HOW BELIEFS MAKE A DIFFERENCE

by

Susan G. Sterrett

B.S. Cornell University 1977

M.A. Mathematics University of Pittsburgh 1987

M.A. Philosophy University of Pittsburgh 1988

Submitted to the Graduate Faculty of

the Faculty of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

1999

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Susan G. Sterrett

It was defended on

November 15, 1999

and approved by

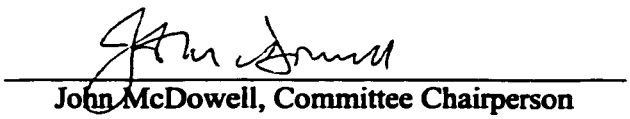
Robert Brandom

Richard Gale

David Gauthier

Rick Grush

Alan Lesgold



John McDowell, Committee Chairperson

John H. McDowell

HOW BELIEFS MAKE A DIFFERENCE

Susan G. Sterrett, PhD

University of Pittsburgh, 1999

How are beliefs efficacious? One answer is: via rational intentional action. But there are other ways that beliefs are efficacious. This dissertation examines these other ways, and sketches an answer to the question of how beliefs are efficacious that takes into account how beliefs are involved in the full range of behavioral disciplines, from psychophysiology and cognition to social and economic phenomena.

The account of how beliefs are efficacious I propose draws on work on active accounts of perception. I develop an account based on a proposal sketched by the cognitive scientist Ulric Neisser. Neisser sketched an active account of perception, on which dynamic anticipatory schemata direct an organism's exploration and action, and are in turn revised as a result of exploration and action. This notion of schema has roots in nineteenth century neurophysiology and in Frederick Bartlett's subsequent work on memory. Neisser appealed to it to unite what he thought was right about information-processing accounts of perception with what he thought was right about ecological accounts of perception. The point that we must anticipate in order to perceive has been recognized by philosophers in the form of the "theory-ladenness of observation." I extend the concept of anticipatory schema to include its role in social perception and social interaction; the concept of anticipatory schema provides a more interactive account of the role of expectations in the maintenance and existence of social institutions, and can be used to enrich the account of convention David Lewis provided. I also show that the concept of rational

expectations, which explains the neutrality of money in terms of the efficacy of anticipatory expectations, is compatible with the proposed account of how beliefs are efficacious.

I discuss how the proposal accounts for the three main modes by which beliefs can be efficacious: (i) via their role in causing intentional action, (ii) via their role in causing economic phenomena and the existence and maintenance of social institutions, and (iii) via their role in causing unintentional physiological responses, including anticipatory physiological responses that can enable perception, cause involuntary actions and give rise to the placebo effect.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND OVERVIEW

Introduction

I chose for my dissertation project a question to which I did not yet have an answer. That question was: "What difference does it make that someone believes one thing rather than another?" The project evolved into answering to the question: "How do beliefs make a difference?"

Dissatisfied with the resources available to answer this question in the philosophy of mind literature to which I'd been exposed, I decided to start by reading William James. Richard Gale led me through James' writings. I read some Dennett, especially The Intentional Stance. Although I frequently use James's and Dennett's views as foils, I am often surprised, in re-reading them now, at how many important distinctions and clarifications I first encountered in reading those two philosophers. Their writings challenged me to think about how one might distinguish habit from mechanism. Reading Wittgenstein's Philosophical Investigations in John McDowell's class on analytic philosophy was an important experience, too; it was in that class that I first felt reassured that one needn't abandon common sense in order to do philosophy.

A course paper on Locke's account of personal identity led me to read experimental and philosophical investigations into memory. In looking for meaningful discussions of memory commensurate with Locke's remarks on the subject, I ended up reading the psychologist F.C. Bartlett's 1932 Remembering, and, later, the philosopher Kathleen Wilkes' more recent Real People. A seminar by John McDowell in the philosophy of mind, much of which later appeared in his "The Content of Perceptual Experience," was especially stimulating.

The answer to the question "How do beliefs make a difference?" presented in this dissertation draws heavily on answers other thinkers have given to different questions: questions about such varied things as perception, action, convention, stereotypes, complex adaptive systems, and rational expectations. In some sense, there is nothing really new here. The intended contribution of the dissertation is in employing various insights from these fields in the service of answering the question I started with.

Overview of the Dissertation

In Chapter II I formulate a more detailed form of the question to which I mean to provide an answer; I distinguish several different ways that belief can make a difference. By considering cases in which the same belief is efficacious in more than one of these ways, we are led to the following requirement: although some physicalist accounts of belief can explain how the holding of a belief could have some physiological effect or other, what's wanted to explain phenomena such as the placebo effect is an account on which the particular effects that actually do ensue can be explained in a way that involves that specific belief.[1] And, the account should apply for effects on social and cultural institutions too, including undesirable effects such as prejudiced actions as well as desirable effects such as giving rise to conditions that enable the ability to communicate succinctly and conditions that enable coordinated actions and cooperative enterprises.
The answer I give to the question of how beliefs make a difference in Chapters III ("Beliefs in Cognition, Perception, and Action") and IV ("Beliefs in Social Interaction") builds on what I find useful in a notion of schema developed by the psychologist-philosopher Ulric Neisser, who, in turn, took the term from Bartlett's Remembering (54). The view Neisser sketched in 1976 in

---

[1] The point is not that the placebo effect presents a contradiction on such a physicalistic view, but that the placebo effect is not taken as something of which such a view aims to give an account, whereas the relation between beliefs and actions is generally taken to be something of which the view aims to provide an account. I discuss the point further in Appendix B "The Placebo Effect and Contemporary Philosophy of Mind".

Cognition and Reality[2] is presented there as an alternative to an information-processing view of cognition, but one that doesn't require the radical rejection of cognitive structures required on J. J. Gibson's view in The Senses Considered as Perceptual Systems. My answer to how beliefs make a difference extends Neisser's notion of anticipatory schemata --- cognitive structures that direct exploration and are revised as exploration progresses --- to obtain a view of belief consistent with the requirements just mentioned: that belief be efficacious in all the various ways that we in fact know belief to make a difference.

Although concerned with very different questions, David Lewis' Convention and David Gauthier's Morals by Agreement and Moral Dealing were helpful in structuring questions about how beliefs are involved in the existence and maintenance of social institutions. I extend Neisser's notion of schema to perception and action in a social context. In extending the notion from perception of inanimate objects to perception of other people, various psychological features of the object (i.e., the person) being perceived and with whom one interacts then become important. Recognition of the importance of one person's expectations of others' expectations led, in turn, to an interest in work by economists, especially Robert Lucas, on inflation, the neutrality of money, and the hypothesis of rational expectations originally proposed by John Muth. What I found notable about Muth's suggestion --- that the best forecasts of certain economic variables are obtained by modeling people's behavior as though they act just as they would if they knew what the econometrician analyzing the model knew --- was that it indicated a need to think of agents as though they found out and responded to changes in their environment more effectively than the models of rational information-processors previously used. The hypothesis of rational expectations is often criticized for unrealistically attributing to agents knowledge they cannot possibly have: i.e., knowledge they would have only after they had "solved their coordination problems" and achieved a rational expectations equilibrium. As I explain in Chapter V, I see such

---

[2] I first read Neisser at the suggestion of a fellow student, Anders Weinstein.

a criticism as an analogue of a certain kind of skeptical attitude about perception: one might as well object to a statement that six-year olds ride bicycles as though they calculate and plan their motions using geometry and theoretical mechanics, by pointing out that it is unrealistic to think that they could have learnt the requisite mathematical theorems by then. Inasmuch as there is some truth to Muth's empirical hypothesis, I see it as reflecting the significance of the interactiveness that in fact exists between an agent and his environment, including his social environment.

There is an interesting philosophical connection between Lewis' work on convention as the solution of coordination problems and Lucas' work on rational expectations that deserves mention here, though: each was trying to capture an insight they attributed to David Hume. Lewis described his theory of convention as "along the lines of" Hume's notion of convention, on which "the actions of each of us have a reference to the other". (4) In "Adaptive Behavior and Economic Theory" Lucas noted that Hume had argued for the neutrality of money --- the view that increasing the amount of money in an economy would have no effect other than to raise prices proportionately --- by reasoning, in effect, that "things that 'ought' not matter to 'rational' people are assumed not to matter in fact" (221-222). Both have, I think, picked up on Hume's recognition of our tendency to anticipate others' behavior. I became aware of this connection to a common insight found in Hume only near the end of writing this dissertation, but it was an affirming revelation: I felt it reinforced the naturalness of extending the importance Hume accorded anticipations in our interactions with our physical environment to our interactions with our social environment as well. Our beliefs make a difference to the things and people with whom we interact by how they affect the attention we pay to them, the information we pick up from them, and how we respond to them, in every way: physiologically, cognitively, and socially.

The multi-modal quality of the anticipatory schemata through which beliefs are efficacious is meant to be construed very broadly: the view I propose here is meant to describe the mental capabilities

of creatures who, though perhaps having the potential of being conditioned to salivate in anticipation of a meal, also blush in embarrassment at an inappropriate automatic response. We are creatures who can follow a path by heart with little cognitive expenditure, as well as figure out an alternate route when that path is blocked, or, even, when we're just curious about finding another way. Whether the activity is using multiplication tables, answering the phone, or avoiding precarious stances, we can not only take advantage of our ability to learn automatic reactions, but can override our training to some extent. Just as memory involves "body memory" in addition to more cognitive abilities, so belief involves all the aspects of the believer's being: physiological, cognitive, and social. That what someone believes makes a difference in all these ways is part of our common sense knowledge of ourselves and others.

CHAPTER II

WAYS IN WHICH BELIEFS ARE CAUSAL

Philosophers of mind have been concerned with how a thinker comes to have the beliefs he or she does, and with how it is that they are about something. These usually involve questions about how beliefs are caused. But causation can go the other way, too. Suppose we ask: what can beliefs cause? One answer is: intentional action. But there are others. Among accounts of belief philosophers have given that do involve this other side of belief, effects other than intentional action are mentioned only incidentally. So, although the causal roles of belief haven't been totally ignored, they haven't been properly taken into account, either.

The notion of cause employed here is a commonsense one: if I can affect Y by changing X, then the change in X has been causally efficacious in effecting a change in Y. This notion of cause certainly applies to beliefs making a difference via rational intentional action: if I can cause more sunscreen to be sold by convincing people that solar radiation is harmful, those people's beliefs about the safety of solar radiation play a role in causing them to purchase more sunscreen. If I can cause a run on sunscreen by convincing people that shortages of sunscreen are expected before August, their beliefs about the future availability of sunscreen play a role in causing them to purchase sunscreen when they do, and may even cause a temporary shortage of sunscreen. The same notion of cause is involved in claims of other modes of beliefs being causal. If I can affect a child's educational development by changing her teacher's beliefs about her capabilities, the belief has been efficacious in improving the child's educational development. Beliefs are efficacious in the establishment and maintenance of social institutions if we can affect the existence of the social institution by changing people's beliefs. If I can cause someone's heart rate to become elevated by getting him to believe that the (placebo) pill he has consumed is a stimulant, his beliefs have played a role in causing his elevated heart rate. Similarly, my thoughts

are causal if I can cause my own heart rate to become elevated by changing what I am thinking about.

