made). Unlike an analytic definition, there is no pretense that the intension and extension of the concepts is the same, so that many of the conceptual connections are either rejected or stand in need of explanation (e.g., "Water is liquid," "Water does not break windows though ice sometimes does," "The water in this lake contains all kinds of dirt and chemical substances").

What all these types of definition have in common is that the definiendum is taken to be of a different logical order than the definiens. But they differ in the extent to which the definition is responsive to preexisting conceptual connections. A reductive-conventional statement does not require any pre-existing conceptual connections. A reductive-analytic statement requires that the conceptual connections overlap entirely. And a reductive-explicatory statement requires a partial overlap.

Construing (I) non-reductively means that we must abandon the idea that one concept is somehow logically prior to another. But a non-reductive reading likewise may be responsive to preexisting conceptual connections to different degrees: it may *establish* a conceptual connection (non-reductive-conventional; NC), it may *report* a conceptual connection (non-reductive-analytic; NA) or it may *partly report* and *partly establish* a conceptual connection (non-reductive-explicatory; NE).

(NC) It may be understood as stipulating that two concepts are to be understood in terms of one another. In this case, the claim is purely conventional. It does not

depend on any prior usage of the concepts. Such a claim cannot be criticized on cognitive grounds, for it coins the conceptual connection.

(NA) It may be understood as reporting a pre-existing conceptual connection. In this case, the concepts are supposed to be well-established with clear areas of application. The purpose of the claim is to report the connection existing between the concepts. If there are any areas of vagueness they overlap.
An example of a statement reporting a conceptual connection is "A bachelor is a married man."

(NE) It may be understood as partly reporting but partly establishing a conceptual connection. This is most likely when the areas of indeterminacy or vagueness of the two concepts do not overlap, so that in some cases, one of the concepts can clarify the other, while in other cases, the clarificatory roles are reversed.

Again, what is common to all these options is that the concepts involved are treated as being of the same order. However, they differ in the extent to which they are responsive to existing connections. A non-reductive-conventional statement does not require any pre-existing conceptual connections. A non-reductive-analytic statement requires that the conceptual connections overlap entirely. And a non-reductive-explicatory statement requires a partial overlap.

2. Ramifications of a Non-Reductive (NE) Reading of (I)

If we give (I) a non-reductive reading, it seems most plausible to construe it as explicatory (NE). It seems obvious that (I) is not intended to *establish* a conceptual connection in the manner of (NC). It is less obvious that (I) is not well construed as simply reporting a conceptual connection as in (NA). In order to show this, it will pay to look at cases where it is plausible to suppose that one concept helps to clarify the other. Among the cases where the concept of being intentional under a description helps our intuitions about the concept of action are cases of negative actions. Among the cases where the concept of action helps our intuitions about the concept of being intentional under a description are cases of spontaneous actions (which philosophers sometimes cast as actions performed only with an intention-in-action, not on prior intention). If so, then to the extent that we want to uphold (I), we ought to recognize it as not simply reporting a

conceptual connection between two concepts, but rather as partly reporting and partly establishing it. Let us see that this is so.

Let us take the class of cases where the concept of action is sharpened by our grasp of the concept of being intentional under a description. The most important cases here are those of negative actions. Prima facie we might think that there are a number of classes of negative actions. Although the following list is not exhaustive it should be sufficiently suggestive. (a) An agent is tempted to eat yet another cookie but has firmly resolved to go on a strict diet. His not eating the cookie (refraining from eating it) is something he does. (b) An agent is obligated to file a report but decides not to do so. In this case his not doing it (intentionally omitting to do it) is something he does. (c) An agent accompanies his friend to a party, where another person attacks his friend in conversation. Our agent does absolutely nothing but without intending to do anything either. He just idly stands there. His failing to come to his friend's help is also something he does (something that his friend will rightfully blame him for). (d) A person oversleeps as a result of which he fails to come to a crucial meeting. Once again, one might think it something the agent does.

Our intuitions regarding cases (a) through (d) are not uniform. It suffices for my purposes here to demonstrate that there are some among us whose intuitions favor the inclusion of all these cases under our actions.[6] There are others, however, who favor the inclusion only of cases (a) and (b). What distinguishes cases (a) and (b) from (c) and (d) is precisely the fact that (a) and (b) are intentional under the negative description of the action, while (c) and (d) are not.[7] Indeed, some authors appeal to the fact that (c) and (d)

---

[6] John C. Hall, "Acts and Omissions," *The Philosophical Quarterly* 39 (1989), 399-408; H.L.A. Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 1968); Steven Lee, "Omissions," *Southern Journal of Philosophy* 16 (1978), 339-354; Patricia G. Smith (Milanich), "Allowing, Refraining, and Failing. The Structure of Omissions," *Philosophical Studies* 45 (1984), 57-67; "Contemplating Failure: The Importance of Unconscious Omission," *Philosophical Studies* 59 (1990), 159-176.