The view of how beliefs are causal proposed in this dissertation arose from examining the causal roles of belief observed mainly in the natural context of everyday life, supplemented with a few observations others have made in experimental psychophysical and psychological laboratories. The first step was to ask what belief can---and cannot---do, both with respect to the range of effects, and with respect to the range of ways (modes) in which it can be efficacious. More generally, one could consider the efficacy of things and activities deemed mental (e.g., attitudes, moods, memories recalled, possibilities imagined, thoughts entertained). The reason for focusing on belief is that current theories of mind seem to regard believing as the primary cognitive function, and so focusing on the notion of belief is a way for this investigation of mental causation to make contact with them.

Accordingly, I categorize the ways in which beliefs can be efficacious into three main categories:

1. Via Intentional Action

   as part of a cause of the actions (or, intentional behavior) of an individual having that belief,

2. Via Social Institutions

   as responsible for the maintenance and existence of some social institutions and socially-constituted states of affairs, and

3. Via Unintentional Action and Physiological Changes

   as causal in effecting physical changes (including bodily motions) unmediated by intentional actions of a rational agent.

I mean here to be employing a commonsense notion of belief, as used in everyday discourse.

The first category covers the most commonly recognized way that beliefs can be efficacious: via intentional action. For many philosophers, this is not problematic for their views; they would say that, when an action is explained by citing the reason for doing it, this counts as a causal explanation. On one well-known account (Davidson 1985), the reason for doing the action is the cause of the action; if the reason involves a belief, the relevant belief is, in such a case, part of the cause of the action. The only effects of beliefs accounted for by this mode of efficacy are the intentional actions of individuals.

But belief can also effect group behaviors when the belief is collectively held. The sort of examples usually cited are that of a run on a bank caused by a collectively held belief (i.e., a belief that is held by many different individuals) that the bank is short on liquid assets. When the group behavior is simply the cumulative effect of individuals' behavior, as some have argued that it is for such cases of collectively held beliefs, one could plausibly argue that the mode of the efficacy of the belief that the bank is short on liquid assets is the same as the mode of efficacy in the individual case: i.e., that the group behavior which is effected is mediated by individual intentional behavior and thus that it is still via being part of a reason for an individual's action that the belief is causal. Certainly at least some cases of behavior by a group of people are properly described as mere aggregates of the behaviors of the individuals in the group; in such cases, the mode of the belief's efficacy is simply via the intentional actions of those individuals.

There are cases, however, in which the mode of the belief's efficacy is not obviously reducible to the mode of efficacy via the intentional action of individuals. The second category of modes of efficacy, which I've described above as efficacy via social institutions, is meant to include cases in which the belief is efficacious not solely in virtue of individuals having the beliefs that they have,

but also in virtue of the additional fact that the individuals having those beliefs are in a certain social arrangement, and conceive of and interact with each other in ways that would not be possible otherwise. On the view of beliefs I will propose, some uses of social stereotypes will fit into this category. Another example of a state of affairs in which the efficacy of the belief is via social institutions might be that the practice of contract-making within a social group exists. Such a practice depends on certain people in the social group believing that certain others are trustworthy; it is not sufficient that people perform trustworthy acts or, even, have developed dispositions to be trustworthy. Yet another example of a state of affairs brought about by beliefs that are efficacious via social institutions might be that something is prestigious, for it seems that a thing is prestigious in virtue of people's beliefs about it, and not solely in virtue of intentional actions caused by the belief.

The third mode for beliefs to be efficacious identified above is characterized by the feature that the explanation of the role of the belief in causing the agent's behavior is not in terms of the belief being part of the reason for an action; i.e., the belief is efficacious, but it is not efficacious in virtue of being part of a reason. I've described this as efficacy via unintentional actions or physiological changes (unmediated by intentional action). This mode of the efficacy of belief includes cases of bodily motions or physiological changes that don't involve acting for reasons, such as making a mistake as a result of being preoccupied with the disastrous consequences of making it. Here, the bodily motion isn't an intentional action done for a reason, but is caused, at least in part, by a belief (that the consequences would be disastrous). It could be argued that other mental activities, such as entertaining thoughts of disaster, are also involved in this case; but this is just to say that it is not only beliefs that are efficacious in this way. I think this is a clear case of the mental being efficacious, but in some other cases, it is not so clear whether we have a case of mental causation or simply of reflex.

An example at the other extreme, of being simply reflex, would be closing your eyes as something unexpectedly approaches your eye; here there seems to be little that is cognitive (i.e., that involves mental activity) in the cause of the behavior. This is in contrast to closing your eyes upon seeing something that fills you with horror; in this case, that you experience the thing you see as something horrible is what makes you shut your eyes. A case where things are not so clear is yawning upon seeing someone else yawn; here it is not clear whether or not mental activity is involved in causing the motion. The relevant point is that there are cases where one's beliefs and thoughts are efficacious in causing behavior that is not an intentional action.

Another kind of case that falls under this third mode is that of an individual's belief effecting physiological changes other than bodily motions. A comprehensive characterization of the category would include psychokinesis, but it is not clear that there are any such cases. One may be interested in what would be required of a theory of mind for psychokinesis to be conceptually consistent, as a matter of pure inquiry, but we need not feel constrained to accounts of how beliefs are causal that allow for psychokinesis. Things are different when it comes to the placebo effect, however; we do not have the luxury of deciding whether or not a theory of mind should allow that an individual's belief affect him physiologically, for there are phenomena such as the placebo effect which cannot be ignored.

Some caution should be used in referring to "the placebo effect." There is certainly a lot of evidence pointing to the conclusion that the patient's belief in the efficacy of the treatment being administered is often efficacious in bringing about some physiological effects. However, Adolf Grunbaum (Validation 76) has pointed out that, as used in medical trials, a placebo effect is any effect on the target disorder attributable to factors that are considered (by the therapeutic theory) incidental to the treatment. Grunbaum also points out that whether a treatment factor is a candidate for causing the placebo effect is thus relative to a particular therapeutic theory. Thus, for many therapies, the placebo effect can be due to factors such as the clinical setting, the

attention focused on the patient's progress, or the patient's decision to seek treatment. When sham surgery is used as a placebo, as it sometimes is in determining the efficacy of a surgical procedure, the effects of the sham surgical intervention, which does involve actually making incisions, would also fall under the rubric of a placebo effect. So, not all placebo effects are due to belief, and not all placebo effects due to belief are necessarily due to the belief in the efficacy of the treatment being administered. If researchers did not think that patient belief was very likely among the incidental treatment factors that produce placebo effects, however, they would not go to the trouble and expense of employing blind and double-blind methods of administering the placebo to the placebo group. Thus, I believe it a recognition of the efficacy of belief in effecting physiological changes that results of medical experiments on the efficacy of therapies are not considered valid unless the identity of those receiving the placebo treatment is concealed from both subject and researcher. It is also now recognized that physiological effects (including effects that are not bodily motions) can be induced by hypnosis, and that organic changes as well as functional ones can result from one's beliefs, attitudes, and moods.

This last mode (efficacy via unintentional actions and physiological changes) has been neglected in most discussions in philosophy of mind. [3] And, while few would deny some role of belief in the second mode (efficacy via social institutions), it likewise goes unmentioned in most discussions

---

[3] Since I wrote the paper in which I first drew the characterization of the three modes of efficacy of belief (1989), the topic has become more reputable in the scientific community. There has been a plethora of books on the topic of the effect of the mind on the body since then; the contributor list to Bill Moyers' anthology Healing and the Mind reflects the coincidence of increased public interest and increased scientific respectability of the topic. This, of course, carries no philosophical weight, but I do take it as a point in favor of the genuine existence of the phenomena of the physiological effects of belief that medical researchers, who would have preferred to discredit the effect, found that they had to account for it in order to conduct meaningful research into the medical efficacy of drugs. This is in contrast to some other claims about mental efficacy, such as ESP, that have not withstood such skeptical investigation: precautions against such modes of causation are generally not employed in trials to determine the efficacy of medical treatments.

about the efficacy of belief. The first mode (efficacy via intentional action) has received much attention, but is often mixed with other philosophical commitments (such as a commitment to materialism or physicalism) so that what taking the operation of rationality seriously without such commitments requires of a theory of mind is often obscured.

An account of belief should allow for all three of the modes of efficacy identified. In the next chapter, I sketch an account of how beliefs are efficacious that is neutral with respect to various competing accounts of what belief is. The account I sketch does provide a constraint on candidate notions of belief, however.

CHAPTER III

BELIEFS IN COGNITION, PERCEPTION, AND ACTION

Though accounts of how beliefs are efficacious seldom center on remembering per se, few would dispute that, whatever beliefs are, and however they are involved in an organism's behavior, a thinker's beliefs reflect, and are affected by, his or her experience. So it should not seem out of place that the view I propose, on which beliefs are efficacious via being incorporated in the dynamic anticipatory schemata of individuals, derives from a notion that arose in trying to capture the nature of cognitive structures employed in remembering.

My proposal draws on Ulric Neisser's notion of schema.[1] Neisser explained his view of how a perceiver's unique schema was involved in perception and cognition in his 1976 work Cognition and Reality. In doing so, he drew on F. C. Bartlett's use of schema in his now classic 1932 treatise on memory, Remembering. In this chapter, I first describe the features of Neisser's notion of schema that I think get things right about cognition and so plan to employ in answering the question of how beliefs are efficacious. I then supplement that account with features drawn from William James' insights about the role played by both the senses and "the ideational centres" in attention. The notion of anticipatory schema that results involves what we would call physiological aspects as well as cognitive aspects of a person; an anticipatory schemata not only enables perception, but may guide one's actions as well.

---

[1] I did not start out knowing of Neisser's work and then decide to apply it to the problem at hand. Rather, while I began working out a notion of belief as expectation, Neisser's Cognition and Reality was brought to my attention by Anders Weinstein.

A. The Concept of Dynamic Anticipatory Schemata

1. Neisser on The Perceiver's Contribution to Perception

Although Neisser reached back to Bartlett's 1932 work for inspiration, the approaches he was concerned to counter in 1976 were of course those that had developed in the forty-odd intervening years. It may help in understanding Neisser's notion of schema to describe the views of perception to which he was supplying an alternative. Neisser presents his view as incorporating valuable insights from two opposing approaches in psychology: the notion of information pickup found in Gibson's ecological approach to perception, and the notion of information-processing found in cognitive psychology. Gibson's explanations focused on the structure of the environment in explaining behavior; he avoided hypothetical cognitive constructs, such as the internal structure of a perceiver, on principle. So, for example, Gibson thought to explain the perception of permanent characteristics of one's environment (e.g., the walls and stationary furniture of a room in which one is walking about) in terms of invariant properties of what he called the optic array (the whole range of available light in one's ambient environment from which one can pick up information) as one moved around; the changing features of one's envrionment would be explained in terms of variant properties of the optic array as one moved around. Rather than explaining perception as complicated processing of information, he explained perception as the picking up of variant and invariant structural features of one's ambient environment. This, Neisser thought, tended to underappreciate what the perceiver has to do with perception.[5]

---

[5] John Heil takes a similar approach in Perception and Cognition; he writes: "I share Gibson's distrust of theories that take perception to be essentially a matter of mental synthesis and construction, theories founded on the conviction that, in perceiving, one is obliged to assemble inside one's head a coherent and unified structure from an inchoate sensory 'input.' At the same time, I wish to call into question the Gibsonian notion that perception does not invove a cognitive component essentially. (p. xi)" Heil is concerned to show the sense in which these cognitive

Information processing views of a perceiver tend to the opposite kind of imbalance: though they do appreciate the contributions of the perceiver, they often "pay little heed to the kind of information the environment actually offers" (53). Neisser's approach, though markedly different from either the ecological approach or the information-processing approach, grants that what each of the two views focuses on is important: Perception is the interaction of something properly regarded as a contribution of the perceiver with his environment: "the perceiver has certain cognitive structures, called schemata, that function to pick up the information that the environment offers" (p. xii).

Neisser discussed how schemata were employed in perception and cognition. Because my topic here is how beliefs are efficacious, I am interested in how the notion of schema would be employed in action as well. I do not think such an interest much of an extension, however, for Neisser grants that "action is organized just like perception, guided by expectancies that in turn are altered by consequences" (52-53). Since he is using a notion that was developed from a notion originally drawn from a neurologist's account of action, it is not surprising that the notion applies to action rather smoothly. In fact, Neisser's discussion of schemata actually does involve action, in that he cites each of perceiving, reasoning, and acting as a phase of a cycle that occurs during perception. He even says that, in some respects, "perceiving is a kind of doing." And, he discusses how schemata are employed as one navigates a certain landscape, recognizing landmarks and attempting to reach a specific place. However, for Neisser, perceiving is generally distinguished in this: ". . . [with some exceptions] the perceiver's effects on the world around him are negligible; he does not change objects by looking at them or events by listening to them" (52).

---

structures are properly regarded as representational; my concern is in investigating the ways in which they are causal. Heil's main conclusion is that "beliefs are connected essentially to patterns of behavior, even though they are not in any sense reducible to such behavior. It is in virtue of these connections that beliefs have whatever content they have, it is here that the notion of mental representation finds a place. (219)"

The perceiver, however, is not unchanged by perception. Neisser's account reflects this in that the cognitive structure he calls a schema is modifiable by perception. Here is his attempt to define what he means by a schema:

> A schema is that portion of the entire perceptual cycle which is internal to the perceiver, modifiable by experience, and somehow specific to what is being perceived. The schema accepts information as it becomes available at sensory surfaces and is changed by that information; it directs movements and exploratory activities that make more information available, by which it is further modified. (54)

Although he makes some general remarks about what a schema might be "from the biological point of view," schemata are characterized "in relation to the perceptual cycle of which they are only a part" (54-55). In saying that a perceiver's schemata are only part of the process, Neisser means to emphasize that the environment (as well as the perceiver) is part of the process of perception. Thus in understanding his claims about what a schema contributes to perception, it should be remembered that perception is not to be understood as the operation of a self-standing schema applied to some "input" or other contributed from outside the perceiver. Rather, schemata, whatever they may be, are those aspects of the perceptual process attributable to the perceiver.

Since my own first attempts at developing an account of the efficacy of beliefs via schemata involved adapting Bartlett's account of the cognitive structures involved in remembering as well, it will be helpful to describe some of Bartlett's work first.

2. Bartlett's Notion of Schema in Remembering

Bartlett's account arose out of dissatisfactions with the work of his mentor Ebbinghaus. Ebbinghaus had tried to make the study of memory more scientific by controlling the conditions under which memories were formed and recalled; he used nonsense syllables in order to reduce the influence that a particular person's interests and knowledge might have on his remembering processes. Bartlett criticized Ebbinghaus for studying the processes of recall and recognition in isolation, and for concentrating exclusively on what occurs at the moment of recall or recognition, rather than examining what precedes that moment. Bartlett instead studied remembering in natural contexts; his research revealed how large a part a particular person's interests and knowledge play in remembering processes. He also thought it important to examine the cognitive functions of perceiving in conjunction with those of recognizing and recalling. Past experience affects all these cognitive activities, but, Bartlett came to believe, past experience doesn't show up as anything that could be construed in terms of individual memory traces, as some researchers had thought. Rather, "the past operates as an organised mass rather than as a group of elements each of which retains its specific character" (197). This "organized mass" is continuously changing to incorporate the thinker's ongoing experiences. As with so many since, Bartlett settled on the term 'schema' in spite of feeling it unsatisfactory; he said that he strongly disliked the term 'schema' because "it does not indicate what is very essential to the whole notion, that the organised mass of results of past changes [. . .] are actively doing something all the time" (201). Bartlett preferred the term 'organised setting', but still found the term 'schema' useful when clarified as follows: " 'schema' refers to an active organisation of past reactions, or of past experiences, which must always be supposed to be operating in any well-adapted organic response" (201). These past experiences and reactions are synthesized, rather than recorded or preserved: the resulting synthesis is to be thought of as, not a "patchwork", but a collection of "living, momentary, settings."

This notion of a "living, momentary, setting" was in turn inspired by a piece of neurological research that struck Bartlett as containing an idea that could explain the phenomenon of remembering (which, he felt, was of a distinctly different character than recognizing). The neurologist Henry Head's research on patients with various pathologies led him to reject the then-current explanation of how normal individuals can so effortlessly carry out skilled movements in which "every movement is carried out as if the position reached by the moving limbs in the last preceding stage were somehow recorded and still functioning". The presumption at the time was that the explanation must be that "a preceding movement produces a cortical image [of movement], or trace [of movement], which, being somehow re-excited at the moment of the next succeeding movement, controls the latter" (Bartlett 198). Head noted that some patients were able to identify the positions of their bodies with their eyes open, but, once their eyes were closed, could not tell if their limbs had been placed differently. They could, however, tell exactly where on their skin surface they were being touched. (Case studies of similar phenomena are recounted in William James' Principles (1123ff); James cites cases showing that the sense we have of our limbs' positions cannot be due to feelings of innervation, in arguing against the existence of feelings of innervation.) Thus, Head concluded, it cannot be the case that the ability to "image" one's bodily position necessarily includes an appreciation of relative changes in posture, or "the capacity to relate serial movements" (199). Thus he rejected the explanation of skilled actions in terms of stored traces of movements. Rather, he said, "[past impressions] form organised models of ourselves which may be called schemata" (quoted in Bartlett 200).

Bartlett thought Head had an important insight, though he seems to have objected to the talk of stored traces or models. Bartlett stripped Head's insight free of ontological commitments he found objectionable, to arrive at a notion of schema he thought could be fruitful in experimental psychology, especially in his research on remembering:

> 'Schema' refers to an active organisation of past reactions, or of past experiences, which must always be supposed to be operating in any well-adapted organic response. That is, whenever there is any order or regularity of behavior, a particular response is possible only because it is related to other similar responses which have been serially organised, yet which operate, not simply as individual members coming one after another, but as a unitary mass. Determination by schemata is the most fundamental of all the ways in which we can be influenced by reactions and experiences which occurred some time in the past. (Bartlett 201)

One's experiences are not retained intact as such; rather, Bartlett proposed, "All incoming impulses of a certain kind, or mode, go together to build up an active, organised setting: visual, auditory, various types of cutaneous impulses and the like, at a relatively low level; all the experiences connected by a common interest: in sport, in literature, history, art, science, philosophy and so on, on a higher level." What's important in extending this notion to psychology is that an organism's experiences affect its ability to generate a certain response, not in virtue of being individual events in a chronological sequence, but in virtue of being constituents synthesized into what Bartlett called "living, momentary settings belonging to the organism" (201). He illustrated how such "organised settings" determined the skilled responses generated by a tennis player during a match: the tennis player may think that he is reproducing "a series of text-book movements" but, says Bartlett, that is not what is happening:

> How I make the stroke depends on the relating of certain new experiences, most of them visual, to other immediately preceding visual experiences and to my posture, or balance of postures, at the moment. The latter, the balance of postures, is a result of a whole series of earlier movements, in which the last movement before the stroke is played has a predominant function. When I make

> the stroke I do not, as a matter of fact, produce something absolutely new, and I never merely repeat something old. The stroke is literally manufactured out of the living visual and postural 'schemata' of the moment and their interrelations. (201-202)

It was the idea that a constantly changing "organised mass" of past experience was employed in action that Bartlett wanted to capture, in developing a notion of schema analogous to that found in physiology. But Bartlett speculated further: he speculated that remembering is a case, not of an organism's being influenced by the past, but of "turning round upon its own 'schemata.' In effect, the organism is saying "This and this and this must have occurred, in order that my present state should be what it is. " Bartlett adds, "I believe this is precisely and accurately just what does happen in by far the greatest number of instances of remembering" (202). There are two features of the phenomenon of remembering that he wanted to be true to: first, that in remembering we are influenced by the past in the form of some organized whole, and, secondly, that we know how to be somewhat selective in how we are determined by the past, so that we can "rove more or less at will in any order over the events which have built up [our] present momentary 'schemata'" (203). To satisfy both these requires that "the 'schema' must become, not merely something that works the organism, but something with which the organism can work" (208). Inasmuch as we may speak of "traces', the traces constantly change and reflect the organism's interests, not just the environment in which it has been living. Memories can thus vary from person to person, "because the mechanism of adult human memory demands an organisation of 'schemata' depending upon an interplay of appetites, instincts, interests and ideals peculiar to any given subject" (Bartlett 213).

What we will be appealing to in Bartlett is a notion of schema as a constantly changing, interest-influenced organised mass reflecting a specific organism's past experiences, which it employs in its activities, including action and perception. Bartlett remarks that his view brings remembering

into line with imagining. I would add that there are some features ascribed to remembering on his view that I think fit perceiving and acting. First, because remembering is, on his account, an imaginative reconstruction constructed in part by our own personal schemata, it is never exact, but neither is it important that it be so. We develop the schemata we need to enable us to follow our interests. Secondly, remembering involves being able to use our schemata, not just be bossed around by them; the same is true for perception. "Theory-laden" perception may be, but, just as we can improve our reconstructed memories by examining what we know, we can in many perceptual activities, from wine tasting to map reading, examine and revise the "theories" with which our perceptions are laden, and thereby enhance our perception. Finally, Bartlett remarks that the reconstruction that occurs in remembering is often constructed beginning with a small detail, upon which the memory is then built, in conjunction with employing the 'whole active mass of organised past reactions or experience'; the same could be said of perception (perceiving a particular object often begins with a small detail distinctive of that kind of object "catching our eye"), except that in perception the schemata also guide us in picking up more details from the current situation.

3. Neisser on Anticipatory Schemata Employed in Perception

Just as no two people's memories are alike, so each perceiver's schema is unique. The process whereby a perceiver interacts with his environment, according to Neisser, is a cycle that proceeds roughly as follows: the perceiver's schema directs exploration. During the perceiver's exploration, the object (available information) is sampled; the perceiver's schema is modified as a result of the information just picked up, and the cycle repeats itself as the modified schema directs further exploration.

During the step in the cycle in which the schema directs exploration:

> At each moment the perceiver is constructing anticipations of certain kinds of information, that enable him to accept it as it becomes available. Often he must actively explore . . . by moving his eyes or his head or his body. These explorations are directed by the anticipatory schemata, which are plans for perceptual action as well as readinesses for particular kinds of optical structure. (Neisser 1976, 21)

In the next parts of the cycle, the perceiver (employing his particular, unique, schema) picks up information during his explorations; the schema employed in doing so is modified by the information picked up. The cycle is then repeated as the perceiver employs the newly modified schema in further explorations.