[7] Bruce Vermazen has argued that it is a special feature of negative actions that they need to be intentional (and so intentional under the negative descriptions) on pain of including too many negative actions. If one allowed unintentional negative actions to count as negative actions, i.e. if one allowed that as long as a performance is intentional under some description, it is a negative action under all negative descriptions, the list of negative actions would be endless. In Vermazen's words: "Certainly we don't want to say that a person is not-ψ-ing just in case he is not ψ-ing. ...It won't help much to add the rider 'if the agent is doing something' to this last, since the agent will then be doing far too many negative acts: Andy, as he sits twisting his buttons, would also be not-sweeping the table clear of canapés, not-preparing for a Channel

are not intentional under any description to dismiss any intuitions to the effect that they ought to be classified as actions.

The second class consists of cases where the concept of action has a firmer grip and helps us with the concept of being intentional under a description. It comprises spontaneous voluntary actions done with no reason, ones that it is reasonable$_A$ though not reasonable$_N$ to expect of the agent (see p. 152). Consider a case mentioned by Michael Bratman of spontaneous voluntary action:

> Suppose you unexpectedly throw a ball to me and I spontaneously reach up and catch it. My catching it is under my control and voluntary; it is not just like a mere reflex blinking of my eye. But my action is relatively automatic and unreflective, so it may seem strained to suppose that its etiology must involve a distinctive attitude of intending, given that we are understanding intending largely in terms of its role in planning.[8]

(What is noteworthy here is Bratman's simultaneous appeal to the concept of "being under the agent's control," "being voluntary" and the contrast with its *not* being like a reflex action. This is exactly the path we ought to follow if we were to determine the applicability of the concept of action in terms of the absence of defeating conditions. The performance is an action as long as it is unlike performances that happen in the wrong kind of circumstances (in this case: that result from the operation of a reflex).)

One might worry here that Bratman has a vested interest in applying the concept of being intentional under a description to coincide with his concept of intention, which, as he admits, is shaped by his planning theory.[9] He suggests in effect that cases of this sort ought to be described as "voluntary but neither intentional nor unintentional."[10] So, one could argue that on an alternative understanding of intention and of the concept of being intentional under a description, we would have no problem in qualifying this case as falling right under it. But even if we ignore the virtues of Bratman's account of intention, still the point is that there *are* intuitions on which it is not obvious that

---

swim, not-attempting to cross the Sino-Soviet border, and so on." ("Negative Acts," in (eds.) Bruce Vermazen, Merrill B. Hintikka, *Essays on Davidson* [Oxford: Clarendon, 1985], p. 96).

[8] Michael E. Bratman, "Moore on Intention and Volition," *The University of Pennsylvania Law Review* 142 (1994), p. 1712.

[9] Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

spontaneous actions are intentional under some description. In order to uphold (I), one must dispense with such intuitions, in other words, one must "argue" that we need to extend the concept of being intentional under a description because such cases would not be included as actions, and that they are actions seems intuitively indisputable.

In summary, I have argued that if (I) were to be construed non-reductively, it is most plausible to understand it as partly reporting but partly tightening the conceptual connection between the concepts. Cases of negative actions are ones where our understanding of what it is to be an action is sharpened by appeal to the concept of being intentional under a description. Cases of spontaneous actions are ones where our understanding of what it is to be intentional under a description are sharpened by our understanding of what it is to be an action. In other words, (I) is best understood as a non-reductive explicatory (NE) claim, but neither as a non-reductive conventional (NC) nor as a non-reductive analytic (NA) claim.

3. Circularity or Inadequacy of a Reductive Reading of (I)

It seems clear that if one wanted to construe (I) reductively, the very same sorts of cases discussed above likewise tell for construing it as an explication of either one or of the other concepts. If one were to take (I) as analyzing the concept of action in terms of the concept of being intentional under a description, the cases of mistakes or spontaneous actions would be problematic, for the concept of being intentional under a description does not straightforwardly apply to them. If one were to take (I) as analyzing the concept of being intentional under some description in terms the concept of action, one may object in a similar way that the concept of action is not firm enough with respect to negative actions.