So perception could be called a constructive process, in that the perceiver's active exploration of his environment results in a modification of what could be called an internal cognitive structure. Neisser says that schemata tend to become tuned to the object being perceived. However, it is not true that the perceiver is constructing an image, or an internal representation of the object perceived. The schema, he says, is not to be thought of as what is perceived, any more than a gene is to be identified with a characteristic of an organism. It is real objects, not internal representations of them, that the perceiver perceives.

Not only is the schema not a "percept", it does not produce "percepts", either. In fact, says Neisser, "I submit that perceiving does not involve any such things as 'percepts'" ("Perceiving" 93. This rejection of percepts is pretty obviously directed against information-processing models of cognition, and is not unique to Neisser. However, besides citing the usual problems associated with information processing accounts of vision (e.g., "To *see* a unicorn is to have one's retina stimulated by unicorn-shaped rays of light and to process the resulting detector activity. . .

To *imagine* a unicorn is to ... begin the processing a little further along. How, then, do we know whether we are seeing or imagining one?") Neisser raises another problem, one that I think can be seen as motivation for the anticipatory character of his proposed alternative: "[H]ow would we go about *looking for* unicorns if we wanted to see them? The [information-processing] model makes no provision for perceptual search" (91). He wants to shift the goal of explaining how perception occurs from a concern about the (relatively rare) occurrence of illusions to explaining that the accuracy of ordinary seeing occurs "despite the inadequacy of every momentary retinal image." He notes that visual perception is not a discrete event, but a continuous activity. I think it important that he takes looking to be part of the act of seeing:

> . . . visual perception is a continuous activity. We look at things over extended periods of time, through many fixations. For this reason, looking must involve the anticipation of information as well as its pickup. I suggest that it depends on certain crucial internal structures, or "schemata," that function as anticipations and plans. It is these schemata, together with the information actually available in the environment, that determine what is seen. (Neisser, "Perceiving" 92)

Such anticipations are involved in all perception. For instance, listening for a phone number or a price involves anticipations that enable one to pick up the specific numbers spoken, but these anticipations are not of any number in particular. The schemata or plans may also involve motor activities such as turning one's head, but they need not. In explaining that anticipations are involved even in perceiving things that appear suddenly and unexpectedly, he describes the schema involved in perceiving a visitor to his office of whom he has no forewarning:

> . . . my visitor would not find me perceptually unprepared. After all, he must appear in the doorway. If I am working in my office, I already know where the doorway is, and what lies beyond it, just as I know the location of other familiar objects. This

> means that I can anticipate the distances and possible motions of any arriving guest. Information about his location and movements fits into a preexisting spatial schema, or cognitive map, and thereby modifies that schema. A visitor who entered through the wall, or materialized in the middle of the room, would be more like a ghost than a person. (96)

This example illustrates how a schema is like and unlike a format. Like a format, the schema specifies the sort of information that will be picked up; unlike a format, the schema does not allow a sharp distinction between form and content: "[In a schema, the] information that fills in [what functions like] the format at one point in the cycle becomes a part of [what functions like] the format in the next, determining how further information is accepted" (Cognition and Reality 98). Thus, details of a particular person such as a distinctively shaped shock of red hair, might at one point fill in my schema for perceiving a person in the doorway; later, after this experience, my plan for locating that person in an airport crowd might include being poised to sight that shock of red hair, and the person's location would be a detail my schema enables me to pick up.

Perception is of particulars, and schemata enable a perceiver to perceive a particular object, in a broad sense of the term 'object'. An example of the employment of a schema may help here. In the case of perceiving a face, the schemata employed depend on what it is one is trying to perceive about the face in front of him:

> It takes time to perceive *any* aspect of an object, whether it be the meaning of your brother-in-law George's smile or the relative lengths of his mouth and his eyebrows. Your schemata develop differently in those two cases, and you execute different exploratory eye movements that make different information available. In one case you look for and find additional facial evidence of smiling, certain patterns of movement that characterize smiles over time, and --- over

> longer periods --- more actions by George that reflect the same feelings on his part. In the other case, you might look for information specifying, say, whether the ends of his mouth reach nearer to the edge of his face than the eyebrows do. (Neisser , Cognition 72-73)

What one perceives cannot thus be characterized as the information processing view puts it, i.e., in terms of a single instantaneous input and its processing by the perceiver. But neither is what is perceived to be explained as simply a picking up of some feature or features in the perceiver's environment, as Gibson tends to put it. For, what it is that the perceiver perceives (meaning of a facial expression versus shape of a facial feature) depends on the whole cycle involved in information pickup.

B. Beliefs as Causal via Dynamic Anticipatory Schemata

1. Anticipatory Schemata in Physiological Responses, Action and Social Interaction

The account just given of what a schema is and how it is employed is that of a schema in the context of perception. Is there an analogue of such a schema in the context of action? In social interaction? In physiological responses that are unmediated by intention (e.g., as in the placebo effect)? At one time, I thought so, and tried to develop the analogies. However, I then realized that, rather than an analogue of a perceiver in each of these three different contexts, what we have is a perceiver in whom a lot of other things are going on as well: he takes action, undergoes physiological responses not mediated by intentions, and interacts socially. Thus, rather than there being different kinds of schemata for the different ways in which belief is efficacious, we have schemata that we employ in all these activities and responses.

To understand how schemata would be involved in action, consider the cycle described earlier for perception, for the case of a perceiver who is also taking action. Consider the schema that would be involved if we include responses of the perceiver/actor in the cycle. In addition to directing the exploration involved in perception (e.g., shifting one's gaze, listening for the beat), a schema would direct other movements as well, such as reaching for a tool or pushing a button. Each of these involves anticipations: of what kind of detail one will see in shifting one's gaze, of there being a beat to pick up, of where the hammer or the button will be, what it will look like, and how it will feel upon being grasped or pushed.

Though the information pickup part of an action-involving cycle will often involve picking up information about objects in the perceiver's environment, such as hammers and pushbuttons, it need not. For instance, if one is singing a familiar song, anticipatory schemata will direct actions, in a continuous manner, even though there is no part of the environment we would say is being explored. For, in singing a song, it is important to sample the sounds made in one step (or at least imagine the sounds that would be made) in order to be able to make the sounds in the next step. It is in general difficult to begin a song at an arbitrary point; one normally must begin at some sort of natural pause. (I am not sure whether the pause would be properly characterized in terms of the muscular motions required or in terms of musical theme, or both. My claim that anticipatory schemata are multimodal would lead one to suspect both are involved.) Often, once one has begun, it is easier to continue than to stop. If asked to pronounce the last syllable of a word on the spot (e.g., the last syllable of "terrible") most people would need to recite the whole word (or at least imagine reciting the whole word) in order to do so. Both these phenomena are easily understood on the view that actions are directed by anticipatory schemata; one needs to get oneself into a certain anticipatory situation that is not achievable save as having already worked through an interactive process.

It might seem that one should be able to characterize getting into a certain anticipatory state as a process wherein one sequential mental state is related to a previous one by employment of a rule. Here it is important to understand the difference between how a schema is employed in singing a song, and how a collection of production rules would be used to explain the same behavior. A production rule is stated as a conditional, the antecedent of which is a certain state of the world, and the consequent of which is a certain bit of behavior the agent produces upon the fulfillment of that state, if the rule is active at the time the antecedent condition is fulfilled. Some proposals for understanding how the mind works are inspired by the complex behavior that can be generated from rather simple production rules. I grant that, if the song contained numerous recurring patterns, it might be possible to analyze the behavior in terms of a handful of production rules each of which is employed more than once. However, in cases where the song is not composed of recurring patterns, the production rules that could generate a singer's performance might have to be so unique that they would be employed only once or a few times. The production rules would state that if a certain pattern of notes already sung had obtained, then a certain pattern of notes is to be produced. Of course, one simple production rule employed only once --- a rule to produce the whole song --- would suffice. Such an account is reminiscent of the neurological explanation of skillful movements that Henry Head rejected, and instead substituted the notion of schema to explain. The anticipatory state is achieved, not by means of a stored trace of the movement required to sing the entire song, but of an interactive schema the singer is constantly employing, and which is changing every moment, as she proceeds through the song. It is entirely possible on this account that no two performances will be exactly the same, just as Bartlett pointed out that, contrary to the feeling one has while playing tennis that he is merely repeating text-book cases of movements, he is generating something anew each time he makes a tennis stroke.

To illustrate the continuous nature of cognitive processes, consider another method that is used to get one into a certain anticipatory mode: the Method of Loci, a well-known method for

remembering. Although often used merely to remember a list of items, it can be used to help one remember all the points to be mentioned in a talk. It works as follows: One associates each point in the talk with an item in a house with which one is familiar; the association of points with items in the house is made in such a way that the order of the points in the talk is the same as the order in which the items would be viewed as one walked through the rooms of the house. As one delivers the talk one then imagines touring the house and viewing the items within it. The method can be used for various kinds of performances; the use I want to consider here is that of someone using it to speak extemporaneously on a subject, to ensure that he will address all the items on a list of topics in a certain order.

Using the method of loci to guide one through a speech is a rather special case of a schema, but I want to illustrate that, even here, where the contribution from the thinker's immediate environment is minimal the process is a continuous one, and takes place in time. The speaker anticipates seeing the first item in the house. But it does not appear isolated or as an instantaneous image. Imagining the stroll through the house puts the speaker in the multimodal anticipatory state he would be in were he looking for the object. He then imagines what he would see, and, just as the anticipatory schema makes him receptive to imagining the object – say, a yellow vase – so the anticipatory schema also allows him to remember the topic he wants to address. If he were trying to remember how the vase looked, then, once he is in the anticipatory state, the schema enables him to imagine seeing (remember) the color, shape, surface texture, and so on, of the vase. When instead used to remember the points one wants to address, imagining the yellow vase at a certain location in the process of imagining himself taking the tour puts him into an anticipatory state for addressing the point he had earlier associated with it. He can now go on to consider the topic he has associated with the yellow vase in more detail: he is prepared to consider the different sides of the point he wants to make, and so in a state of readiness to form suitable statements for explaining it.

The Method of Loci requires no special talents; it will work for any normal perceiver. Even blind people can use it. The image need not be visual; on this understanding of image, the image can direct touching, listening, and tasting as well as looking. Neisser cites the success of the method in support of his view that images are anticipatory schemata like those used by everyone in perception. Images, he says, are not pictures to be examined in some internal manner, but plans for picking up information. The image of the house used in the method of loci isn't a sort of wall map with all the items in the house pictured on it; it's a plan for touring a house. Taking the tour (or even imagining taking it) takes time, and it is often the case that only some of the items would be open to one's view from any one standpoint. A mental image is "the inner aspect of a spatial anticipation. When a subject reports verbally about an image, he is really reporting quite literally what he . . . is prepared to see. The referents of language about images are possible perceivable objects in the environment, not phantasms in the head" (100). Neisser considers this point --- that images are plans for perceiving possible perceivable objects in the world, rather than internal pictures of some sort --- to be a general point about the nature of images; he cites the fact that eye motions accompany dream images as evidence that images are plans for picking up information; he regards the eye motions as attempts to carry out those plans inasmuch as is possible. On this account of perception, the phenomenon of expectancy effects measured in psychological laboratories (e.g., subjects told to anticipate a blue square actually do perceive it more readily when it actually does appear on the screen as compared with subjects who are not told what to anticipate) is no curiosity, but merely reflects the fact that perception involves the kind of receptivity on the part of the perceiver that the employment of anticipatory schemata would provide.

That expectancy effects have been observed by experimental psychologists has sometimes been cited by epistemologists and philosophers of science to make the (mostly negative) point that experience is "theory-laden". Here the kind of experiments that are stressed are those in which a subject's expectations interfere with his ability to perceive or reason about anomalous

objects (e.g., red spades on playing cards), or where a gestalt shift is required in order to perceive a certain pattern, shape, or solution to a problem. But there is a positive point: "subjects actually perceive the relevant information more quickly when they are appropriately prepared." (Neisser, "Perceiving" 103).

Slightly rewording, we might say that anticipatory schemata enable perception by directing a perceiver's information pickup. As explained, information pickup is only part of the cycle; the anticipatory schemata develop as information is picked up, and are revised as a result of the information picked up. To indicate how beliefs could then be understood in terms of anticipatory schemata, consider an example used by the ancient skeptics to illustrate the "probable impression": a coiled rope that looks like a snake, or a snake that looks like a coiled rope. A perceiver's having an "impression" that the thing is a snake can be seen as his employment of a schema for anticipating a snake. That is, the "impression" of the snake functions as Neisser says an "image" does: as a plan for picking up information about the snake. Picking up information then allows one to perceive the snake; we might even say that this is what perceiving the snake amounts to. If the thing in the corner one imagines to be a snake is actually a coiled rope, the sampling of the object directed by the anticipatory schemata leads one to revise the schema; e.g., the information picked up is that the surface is not shiny enough to be a snake, there is no head where one is expected, and so on. This example will be developed in more detail later.

The point can then be extended: people take actions as well as perceive, according as their anticipatory schemata direct their actions and facilitate perception. The anticipatory schemata develop and are revised in response to the information picked up, including the effects one observes and perceives as attributable to the action taken. The point about schemata can be further extended to include among effects other physiological responses not mediated by intention, such as blushing upon being embarrassed, becoming weak-kneed upon being frightened, or losing one's appetite upon witnessing an act of cruelty. And, finally, we can see

social interactions as directed by such anticipatory schemata as well; thus one's schemata will reflect various social stereotypes and social conventions.

It is the identification of beliefs with features of schemata that makes contact with the question of how beliefs are efficacious. On the view that the thinker's contribution not only to perception, but to physiological responses, action, and social interaction is best thought of in terms of the anticipatory schemata that direct his perceptual activity, initiate nonintentional physiological responses, mediate the actions he takes, and direct his social interactions, the appreciation that beliefs are incorporated into a thinker's anticipatory schemata provides us with an account of the efficacy of belief. This is the answer, in a nutshell, to the search for an accountof how beliefs are efficacious in the ways we know our beliefs to make a difference. The following sections develop this answer.

2. How Beliefs Make A Difference to The Believer: Belief Revision and Dynamic Schemata in Perception and Action

On the view I'm proposing, beliefs are incorporated into schemata, and schemata are dynamic. Some features of one's schemata don't vary much over the long term, i.e., those features that incorporate general knowledge about the world and give rise to anticipations such as that a ball of snow will melt at room temperature, whereas the glass bowl it is in won't. Others are transient and short-lived, such as information one picks up and incorporates into one's schema that the glass bowl one is eating out of is a safe distance from the edge of the table. Both kinds of beliefs are incorporated into one's schemata such that one has the associated anticipations (e.g., anticipating a puddle of water in the glass bowl, expecting the bowl to remain stably situated while one continues to spoon soup out of it without undue delicacy). That schemata have both long-range and transient features means regarding beliefs as incorporated into one's schemata fits well

with our everyday notion of belief. For, some beliefs are long-term, and, once acquired, change slowly if at all, whereas other beliefs are relevant to a particular situation and play a short-lived part in our lives. Whenever someone acts, thinks, or perceives, there are likely to be beliefs at each extreme, and many in-between, that make a difference to how the person acts, thinks, and perceives. Although many of the short-lived beliefs will not survive as recoverable wholes in one's schemata for very long after the moment when they were important has passed, most will have left a mark behind in the form of some variation in the perceiver's schemata , i.e., in what Bartlett called the "organized mass" of experience.

In order to act -- for example, to return a serve in a tennis game, respond to a question, or turn one way or another at an intersection -- we often have no choice but to act on less than a full determination of a situation. As we act, we pick up more information, and both our long-term and more transient beliefs may be amended and revised. If someone who believes (all) sand is a yellow-ish tan color visits an island where the sand is black, not only will the fact that the sand is black be part of his schema for viewing the landscape during his visit to the island (e.g., he will anticipate black sand wherever he expects a beach, and so employ such a belief in successfully perceiving a beach; if searching for a sand beach, he may scan for anything black to help in finding one), but he will revise his more general schema for geographical landscapes to allow for other cases in which sand can be black. On our view that beliefs are features of schemata that give rise to anticipations in directing perception and action, both these short-term and long-term revisions to schemata are cases of belief revision. Not all cases of schema revision will be occasioned by a contradiction between the information one anticipates picking up, and the information picked up. Sometimes the occasion for revising a schema is active exploring of some already familiar object, an experience with an object never before encountered, or learning to make a new kind of discrimination, and the resulting revisions are just elaborations on one's schema. For instance, someone who is studying the various styles of hand-made rugs may be incorporating information in her schema that allows her to make distinctions she was not able to

make before. For example, she may be taught about, and learn to make, subtle discriminations in colors, such as whether the color is from a vegetable dye or not; she may learn to search for features she never even noticed on a rug before, such as how yarn knots are tied; she may learn the significance of pattern asymmetries that, although she could have noticed them before training, were not til then deemed important enough to demand her attention. On our account of beliefs, she may be said to have acquired new beliefs, for the schema she employs in perceiving rugs now not only directs her information pickup in ways that enable her to perceive things she was not able to perceive before, but, further, to ancticipate things based on what she has perceived, which she was not able to anticipate before learning about such subtleties. Perception, cognition, and action are by their very natures dynamic processes, and the cognitive structures of the perceiver, thinker, or agent change in the process as well. The proposed view that beliefs are incorporated into schemata offers a very natural account of how beliefs make a difference to the believer's perceptions, actions, and physiological responses.

I referred earlier to the Ancients' example of trying to make out what a snake-like coil of rope in the corner of a dark room was to illustrate how one's anticipations direct information pickup in perception. The example was used by the ancient sceptics to illustrate the reasonableness of acting on impressions that do not meet the Stoics' criterion for certain knowledge. But their discussion of this case illustrated some other features of perception that are a matter of commonsense knowledge about belief, and our account is true to all three of them: (i) They explicitly argued for the role of anticipations in guiding actions, (ii) They argued that a person "was moved" by such anticipations, and (iii) They discussed the revision of beliefs and anticipations as a result of interaction with one's environment.

The explanation of how one could act on "the probable impression" involved an impression that was revised as one accounted for factors that affected one's impression. On the ancient account, the impression bore relations to both the perceiver and what was being perceived: the

"impression" bore a relation of truth or falsehood to the object being perceived (i.e., either it was true to the way things were or it was not), and a relation of being probable or improbable to the perceiver (the perceiver found the impression plausible and was guided by it, or he did not find it sufficiently convincing to be moved by it). On our account, the beliefs of the perceiver are part of an individual perceiver's unique dynamic, changing schema that develops and is revised as information is picked up. So, on the proposed account what is identified as being true (or not) to the way things are are these parts, or aspects, of one's schema. Since we say that the perceiver is guided by the schema, of course it can also correct to say that, since they are incorporated into one's schema, beliefs are efficacious in guiding or "moving" the perceiver. So, on our account, beliefs also have these two sides. Although the ancient account does not explicitly identify a notion corresponding to the notion of schema on the proposed account, the proposed account captures the main aspects of belief that were present in the ancient sceptics' view.

On both our account and the ancient account, the explanation of what is going on in the perceiver who is trying to make out whether the thing in the corner of the room is a coiled rope or a snake is an account of belief revision. I said before that "impressions" played the role of beliefs on the ancient account. I said this because impressions can be obtained as a result of perception, and the agent is moved by, and bases his decisions to act upon, these impressions. It is for just these reasons that I say that features of anticipatory schemata play the role of belief in the proposed account. Although the proposed account does not have anything corresponding exactly to the ancients' "impression", the ancient sceptics' account actually fits well with the proposed account in several ways. The ancient account of how an agent based his actions on the probable impression can be seen as describing how anticipations guide one's actions. When the wise man first boards his ship, "... he has not grasped with his mind or perceived that he will sail away as intended. Who could? But if he were to set out from here to Puteoli, a distance of 30 stades, with a sound ship, a good captain, waters calm as they are now, it would seem plausible to him that he would arrive safely at his destination. In this manner, he takes presentations as guides

for acting and not acting" (Cicero, Academia 2.101). Of course, if the wise man's expectations that arise from his schema based on his initial beliefs that the ship is sound, the captain is good, and the waters are calm are contradicted, his anticipation of reaching the intended destination would be revised, and replaced by other anticipations which would then guide him as to what actions to take next. These revised anticipations (in the ancients' terminology, the "tested impression") might lead a person to regard his situation as vulnerable to things that were not relevant in the situation initially expected, and thus, guide one to pay attention to things that will suddenly be relevant to sizing up the situation now anticipated, and helpful in guiding his next actions. Thus two of the most important features of the account I'm proposing --- that beliefs are expectations that guide one in action and perception, and the interactive, dynamic nature of an agent's beliefs --- can be seen in the ancient account.

The proposed account revives another part of the ancient account: the physiological responses of the perceiver that accompany various phases of the process of perception. Physiological responses include anticipatory muscular configurations, such as extending one's foot in a certain way in anticipation of going down one step, or holding out one's hand in a certain way in anticipation of receving a light object. In general, these will be more properly thought of as anticipations or expectations than as foretellings or predictions; they will have more of the flavor of one's preparedness to respond than of an analysis of an external situation. I include what we would call involuntary as well as voluntary responses. I think the ancient account includes involuntary responses as well as voluntary ones, if we use the modern meanings of those terms. For, the wise man's capacity for action in spite of the impossibility of infallible perception is explained in terms of his physiological as well as cognitive responses: the wise man "will be moved by whatever thing strikes him as a presentation that is plausible and not impeded by any other thing. For he is not carved out of stone or hewn from wood; he has a body, a soul, he is moved with his mind and senses. . . " (Cicero, Academia 2.101). On the proposed account, a person has beliefs whether or not he is currently perceiving or acting, and these are, rather than a

matter of possessing a certain mental property, of having "a particular potentiality" (Neisser, Cognition 62). One need not appeal to the part of Neisser's explanation that explains schemata in terms of the structure of one's nervous system; the main point to draw from his explanation is that the "information" preserved in one's mind is "just manifested by the specificity of his anticipation when a schema is used" (63).