In fact, if the considerations that I invoked against taking (I) as merely reporting a conceptual connection, and in favor of taking (I) as being non-reductively explicatory, are sound, one may argue against any attempt to take (I) to be reductive. For to the extent that it is true that our grip on the concept of being intentional under a description is sharpened by appeal to the concept of action in cases of spontaneous actions, taking (I) as

---

[10] M.E. Bratman, "Moore on Intention and Volition," *op. cit.*, p. 1712.

analyzing (or even as explicating) the concept of action is either question-begging or fails to capture those cases as actions. If one takes the concept of being intentional under a description to apply to spontaneous actions, its use as an explicans of action is circular. For the concept of action is crucially involved in our having a firm enough grip on the concept of being intentional under a description to apply to those cases (see above). Alternatively, if one takes the concept of being intentional under a description to apply only to cases where we have a firm enough grip on that concept (without any enrichment from the connection with the concept of action), the explication will fall short of capturing the concept of action, for it will not straightforwardly apply to spontaneous actions. Thus the explication of the concept of action in terms of the concept of being intentional under a description is either circular or faulty.[11]

Another way of putting the point concerns the "history" of the notion of intention-in-action. Cases of actions that are not done on a prior intention are usually supposed to involve an intention-in-action.[12] If one takes the concept to be involved in a reductive analysis of the concept of action then its status will seem rather peculiar. It is most plausible to suppose that the concept of a performance being intentional under a description applies paradigmatically to cases of actions done on a prior intention, perhaps preceded by a distinct stage of deliberation. If the concept of being intentional under a description were to be limited in application to this class of cases then the class of spontaneous actions, not done a prior intention, would not be covered by it. In order for the concept of action to be properly delimited then one needed to stipulate another notion of intention, which is present even if there is no prior intention on which the agent acts, intention-in-action. Clearly, however, such an invocation of the concept renders its use in the explication circular.

---

[11] These claims are made in abstraction from any further attempts to explicate the concept of being intentional under a description. However, in view of the fact that the literature is full of controversy regarding the concept of intentional action, we might think it pretty safe to say that the concept of being intentional under a description or any of its potential explicanda are far from being precisely settled.

[12] The term is first introduced by G.E.M. Anscombe in *Intention, op. cit.* However, it has since acquired very different interpretations including teleological (George M. Wilson, *The Intentionality of Human Action* [Stanford: Stanford University Press, 1989]), causal (John R. Searle, *Intentionality. An Essay in the Philosophy of Mind* [Cambridge: Cambridge University Press, 1983]) and social-normative (Robert Brandom, *Making It Explicit* [Cambridge: Harvard University Press, 1994]).

Similarly, it could be argued that if (I) were taken to be an explication of the concept of being intentional under a description, the explication would be either circular or inadequate. For to the extent that it is true that our grip on the concept of action is sharpened by appeal to the concept of being intentional under a description in cases of negative actions, taking (I) as analyzing (or even as explicating) the concept of being intentional under a description is either question-begging or inadequate. If one takes the concept of action not to cover unintentional omissions, for instance, its use as an explicans of the concept of being intentional under a description is circular. For the concept of being intentional under a description is crucially involved in our having a firm enough grip on the concept of being an action not to apply to those cases. Alternatively, if one takes the concept of action to apply to cases of unintentional omissions, for instance (thus ignoring the way it is sharpened by the tie with the concept of being intentional under a description), the explication will fall short of capturing the concept of being intentional under a description because unintentional omissions are not intentional under any descriptions.

It thus seems that only two avenues are open if one wants to uphold (I). Either one abandons the reductive project and treats (I) non-reductively, or one treats (I) reductively but abandons some of the intuitions. For instance, if one were to uphold (I) as a reductive explication of the concept of action, one would have to give up the intuition that mistakes are actions, for instance.

4. Why abandon (I)?

An alternative avenue is to abandon (I) altogether. Why? We might first ask why one should adopt (I) in the first place (assuming the most plausible non-reductive explicatory reading). Since, as I argued, it would be implausible to construe (I) as representing even a non-reductive analytic link between the two concepts, the question that is reasonably asked is what it is that makes (I) even prima facie plausible (before one looks to cases where the concepts have a less clear application). The most plausible answer to this question is that (I) holds for cases that are sometimes considered to be

paradigms of actions: intentional movements.[13] Thus, if one takes the category of intentional movements to be central to our concept of action, (I) will be immediately plausible. It will then seem worth the while to sharpen our intuitions on either one (as the reductivist would propose) or on both sides (as the non-reductivist would propose). The acceptability of (I) is thus conditional on the commitment to an investigation of the category of intentional behavior.