Translating from the ancient account to the account I've proposed: the perceiver's having an "impression" that the thing in the corner is a snake can be seen as his employment of an anticipatory schema for perceiving a snake. What does this mean? It means one does in fact have an "image" of a snake, but here "image" is understood as a plan for picking up information about the snake. If there is a snake in the corner, the employment of the schema will enable one to make out the details: perhaps one finds oneself looking for the snake's head or eyes, looking and listening for evidence of slithering motions, trying to make out the texture of the skin, looking for reflections off the skin surface, and so on. In short, he anticipates seeing the snake. He is expecting, or "waiting on" receiving this information about the snake. William James describes the attentive process as involving two coexistent processes: "1. The accommodation or adjustment of the sensory organs; and 2. The anticipatory preparation from within of the ideational centres concerned with the object to which the attention is paid" (James, Principles 411). James refers to both of these as physiological processes which, in combination, explain what is going on in the act of attention. He writes:

> The sense-organs and the bodily muscles which favor their exercise are adjusted most energetically in sensorial attention, whether immediate and reflex, or derived. [...]
>
> That [sensorial adjustment] is present when we attend to sensible things is obvious. When we look or listen we accommodate our eyes and ears involuntarily, and we turn our head and body as well; when we taste or smell we

> adjust the tongue, lips, and respiration to the object; in feeling a surface we move the palpatory organ in a suitable way [...]
>
> But in intellectual attention, [...] similar feelings of activity occur. (James, Principles 411)

Thus, in perception, i.e., in picking up information, there are physiological aspects to one's responses (one's eye muscles focus the eye on certain things, one holds still so as to be able to hear with more precision) as well as cognitive aspects to one's responses (the response employs one's knowledge, such as how a snake's head is shaped, where the eyes would be located were the thing in the corner a snake, what sounds might be indicative of a snake's motion, which changes in light bouncing off the object are attributable to slithering motions and which to changes in the object's position relative to the perceiver as he moves closer to the object, and so on). In the anticipatory schema employed by the perceiver as he tries to make out whether the thing in the corner is a coiled rope or a snake, the physiological aspects of the responses that arise in the perceiver due to the schema are intimately connected with the cognitive ones.

What is meant by cognitive as opposed to physiological aspects of responses? My whole point here is that these aspects cannot always be separated, but some examples will indicate both what the distinction is meant to capture, and why such a distinction cannot always be maintained. Consider a person listening to a poem being read aloud. Suppose we want to categorize his responses into either cognitive or physiological ones. Probably it could be maintained that being poised to pick up a particular tone (frequency) should be classified as a physiological aspect of a person's response; it could perhaps even be maintained that being poised to pick up a particular word can be categorized as a physiological aspect, as might being poised to pick up a particular rhyme. But being poised to pick up a pun, a punchline, or an ironic tone, would be classified as cognitive aspects of the anticipation. In most cases, associating a certain current state of affairs with what is likely to happen next would be categorizable as a cognitive aspect of a response.

Emotional responses such as feeling calmed upon smelling lavender, might conceivably be categorized as a physiological aspect of a person's response, especially if the calm feeling doesn't depend upon recognizing what the odor is, and the response is fairly widespread, indicating it is independent of a person's past experience. Similarly, it might be maintained that feeling peppy upon hearing marching music should be categorized as a physiological aspect of a person's response. However, making associations dependent on one's personal experience, such as associating a certain tune one hears with an event of some emotional significance and feeling a welling up of emotion, contains both cognitive and physiological aspects: cognitive because it requires recognition of the song and remembering something connected to it, not just responding to patterns of soundwaves. Yet, the response is physiological as well; e.g., one's eyes may fill with tears. And these are not different "levels of description" of the same response. The response includes things that would be categorized as cognitive and things that would be categorized as physiological, intertwined in the person's response, but not reducible to each other, on analogy to the way in which visual and aural perceptions are intertwined in viewing a movie.

Since what is physiological and what is cognitive are sometimes inseparably intertwined, one cannot study a person's responses solely in physiological terms, anymore than one can study a person's responses solely in cognitive terms. What I am saying here is incompatible with the way many other philosophers have applied the information-processing framework imported from computer science to cognitive science. In his influential Vision, David Marr suggested that cognitive science take advantage of a methodology used in the design of large computer science projects. In the approach he describes, there are three levels of description: the abstract theoretical level (the goal of the activity), the algorithmic level (including the characterization of input and output, and the hardware, or implementation, level (25). Marr located the perceptual abilities of a person at the top, abstract level and followed a functionalist approach (although the way the tri-level framework is defined does not dictate how to apply that framework to the study of

how beliefs are efficacious). The literature related to the topic of functionalist accounts of beliefs is large, and I do not intend to address it here. In Appendix B, I treat the specific question of the kind of explanations one is able to give of the placebo effect on views that employ such a framework. The point of Appendix B is to motivate interest in an alternative view by showing that one can only explain so much on such functionalist accounts. What's relevant to the present discussion of the proposed view is that although one could go down the path of treating the cognitive aspects of attention along the lines of Marr's framework by treating the cognitive aspects of attention as functions implemented by sensorial adjustments, my neglect of a functionalist approach to beliefs (i.e., beliefs as higher level, functionally characterized entities instatiated by lower level, physically characterized entities) is deliberate and not unmotivated.

I think what James says about intellectual attention and sensorial adjustments in the excerpt from the Principles cited above just describes what we might suspect about them based on our everyday experience. What James said is that these processes -- the sensorial adjustments as well as the intellectual attunements --- are coexistent. There is no reason whatsoever to infer from this that these two processes yield alternate "parallel" accounts of the phenomena of attention. James' remarks about sensorial adjustment cited above are uncontroversial observations, and are based not onlyon appeal to everyday experience, but from researchers of different persuasions; in the discussion from which it was taken, he cites research by Mach, Helmholtz, Fechner, and Hering. He quotes from one of Mach's research papers:

> In an early writing of Professor Mach, after speaking of the way in which by attention we decompose complex musical sounds into their elements, this investigator continues: 'It is more than a figure of speech when one says that we "search" among the sounds. This hearkening search is very observably a bodily activity, just like attentive looking in the case of the eye. If, obeying the drift of physiology, we understand by attention nothing mystical, but a bodily disposition,

it is most natural to seek it in the variable tension of the muscles of the ear."
(James, Principles 413 fn.)

But, although the searching involves bodily activity, it involves cognitive effort as well: James refers to Wundt's observations that "one will always find that one tries first to recall the image in memory of the tone to be heard, and that then one hears it in the total sound. The same thing is to be noticed in weak or fugitive visual impressions" (416).

The cognitive contribution to perception is also illustrated by the phenomenon of "seeing as", the phenomenon in which an object can be perceived as either of two different, inconsistent, objects, or as representing two different, inconsistent, states of affairs. Philosophers are most familiar with examples in which a line drawing such as the "duck-rabbit" or the Necker cube can be seen as representing two different objects. The twin faces-twin vases drawing is also common. With some practice, one can at will switch between seeing the "duck-rabbit" line drawing as a sketch of a duck and as a sketch of a rabbit. Likewise, one can switch between seeing the Necker cube as representing a cube with one orientation, and a cube of another orientation; similarly, one can switch between seeing a drawing as of two facial profiles and as of outlines of two vases. The phenomenon was much investigated in nineteenth century experimental philosophy laboratories and psychophysical laboratories. But the nineteenth century researches were not restricted to perceptions of line drawings, which, I think, are often discusssed as cases of different "interpretations" of an ambiguous representation. Many, including James and Mach, were interested in the effect of anticipations on, and physiological responses to, the perception of objects, not just sketches of them. One of the most striking is that of a folded card; Mach discusses it in The Analysis of Sensations and James mentions it in The Principles of Psychology (886) The reader can easily verify the phenomenon now referred to as the "Mach Card": Bend a card (a 3 x 5 card or a business card will do) in half and place it on the table so that the halves are at about 90 degrees to each other, and the card is opened to face the viewer. View the card with

one eye only; you will find that you can at will perceive it either as opened towards you or as opened away from you. The sensation is striking, especially as the object being perceived is not an ambiguous line drawing, but one you have just held in your hand. James also discusses the phenomenon of perceiving human figures in intaglio as though they were in bas-relief, and inverted (concave) masks of human faces as though they were convex (885). Of the latter, he writes:

> Our perception seems wedded to certain total ways of seeing certain objects. The moment the object is suggested at all, it takes possession of the mind in the fulness of its stereotyped habitual form. This explains the suddenness of the transformations when the perceptions change. The object shoots back and forth completely from this to that familiar thing, and doubtful, indeterminate, and composite things are excluded, apparently because we are unused to their existence. (James, Principles 889)

This explanation, inspired by Mach's remarks, appeals to the perceiver's experience; in the case of line drawings, "The reason why one solid may seem more easily suggested than another, and why it is easier in general to perceive the diagram solid than flat, seems due to probability." Of course, in the case of line drawings, one is not really perceiving a spatial object at all, but merely imagining one prompted by the line drawing. The intaglio figures seen as in bas-relief are a different case of illusion, yet, he says: "Habit or probability seems also to govern the illusion of the intaglio profile, and of the hollow mask. We have never seen a human face except in relief --- hence the ease with whch the present sensation is overpowered" (889).

More recent experiments on cross-cultural differences in perceptions of illusion have borne out James' and Mach's intutions about this. For example, Richard Gregory cites research supporting the claim that there are cultural factors in optical illusions: "[The Zulus'] world has been described

as a 'circular culture' --- their huts are round, they do not plough their land in straight furrows but in curves, and few of their possessions have corners or straight lines. . . . It is found that they experience the Muller-Lyer arrow illusion to only a small extent, and are hardly affected at all by other distortion illusion figures" (Gregory 170). While some have questioned this explanation of the cultural difference, the existence of a cultural difference is not disputed.[6] The cultural influence is not limited to line drawings nor to the effect of man-made artefacts: People who have always lived in dense forest have different experiences than those who have not, "in that they do not experience distant objects, because they live in small clearances in the forest. When they are taken out of the forest, and shown distant objects, they see these not as distant, but as small." (ibid.) Lest we be patronizingly amused at the illusion of seeing objects as small rather than as far away, and think ourselves not susceptible to such an illusion: "People living in Western cultures experience a similar distortion when looking down from a height. From a high window objects look too small, though steeplejacks and men who work on the scaffolding and girder structure of skyscrapers are reported to see objects below them without distortion" (Gregory 170). Hence the common exclamation: "They look like toy houses and cars!", made by people who do not expect to see anything too surprising upon looking down at the ground on their first airpline flight or down onto a street from a very tall building. After all, they are familiar with the objects they are looking at.

James' explanation for the "Mach Card" illusion does not appeal to our experience, for we are just as likely to see a card opened toward us as away from us, but it is similar to Mach's explanation in that he identifies the difference between the two ways of perceiving the card as a difference in "complements from the mind." Mach describes another phenomenon that illustrates the dependence of the contribution of the perceiver on what he chooses to focus his attention upon: if one looks down at a moving stream from the vantage point of a bridge, one can suddenly have the sensation of being in motion (i.e., it feels as though the bridge one is standing on is moving.)

---

[6] See Chapter Three, "Visual Illusions" of Cross-cultural studies. Ed. D. R. Price-Williams (Baltimore, Maryland: Penguin Books Inc., 1970) for other studies and a critical survey.

Mach designed a laboratory apparatus to illustrate that one could at will switch between the two sensations by choosing which part of the apparatus to focus one's attention upon.

I have so far avoided mentioning the role James gives to "retinal images"; on our account, we follow J. J. Gibson in neglecting such notions in favor of a more active account of perception wherein the perceiver picks up information from his environment. The points James makes about attention, though, and for which he marshals lots of experimental evidence from researchers of different schools, express quite well that there is something that functions like the anticipatory schemata Neisser described as the perceiver's contribution to perception. However, James at times seems a bit too focused on the perceiver's contribution. Whereas he is right to note that "the lying in wait for impressions, and the preparation to react, consist of nothing but the anticipatory imagination of what the impressions or reactions are to be" (Principles 415), he exaggerates that contribution at times. For example, surely he is overstating things when he writes: "When watching for the distant clock to strike, our mind is so filled with its image that at every moment we think we hear the longed-for or dreaded sound. So of an awaited footstep. Every stir in the wood is for the hunter his game; for the fugitive his pursuers" (419). For, although such false starts undoubtedly occur (James actually cites laboratory research by Wundt and others that the "reaction time" for a sound one has been briefed to expect can actually become negative), in general a hunter can and does learn to distinguish between different kinds of stirs in the wood.

Likewise, James' statement that "men have no eyes but for those aspects of things which they have already been taught to discern. [...] In short, the only things which we commonly see are those which we preperceive" (420) needs to be qualified to explain that what is preperceived is dynamic, for a perceiver's anticipations develop, are elaborated upon, and are revised as he interacts with his environment and picks up more information. If the thing in the corner one imagines to be (i.e., tentatively regards as) a snake is actually a coiled rope, the sampling of the

object directed by the anticipatory schemata leads one to revise the schema: e.g., suppose the information picked up is that the surface is not shiny enough to be a snake, there is no evidence of slithering motions, there are no eyes where eyes would be expected, and so on. Then one's beliefs and anticipations change. One might then have an image of a rope in mind, and (since, on our view, an "image" of a rope is a plan for picking up information about a rope) use it to pick up information about the thing in the corner that fits the image of a coil of rope.

Similarly, although James is right to recognize the role of culture in training us in the ability to perceive ("Any one of us can notice a phenomenon after it has once been pointed out, which not one in ten thousand could ever have discovered for himself. Even in poetry and the arts, someone has to come and tell us what aspects we may single out, and what effects we may admire, before our aesthetic nature can 'dilate' to its full extent and never 'with the wrong emotion'" (420)), he again overstates the point when citing his experience that kindergarten children given a picture of a bird will not identify any parts of a bird for which they haven't already been told the name in support of his claim that "the only things which we preperceive are those which have been labelled for us, and the labels stamped into our mind" (420).

We want to grant the points James makes about the perceiver's contributions to perception, yet incorporate Gibson's insight that a perceiver picks up information from his environment that has to do with features of the structure of the environment and various means those features afford him. Consider, for example, J. J. Gibson's analysis of how people use the varieties of perceptual information available to them in driving an automobile in real-life traffic situations.[7] Gibson concluded that the driver perceives the "field of safe travel", which "consists, at any given moment, of the field of possible paths which the car may take unimpeded. Phenomenally it is a

---

7 "A Theoretical Field-Analysis of Automobile Driving" in Reasons for Realism: Selected Essays of James J. Gibson. Ed. Edward Reed and Rebecca Jones (Hillsdale, NJ: Lawrence Erlbaum Associates, 1982).

sort of tongue protuding forward along the road" (120). (More precisely, it is the "field in which the time to contact of the driver's vehicle and any object is sufficiently long to afford manuevering around the object.") On Gibson's account, automobile drivers perceive such a field. However, pace William James' claim that we need to be told the name of something in order to preperceive it, it is not true that every driver needs to be taught to perceive, or need even be able to name, the field of safe travel. Most people's driving behavior reflects that they do perceive the tongue-shaped area in front of their car delineating the space they need to keep clear in order to make a safe stop, whether or not such a field has been pointed out to them.

Neisser's statement is a qualified version of James' view here; he says: "We cannot perceive unless we anticipate, but we must not see only what we anticipate" ("Perceiving" 97). On the proposed account, we follow Neisser in this as well as in saying that perceiving amounts to picking up information as guided by one's anticipatory schema. James' physiological investigations, however, add to Neisser's work in filling in the kinds of physiological processes involved in the employment of anticipatory schemata. Making perception possible by directing information pickup is thus one way in which the belief about what the thing in the corner is is efficacious.

3. Efficacy of Belief -- Effects on the Believer

What does it mean to say that the belief is efficacious if, as is becoming clear, on the view of belief I am proposing there is no clean separation between cognitive and physiological responses? One may object that, as the cognitive and the physiological seem to be all mixed up in the schema that Bartlett aptly referred to as an "organized mass" of experience, I am masking the difference between the cognitive causing the physiological and the cognitive being comprised of the physiological. The picture many work with is one in which there are two separate causal nexuses -- a physical (or physiological) one and a mental (or psychological) one. The issue of whether or not the mental causes the physiological is, in such pictures, addressed either by identifying the

mental with some physical entities, or identifying psycho-physical laws. Whereas, the notion of beliefs as features of a schema that lives in a realm where something, e.g., anticipations, are both psychological and physical does not allow one to identify a realm of mental events and ask what they can cause. This is actually the seed of the answer; a schema lives in both the psychological and the physical worlds, as do we. Yet we can still address the question of the efficacy of belief. For, although 'cognitive' and 'physiological' are used here as inherited categories, we will try to make sense of them as certain kinds of features of schemata in order to clarify the proposed view. Otherwise, we have no commitment to them.

As with Neisser's explanation of images as plans for picking up information, so too we can describe cognitive responses in terms of certain of the perceiver's activities: for example, his deeming certain features of a situation relevant in sizing up whether it is precarious. On our view of beliefs as features of anticipatory schemata, the perceiver's beliefs will be efficacious in bringing about not only his actions (including exploratory ones) but other physiological and bodily responses. To those who might object that we cannot talk about beliefs causing physiological responses unless beliefs can be characterized independently of the physiological processes of the agent who is thinking, attending, perceiving, or imagining, I grant that the features of schemata we've demarcated as cognitive are tied up with the perceiver's bodily changes and motions, both during employment of cognitive structures, and as a result of the employment of cognitive structures. But that is no obstacle for our view that beliefs can be efficacious. We do not need to develop a lower-level (i.e., non-intentional, or sub-intentional) ontology of events and entities in order to make good our causal claims about beliefs. If it turns out that the features of schemata that we would identify as beliefs are mongrels, i.e., comprised in part of elements that physiological effects are comprised of, this does not prevent us from asking whether beliefs are causally efficacious in producing physiological effects. This is, I believe, the point of Mach's maintaining that ultimate elements, if there are any, should be regarded as <u>both</u> physical and

psychological.[8] Such a view about ultimate elements, whether they be events, atoms, or something hitherto unknown, disarms this type of objection to monistic views. (In Appendix A, I describe nineteenth century precedents of the mind-body problem, and show how the proposed view is the kind of response Mach gave to philosophical debates I find uncannily similar to twentieth-century ones.) It is people who believe things. We can establish, say, by showing the effects of persuading or convincing someone, that it is the belief that makes a difference to the cognizer's physiological changes or bodily motions. When we can effect changes or bodily motions by getting someone to have new or different beliefs, we rightly say that the belief has been efficacious in causing those changes or motions. [9]

The perceiver will undergo other physiological changes during and after perceiving, in addition to the anticipatory physiological responses necessary for perception. Suppose the perceiver has not yet determined whether the thing in the corner is a coiled rope or a snake. He knows that if he determines it is a snake, he will have to make a judgment of whether to stay and chase the snake away, or whether to leave the snake unattended and get help. He also worries that, if he chooses the latter, and it turns out that the object in the corner was not a dangerous snake, or even a snake at all, he will appear cowardly. The belief that the coiled thing in the corner is a snake carries with it this prospect as well as the prospect of making the choice to attempt to deal with the snake but failing because of insufficient skill or strength. Should these thoughts cause the person to

---

[8] Of course on Mach's view this just meant that the same entity figured in physical explanations as figured in psychological explanations.

[9] This point is made by Adolf Grunbaum, in a chapter entitled "Motives as Reasons and Causes" in The Foundations of Psychoanalysis. He refers to what he calls an "ontologically reductive physicalistic error" : overlooking the fact that "the causal relevance of an antecedent state X to an occurrence Y is not at all a matter of the physicality of X; instead, the causal relevance is a matter of whether X -- be it physical, mental, or psycho-physical -- MAKES A DIFFERENCE to the occurrence of Y, or AFFECTS THE INCIDENCE of Y. Why, one is driven to ask, is the ontological neutrality of X as between being physical or ideational (conative) not a banality among any and all students of psychoanalysis?" (72)

become weak-kneed, we rightly say that the belief that the thing in the corner could be a snake has been efficacious in making him weak-kneed. The snake itself is not the cause of the physiological response, for one can become weak-kneed whether or not a snake is present; it is the belief (which may or may not owe its existence to the presence of the snake) that a dangerous snake is, or might be, present that causes the weakness. The phenomenon of physiological responses such as tensed muscles or the sensation of losing one's control or strength upon perceiving the height one is at is known to those who work in elevated open structures, e.g., bridge inspectors who must walk on bridge surfaces located far off the ground. Here again, being at that height may have contributed to the person having the belief, for his belief may have come about because he actually perceived the height he was at. That we rightly consider the belief to be efficacious in such cases, however, is illustrated by two facts: (i) bridge workers who train themselves not to look down and perceive the height, or to think about it, can avoid the physiological responses in question, and (ii) if a person is sent to walk across a beam in the dark and he is made to believe that the beam is forty feet off the ground, his physiological responses are very different than if he is made to believe that the beam is only four inches off the ground.

The fact that such physiological responses follow from belief is really no more mysterious than the fact that physiological responses follow upon one's attempts to act. It is hard to imagine anyone arguing with the latter point. Yet, in most philosophical discussions, the efficacy of belief has been pretty much equated with the efficacy of belief in reaching a decision about what to do. On the account I'm proposing, beliefs become much more recognizable as the beliefs we know from our daily mental life; we know beliefs can affect not only our decisions and our verbal behavior, but that beliefs also affect our perception (at times even literally "blind" us to some fact or other), influence where our attention becomes focused, make us weak-kneed, make us blush, or make our hearts race.

CHAPTER IV

BELIEFS IN SOCIAL INTERACTION

A. How Beliefs Make A Difference To A Believer's Social Environment

How does the view of how beliefs are causally efficacious proposed in the previous chapter address the question of how beliefs are causal in the existence and maintenance of social institutions in a believer's environment? Our starting point for answering this question will be a commonsense view about social institutions and conventions: social institutions and conventions are brought about and maintained by a certain pattern of beliefs and expectations being developed and sustained among individuals who are in some sort of social arrangement. The general idea is that beliefs are causal here because we can change whether the social institution or convention exists by changing people's beliefs. On the proposed account, beliefs are incorporated into anticipatory schemata, which give rise to expectations. As the terms are used here, there is often no clear delineation between beliefs and expectations in the cyclic process during which we interact with others.