In the present project, I have undertaken an investigation of a related but in important ways distinct concept of *conduct*. The concept of conduct covers both intentional behavior as well as omissions (including unintentional omissions). Since one of the costs of upholding (I) involves tightening our concept of action to precisely exclude unintentional omissions, (I) is not even prima facie acceptable to someone who intends to understand the concept of conduct rather than intentional behavior.

A defender of (I) might object at this point that the initial plausibility of (I) ought to not only constitute reason for letting both categories involved be sharpened, but it also ought to throw doubt on the viability of the very project of trying to understand the concept of action as part of conduct rather than as part of intentional behavior. Such a theorist might argue that the intuitive appeal of (I) actually *shows* that unintentional omissions are not actions and so the very precept of the present project is called into question. In other words, it is illegitimate to reject (I) on the basis that it does not capture unintentional omissions, for (I) (fortified by its intuitive plausibility) actually demonstrates that unintentional omissions are not actions.

The objection fails. After all, the reason for thinking that unintentional omissions are not actions relies on their missing the connection with the concept of being intentional under a description. They have other conceptual connections that they share with what we recognize as actions (among them two prominent facts: we are held responsible for them and they form the basis on which we attribute character traits to people). It is not clear therefore that our unsharpened concept of action excludes the foundation of the

---

[13] Recall from the Introduction that the category of intentional movements is extensional. It will thus be roughly coextensive with intentional actions and unintentional actions (which on Davidson's rendition are not extensional categories).

present project. What of the claim that we should reject the thesis that the concept of action covers unintentional omissions on the basis of the theoretical adequacy of (I)? I have already suggested that from the point of view of the present account, (I) does solidify *some* of our intuitions about the concept of action, viz. those that pertain to the idea of action as part of our intentional behavior. But there are other intuitions about the concept of action, viz. those that pertain to the idea of action as part of our conduct. It seems reasonable therefore to treat (I) as offering an adequate account of a narrower category than the concept of action that is under investigation here. But if so then it would be preposterous for however ardent a proponent of (I) to criticize the present account for not being narrow enough.

I have considered in some detail the thesis that to be an action is to be intentional under a description. We have seen that under the most plausible reading the thesis is understood as partly reporting but partly sharpening a conceptual connection that exists between the two concepts. I have suggested that there are cases (of negative actions) where the concept of being intentional under a description is used to sharpen our intuition of what negatively described performances ought to count as actions. It is on this ground that unintentional omissions are argued not to be our actions. And there are cases of spontaneous actions where the concept of action is used to sharpen our sense that there is some description under which they are intentional.

I have also suggested that what makes (I) so intuitively appealing is the fact that it unproblematically reports a preexisting conceptual connection for a range of cases, viz. intentional movements. Most intentional movements count as actions and they have some description under which they are intentional. For a theorist who aims to capture the concept of action understood as a unit of our intentional behavior, (I) will and ought to be a central thesis that is worth sharpening in any areas of unclarity. For a theorist who, as I do in this project, aims to capture the concept of action understood as a unit of our overall conduct, especially one of the costs of adopting (I) will be particularly unacceptable. Accepting (I) means that one has to deny that unintentional omissions and idle negative actions are to count as actions. But, as I insisted at the very outset, an account that attempts to capture the concept of action as a unit of conduct must include those cases. We are thus committed to rejecting (I).

BIBLIOGRAPHY

BIBLIOGRAPHY

Robert Merrihew Adams, "Involuntary Sins," *Philosophical Review* 94 (1985), 3-31.

G.E.M. Anscombe, *Intention* (Ithaca: Cornell University Press, 1957).

G.E.M. Anscombe, "Under a Description," *Nous* 13 (1979), 219-233.

Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis: Hackett, 1985).

Annette C. Baier, "The Search for Basic Actions," *American Philosophical Quarterly* 8 (1971), 161-170.

Annette C. Baier, "Ways and Means," *Canadian Journal of Philosophy* 1 (1972), 275-293.

Annette C. Baier, *Postures of the Mind: Essays on Mind and Morals* (Minneapolis: University of Minnesota Press, 1985).

Annette C. Baier, "Rhyme and Reason: Reflections on Davidson's Version of Having Reasons," in LePore and McLaughlin, *Actions and Events*, pp. 116-129.

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

Kurt Baier, "Moral and Legal Responsibility," in Mark Siegler, Stephen Toulmin, Franklin E. Zimring and Kenneth F. Schaffner (eds.), *Medical Innovation and Bad Outcomes* (Ann Arbor, MI: Health Administration Press, 1987), pp. 101-129.

Lynne Rudder Baker, *Explaining Attitudes: A Practical Approach to the Mind* (Cambridge: Cambridge University Press, 1995).

Nuel Belnap, "Before Refraining Concepts for Agency," *Erkenntnis* 34 (1991), 137-169.

Nuel Belnap and Michael Perloff, "Seeing to It that: A Canonical Form for Agentives," in Kyburg, Jr. et. al., *Knowledge Representation and Defeasible Reasoning*, pp. 175-199.

John Bishop, *Natural Agency: An Essay on the Causal Theory of Action* (Cambridge: Cambridge University Press, 1989).

Robert Boyd and Peter Richerson, *Culture and the Evolutionary Process* (Chicago: Chicago University Press, 1985).

Myles Brand, *Intending and Action: Toward a Naturalized Action Theory* (Cambridge, MA: The MIT Press, 1984).

Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994).

Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

Michael E. Bratman, "Moore on Intention and Volition," *The University of Pennsylvania Law Review* 142 (1994), 1705-1718.

Tyler Burge, "Individualism and the Mental," in Peter A. French, Theodore E. Uehling, Jr. and Howard K. Wettstein (eds.), *Studies in Metaphysics* (Minneapolis: University of Minnesota Press, 1979), pp. 73-122.

Ronald Butler, "Report on Analysis Problem No. 16," *Analysis* 38 (1978), 113-114.

Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983).

Christopher Cherry, "The Limits of Defeasibility," *Analysis* 34 (1974), 101-107.

William Child, *Causality, Interpretation and the Mind* (Oxford: Clarendon Press, 1994).

Roderick M. Chisholm, "Freedom and Action," in Keith Lehrer (ed.), *Freedom and Determinism* (New York: Random House, 1966), pp. 11-44.

Roderick M. Chisholm, *Person and Object: A Metaphysical Study* (La Salle, IL: Open Court, 1976).

Randolph Clarke, "Ability and Responsibility for Omissions," *Philosophical Studies* 73 (1994), 195-208.

G.A. Cohen, *Karl Marx's Theory of History: A Defence* (Princeton: Princeton University Press, 1978).

Rachel Cohon, "Are External Reasons Impossible?," *Ethics* 96 (1986), 545-556.

Rachel Cohon, "Hume and Humeanism in Ethics," *Pacific Philosophical Quarterly* 69 (1988), 99-116.

Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

Donald Davidson, *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).

Lawrence H. Davis, *Theory of Action* (Englewood Cliffs: Prentice-Hall, 1979).

Daniel C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, MA: Bradford Books, 1981).

Daniel C. Dennett, *The Intentional Stance* (Cambridge, MA.: Bradford Books, 1987).

John Dewey, *Human Nature and Conduct* (New York: The Modern Library, 1957).

Joel Feinberg, "Action and Responsibility," in White, *The Philosophy of Action*, pp. 95-119.

John Martin Fischer, "Responsibility and Failure," *Proceedings of the Aristotelian Society* 86 (1985/86), 251-270.

John Martin Fischer, "Responsiveness and Moral Responsibility," in Schoeman, *Responsibility, Character, and the Emotions*, pp. 81-106.

John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control* (Oxford: Basil Blackwell, 1994).

John Martin Fischer and Mark Ravizza, "Responsibility and Inevitability," *Ethics* 101 (1991), 258-278.

Anthony Flew (ed.), *Essays on Logic and Language* (Oxford: Blackwell, 1951).

Harry Frankfurt, "An Alleged Asymmetry between Actions and Omissions," *Ethics* 104 (1994), 620-623.

Harry G. Frankfurt, *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988).

Bernard Gert, "Coercion and Freedom," *Nomos* 14 (1972), 30-48.

Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge: Harvard University Press, 1990).

Carl Ginet, *On Action* (Cambridge: Cambridge University Press, 1990).

Erving Goffman, *Stigma: Notes on the Management of Spoiled Identity* (New York: Simon & Schuster, 1963).

Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970).

Patricia Greenspan, "Behavior Control and Freedom of Action," *Philosophical Review* 87 (1978), 225-240.

Patricia Greenspan, "Unfreedom and Responsibility," in Schoeman, *Responsibility, Character, and the Emotions*, pp. 63-80.

Ishtiyaque Haji, "A Riddle Regarding Omissions," *Canadian Journal of Philosophy* 22 (1992), 485-502.

John C. Hall, "Acts and Omissions," *The Philosophical Quarterly* 39 (1989), 399-408.

Gilbert Harman, "Moral Relativism Defended," *Philosophical Review* 84 (1975), 3-22.

Gilbert Harman, "Practical Reasoning," *The Review of Metaphysics* 29 (1976), 431-463.

Gilbert Harman, "Relativistic Ethics: Morality as Politics," in Peter A. French, Theodore E. Uehling, Jr. and Howard K. Wettstein (eds.), *Studies in Ethical Theory* (Minneapolis: University of Minnesota Press, 1980), pp. 109-121.

Gilbert Harman, *Change in View: Principles of Reasoning* (Cambridge, MA: The MIT Press, 1986).

H.L.A. Hart, "The Ascription of Responsibility and Rights," in Flew, *Essays on Logic and Language*, pp. 145-166.

H.L.A. Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 1968).

H.L.A. Hart and Tony Honore, *Causation in the Law* (Oxford: Clarendon Press, 1985).

John Heil and Alfred Mele (eds.), *Mental Causation* (Oxford: Clarendon Press, 1993).

Carl G. Hempel, "Provisos," in Adolf Grunbaum and Wesley C. Salmon (eds.), *The Limitations of Deductivism* (Berkeley: University of California Press, 1988), pp. 3-22.

Jaakko Hintikka (ed.), *Essays on Wittgenstein in Honor of G.H. von Wright: Acta Philosophica Fennica 28* (Amsterdam: North-Holland, 1976).

Brad Hooker, "Williams' Argument Against External Reasons," *Analysis* 47 (1987), 42-44.

Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980).

Jennifer Hornsby, "Which Physical Events are Mental Events," *Proceedings of the Aristotelian Society* 81 (1980-1), 73-92.

Jennifer Hornsby, "Agency and Causal Explanation," in Heil and Mele, *Mental Causation*, pp. 161-188.

I.L. Humberstone, "Direction of Fit," *Mind* 101 (1992), 59-83.

H.E. Kyburg, Jr., R.P. Loui and G.N. Carlson (eds.), *Knowledge Representation and Defeasible Reasoning* (Dordrecht: Kluwer, 1990).

Marc B. Lange, *The Design of Scientific Practice: A Study of Physical Laws and Inductive Reasoning* (Ph.D. Dissertation: University of Pittsburgh, 1990).

Steven Lee, "Omissions," *Southern Journal of Philosophy* 16 (1978), 339-354.

Ernest LePore and Brian P. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell, 1985).

David Lewis, *Counterfactuals* (Oxford: Basil Blackwell, 1986).

Michael J. Loux (ed.), *The Possible and the Actual: Readings in the Metaphysics of Modality* (Ithaca, NY: Cornell University Press, 1979).

J.L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin Books, 1977).

Hugh McCann, "Volition and Basic Action," *Philosophical Review* 83 (1974), 451-473.

John McDowell, "Functionalism and Anomalous Monism," in LePore and McLaughlin, *Actions and Events*, pp. 387-398.

John McDowell, "Might There Be External Reasons?," in J.E.J. Altham and Ross Harrison (eds.), *World, Mind, and Ethics* (Cambridge: Cambridge University Press, 1995), pp. 68-85.

Alison McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency," *Philosophical Review* 103 (1994), 453-488.

A.I. Melden, *Free Action* (London: Routledge & Kegan Paul, 1961).

Alfred Mele, "Motivational Internalism: The Powers and Limits of Practical Reasoning," *Philosophia* 19 (1989), 417-436.

Denise Meyerson, *False Consciousness* (Oxford: Clarendon Press, 1991).

Stanley Milgram, *Obedience to Authority* (New York: Harper & Row, 1969).

Adam Morton, "Because He Thought He Had Insulted Him," *Journal of Philosophy* 72 (1975), 5-15.

Carlos J. Moya, *The Philosophy of Action* (Cambridge: Polity Press, 1990).

Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970).

Leszek Nowak, "Theory of Socio-Economic Formations as an Adaptive Theory," *Revolutionary World* 14 (1975), 85-102.

Leszek Nowak, *The Structure of Idealization* (Dordrecht/Boston: Reidel, 1980).

Leszek Nowak, "Man and People," *Social Theory and Practice* 14 (1987), 1-17.

Leszek Nowak (ed.), *Dimensions of the Historical Process. Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 13 (Amsterdam: Rodopi, 1989).

Leszek Nowak, *Power and Civil Society: Toward a Dynamic Theory of Real Socialism* (New York: Greenwood Press, 1991).

Robert Nozick, "Coercion," in Sidney Morgenbesser, Patrick Suppes and Morton White (eds.), *Philosophy, Science, and Method* (New York: St. Martin's Press, 1969), pp. 440-472.

George Orwell, *Nineteen Eighty-Four* (New York: Harcourt, 1949).

Christopher Peacocke, *Holistic Explanation: Action, Space, Interpretation* (Oxford: Clarendon Press, 1979).

R.S. Peters, *The Concept of Motivation* (London: Routledge & Kegan Paul, 1958).

Philip Pettit, *The Common Mind: An Essay on Psychology, Society, and Politics* (Oxford: Oxford University Press, 1986).

Philip Pettit, "Humeans, Anti-Humeans, and Motivation," *Mind* 96 (1987), 530-533.

Philip Pettit and Michael Smith, "Backgrounding Desire," *Philosophical Review* 99 (1990), 565-592.

George Pitcher, "Hart on Action and Responsibility," *The Philosophical Review* 69 (1960), 226-235.

Hilary Putnam, "The Meaning of Meaning," in Hilary Putnam (ed.), *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975), pp. 215-271.

Nicholas Rescher, *A Useful Inheritance: Evolutionary Aspects of the Theory of Knowledge* (Savage, Md.: Rowman & Littlefield, 1990).

Nicholas Rescher, *Standardism*, forthcoming.

Ferdinand Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (Cambridge: Cambridge University Press, 1987).

G.F. Schueler, "Pro-Attitudes and Direction of Fit," *Mind* 100 (1991), 277-281.

John R. Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

Holly Smith, "Culpable Ignorance," *Philosophical Review* 92 (1983), 543-571.

Michael Smith, "The Humean Theory of Motivation," *Mind* 96 (1987), 36-61.

Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994).

Michael Smith, "Internal Reasons," *Philosophy and Phenomenological Research* 55 (1995), 109-131.

Patricia G. Smith (Milanich), "Allowing, Refraining, and Failing. The Structure of Omissions," *Philosophical Studies* 45 (1984), 57-67.

Patricia G. Smith, "Ethics and Action Theory on Refraining A Familiar Refrain in Two Parts," *The Journal of Value Inquiry* 20 (1986), 3-17.

Patricia G. Smith, "Contemplating Failure The Importance of Unconscious Omission," *Philosophical Studies* 59 (1990), 159-176.

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

Robert M. Stewart and Lynn L. Thomas, "Recent Work on Ethical Relativism," *American Philosophical Quarterly* 28 (1991), 85-100.

Rowland Stout, *Things that Happen because They Should: A Teleological Approach to Action* (Oxford: Clarendon Press, 1996).

Frederick Stoutland, "Oblique Causation and Reasons for Action," *Synthese* 43 (1980), 351-367.

Frederick Stoutland, "The Causation of Behavior," in Hintikka, *Essays on Wittgenstein in Honor of G.H. von Wright*, pp. 286-325.

Frederick Stoutland, "Von Wright's Theory of Action," in P.A. Schilpp and L.E. Hahn (eds.), *The Philosophy of Georg Henrik von Wright* (La Salle, IL: Open Court, 1989), pp. 305-332.

Steven Sverdlik, "Pure Negligence," *American Philosophical Quarterly* 30 (1993), 137-149.

Richard Taylor, *Metaphysics* (Englewood Cliffs, NJ: Prentice-Hall, 1983).

Sergio Tenenbaum, *The Object of Reason: An Inquiry into the Possibility of Practical Reason*. Ph.D. Dissertation: University of Pittsburgh, 1996.

Judith Jarvis Thomson, "The Time of a Killing," *Journal of Philosophy* 68 (1971), 115-132.

J. David Velleman, *Practical Reflection* (Princeton, NJ: Princeton University Press, 1989).

J. David Velleman, "What Happens When Someone Acts," in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press, 1993), pp. 188-210.

Bruce Vermazen, "Negative Acts," in Vermazen and Hintikka, *Essays on Davidson*, pp. 93-104.

Bruce Vermazen and Merrill B. Hintikka (eds.), *Essays on Davidson: Actions and Events* (Oxford: Clarendon, 1985).

Barbara von Eckhardt, "The Empirical Naiveté of the Current Philosophical Conception of Folk Psychology," delivered at the Central Division of the American Philosophical Association and the Third Meeting of the Pittsburgh-Konstanz Colloquium in the Philosophy of Science (1995).

G.H. von Wright, *Norm and Action* (London: Routledge & Kegan Paul, 1963).

Georg Henrik von Wright, *Explanation and Understanding* (Ithaca: Cornell University Press, 1971).

Georg Henrik von Wright, *Practical Reason: Philosophical Papers, vol. I* (Ithaca: Cornell University Press, 1983).

R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994).

Alan R. White (ed.), *The Philosophy of Action* (Oxford: Oxford University Press, 1968).

Bernard Williams, *Moral Luck* (Cambridge: Cambridge University Press, 1981).

George M. Wilson, *The Intentionality of Human Action* (Stanford: Stanford University Press, 1989).

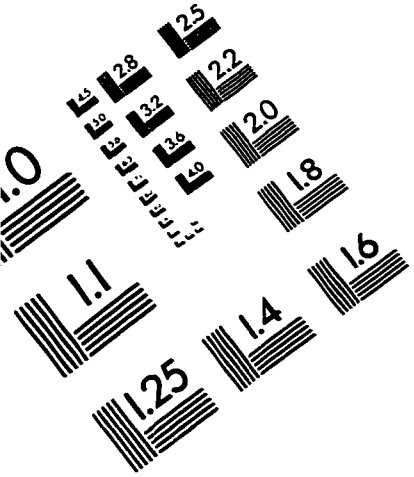
Ludwig Wittgenstein, *Philosophical Investigations* (New York: Macmillan, 1958).

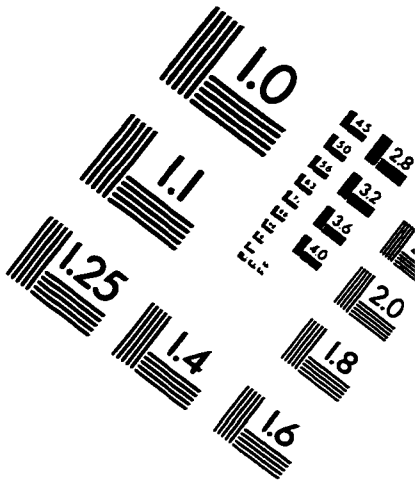
Keith D. Wyma, "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34 (1997), 57-70.

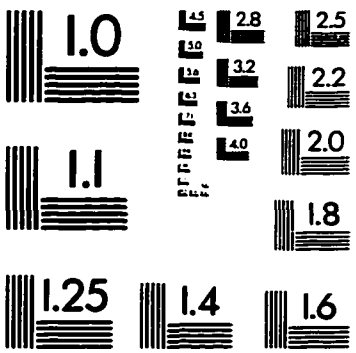
David Zimmerman, "Acts, Omissions and 'Semi-Compatibilism'," *Philosophical Studies* 73 (1994), 209-223.

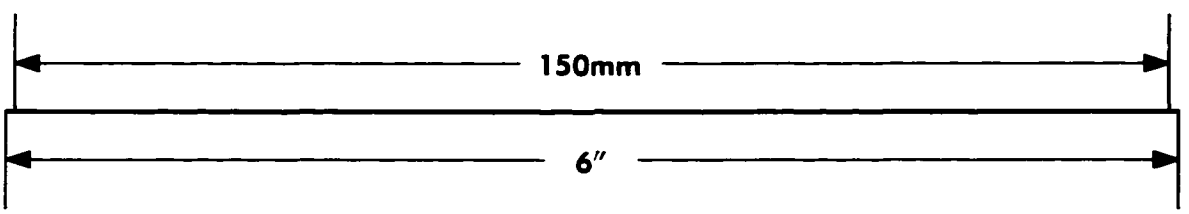
Michael Zimmerman, "Negligence and Moral Responsibility," *Nous* 20 (1986), 199-218.

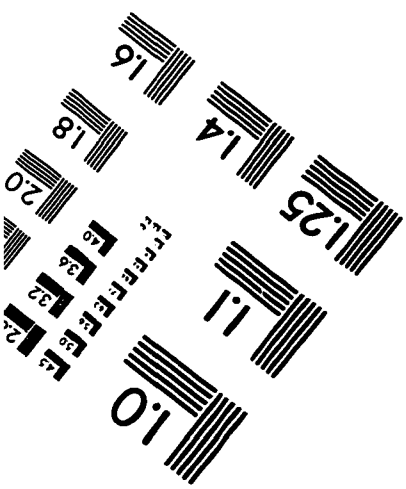
IMAGE EVALUATION TEST TARGET (QA-3)

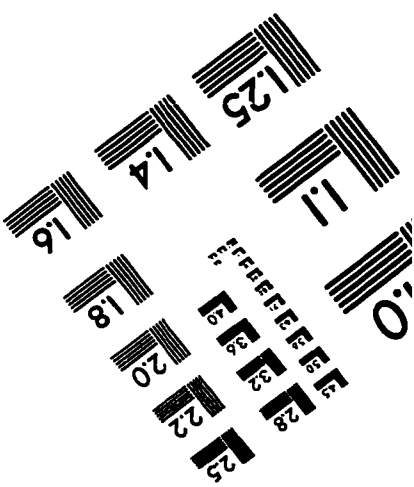












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989