A much-examined example of a social institution is the practice of contract-making. Whether the practice exists or not is a matter of whether or not one can conduct his business by entering into contracts, or whether things have deteriorated to the point where he can't, because of not being able to count on people honoring their contracts. The intuitive idea is that there is a clear difference between the state of affairs indicated by: "Of course you can buy on time around here," and the one indicated by: "Nobody around here is going to take another person's word for it; you'll have to have the cash up front." So, we can imagine a situation changing from one where you can make deals, and count on people to honor them, to one where you can't. And we can

imagine the change being due to the intervention of an Iago-like character who plants seeds of suspicion throughout the social group, with the result that people no longer trust certain of the others enough to want to enter into contracts, and feel sufficiently betrayed so that they no longer hold up their end of the contracts they've made. Then, the state of affairs---business in that culture or group can be conducted by striking contracts---has been changed, and the change has been effected by influencing beliefs (getting people to be suspicious of others, by innuendo). [10]

So, beliefs have played a causal role in effecting the change: by influencing beliefs in some manner, a change in a state of affairs has been effected. There are other ways than by influencing someone's estimation of others' characters for this change to have come about---for instance, by the practice of bribery becoming so common that you could never count on someone to carry out a contract, as they might be offered an irresistible bribe to renege. Then beliefs would be involved in causing the change, too, though in the latter case the belief would involve a change in the estimation, not of people's characters, but rather of the existence of the temptations they are likely to face. The description of the practice of contract-making is sometimes modified so as to include various refinements: it might matter who you are, and it might be that it is only with some people, and not with others, that you can count on the other person to make and honor deals with you. Then, a means for these kinds of changes to come about would be by the establishment of emerging or evolving social stereotypes, or the disintegration of existing social stereotypes. "Stereotype" is used here in the broadest sense; stereotypes are not necessarily negative and their use is not necessarily unwarranted. We have stereotypes for cups and trees, and these aid in perceiving cups as cups and trees as trees.

---

[10] Richard Gale has rightly pointed out to me that, alternatively, the result of Iago-like interference might be that people arrange for some means of enforcement of contracts. That is, although people will enter into into contracts only if they have confidence the other party is going to fulfill his or her obligation, this confidence can arise from trust in an enforcer even in the absence of confidence in the trustworthiness of the other party to the contract.

One feature common to philosophical treatments of examples of social institutions such as conventions and the practice of contract-making is that the establishment of the practice is said to involve individuals having what are often called higher-order beliefs or expectations: beliefs or expectations about other individuals' beliefs or expectations. The essential involvement of higher-order beliefs might thus be thought to provide a clue to the character of the causality involved when beliefs are causal in establishing, maintaining, or changing social institutions.

In this chapter, we extend the account of belief developed so far to include social perception and interaction. In the example above of a person's social environment changing from one in which he can count on others to enter into agreements with him and to carry them out, to an environment in which he can't, the person whose social environment has changed might describe the change by saying that his opportunities for certain kinds of interactions have disappeared. Alternatively, if the environment were to change from one in which the institution of contract-making does not exist to one in which it does, he might describe it as one in which opportunities have come into existence.

On our account of how beliefs are involved in perception and action, a person's anticipatory schemata direct his exploration, information pickup, and actions. So the person's opportunities for interactions depend in part on his anticipatory schemata, which incorporate his knowledge and beliefs as well as his skills. Since the social environment includes other people, a person's opportunities for interaction with others will also depend upon the ability of the other people in his social environment to perceive him or her as having certain capabilities and characteristics. The anticipatory schemata employed by various other people in a person's social environment will not only enable perception of that person as a person in that social arrangement, but also determine other information that these other people pick up about him or her. Just as a perceiver employs (cognitive) stereotypes in perceiving a tree as a tree, and may even employ them in determining

what type of tree it is, and whether it might be capable of supporting a hammock, so the perceivers in one's environment employ social stereotypes in perceiving him or her; social stereotypes can be used in perceiving others as trustworthy or untrustworthy. Further, the social institutions in one's culture, such as the practice of contract-making, have their roots in other people's schemata for interaction.

To take a very practical example, say you want to purchase a time-share property. To do so requires that there be other people in your social environment who have the notion of such a business transaction and know what to expect of, and how to interact with, banks and legal authorities to effect such a transaction. If the institution of buying, selling, and owning time-share properties exists among enough people in the economic-social environment you are in, and you have the sort of characteristics such that you will be perceived as a potential party to such a transaction, you have opportunities open to you that do not exist if no others in your social environment have such a social institution incorporated into their schemata. Similarly, the possibility of participating in a social convention is open to you only if others in your social environment have the right kind of expectations of your behavior, and of your expectations of their expectations of your behavior. We will be able to be more precise about how anticipatory schemata are involved in social interaction later in this chapter, after suitably extending the account of belief developed in the previous chapter.

However, questions such as the origins of justice or the morality of certain strategies for interacting with others are outside the scope of this inquiry. The proposed account of belief does not address what counts as a fair institution. In spite of the fact that institutions create beneficial opportunities not available otherwise, some institutions may exclude individuals for no discernible reason, or exact a disproportionate amount of sacrifice from some. There is nothing about the notion of belief or the ways in which it is efficacious to prevent such consequences. This is analogous to the point that, although social stereotypes incorporated in anticipatory schemata

may enable perception of people's personalities, talents, and characteristics, they may also cause a perceiver to misperceive a person's characteristics, or "blind" a perceiver to some characteristics of a person. Our interest here is in how beliefs are causal, and our present aim is to give an account of how beliefs are causally efficacious in the existence and maintenance of social institutions. We begin by extending the account of anticipatory schemata from perception of one's immediate physical environment to social perception.

1. Social and Non-Social Aspects of Perception

Neisser remarks that perception differs from other skilled activities in what it affects. Although perception does change the perceiver, he says, "[Perception] differs from performances like sculpting and tennis playing in that the perceiver's effects on the world around him are negligible; he does not change objects by looking at them or events by listening to them" (Cognition 52).

No one would disagree that we sometimes affect the environment when we are perceiving it; we may turn the pages of a book we are reading, or step on a leaf as we perceive a sunset. Most would also agree with Neisser's point in the passage quoted above: i.e., of activities that involve an individual interacting with his environment, the effects on the thing perceived that might arise during the process in which it is perceived are not ineliminable results of being perceived. However, we should clarify the claim here: although it is easy to imagine a process of perceiving a stone in which the stone is not affected, we cannot always arrange a process of perceiving a person such that the person perceived is not affected during the process of being perceived by someone else. There are sometimes ineliminable, nonnegligible effects on the one perceived due to the processes employed in social perception; that what is perceived is affected during the process of being perceived is a symptom of the interactiveness of the perceptual cycle in a special context. The cases in which it arises involve other perceivers. Neisser does not offer this kind of

clarification of the passage just quoted, but I think it is at least consistent with his view; the clarification was not important in the situations in which he was discussing perception.

The clarification is important for the cases I want to consider here, however. This is because I want to extend Neisser's view of anticipatory schemata to their employment in social perception, whereas he did not develop how his view of perception would work in contexts involving social interaction very far. In fact, comments he made at a symposium dedicated to "social knowing" many years after Cognition and Reality appeared hint that he would agree that his remark does not apply to cases of perception involving social interaction. The symposium was organized around the idea that social knowing occurs along a continuum between direct, "perception-based knowing" and indirect, "cognition-based knowing" of persons. In explaining why, despite sympathizing with that idea, he was dissatisfied with the work presented there, Neisser commented: "The theories and experiments described here all refer to an essentially passive onlooker, who sees someone do something (or sees two people do something) and then makes a judgment about it. . . . He watches and makes judgments." Whereas, he pointed out:

> Realistically, we should face the fact that social judgments are not universal, but occur in particular contexts. Watching television and other performances is one such context [...] being in a room full of strangers may be another. [...] The people who participated in the studies cited in the symposium were allowed to watch slides, see films, and read descriptions, but only rarely to do anything besides preferring and judging. My guess would be that 'social knowing' when it does occur, takes different forms and depends on different variables in various situations. ("On Social Knowing" 604)

However, he implied his own theory of perception was of limited usefulness in such an investigation. He said then that although the perceivers of Gibson's theory and of his own theory

"do more than those of other cognitive theorists . . . still, they only perceive" ("On Social Knowing" 604), and he made other comments lumping the perceivers of his own theory of perception in with those of other theories in which the perceivers and knowers are "detached". The "detached perceivers and knowers", he seemed to think, were inhabitants of a theory of perception, but not suitable for use by social psychology as models of human nature (605). I see more usefulness for employing Neisser's account of perception in social psychology, when appropriately developed, than he seems to have seen.

I think the way to explore the usefulness of Neisser's account of perception in social psychology is to consider cases in which perception does not leave the thing perceived unchanged. That is, we can consider cases in which perception occurs as part of an interactive process that does not leave the one perceived unchanged. The point that perception is guided by anticipations that are revised as perception proceeds can be illustrated, as it was in the previous chapter (by the example of how perception is involved in making out whether the object in the corner is a snake or a coil of rope), without thinking about the difference between cases where the object perceived was affected by a perceiver and cases where it was not. We will be missing a qualitatively different kind of case of a perceiver's interactions with his environment, though, if we consider only the kinds of cases in which the perceiver does not affect what is perceived. Thus, we will proceed as follows: we begin with cases in which other perceivers may be present, and regard the simplest cases, in which only the perceiver is affected, as limiting cases of the more complex one. Whether or not the thing perceived is affected depends on the kind of interaction, if any, between the perceiver and the perceived.

I am not making any presumption here as to whether or not all contexts in which other perceivers are involved should be called social contexts. There is some disagreement about what constitutes a social interaction or a social context, how the social is related to the ability of the perceiver, what perceptual abilities are requisite for social interaction, and to what extent perceptual abilities

depend upon socialization. But these disagreements need not be sorted out here; I take it as uncontroversial that applying the terms "social interaction" and "social context" to interactions and contexts presupposes that those involved in the interactions and contexts have the ability to perceive. Thus, I am not assuming that all interaction involving perceivers should be regarded as social interaction, but I am assuming that all cases of social interaction involve perceivers.

There is an objection one might plausibly raise to my statement that one perceiver can be affected by another's perception of him or her which is, I think, due to dwelling on the kind of cases in which there is no social interaction between the perceiver and the perceived, and not appreciating that qualitatively different cases exist. The objection is that the perceiver who becomes self-conscious or inhibited while being perceived is affected by his belief that he is being perceived. We are not affected when being monitored by hidden cameras or tape recorders of which we have no suspicion; thus, the objection goes, the fact that a perceiver who is unaware of being perceived will not be affected during the process of perception proves that there is nothing about the process that is efficacious in producing whatever effects ensue. Rather, all the ensuing effects can be ascribed instead to the awareness that one is being perceived. For cases in which there is no interaction between the perceiver and perceived, these points can be granted. But cases in which there is social interaction are different.

Compare the case of someone perceiving a stone doorstop with the case of someone perceiving a person leaning against a door. On our account of perception, the perceiver employs anticipatory schemata, whether the object being perceived is a doorstop, an animal, or a person. Employing these schemata involves physiological anticipations on the part of the one employing them. We have said that beliefs are features of anticipatory schemata. These beliefs will show in the physiological anticipations and responses in a perceiver's face and posture, even if he does not speak. I have made a point of saying that physiological configurations and bodily motions need not be the result of intentional action; employing anticipatory schemata involves the person

employing them physiologically. Anyone who is not too far removed from the culture of the perceiver will be able to perceive something of the perceiver's beliefs, thoughts, and anticipations. The point is not that the perceiver is not hidden any more; a perceiver in full view differs not only from hidden cameras and tape recorders, but from cameras and tape recorders in full view.

Although Neisser does not discuss cases in which what is perceived is affected by the perceiver, he does discuss physiognomic perception. He takes the phenomenon of being able to quite literally see how another person feels as something to be explained: "physiognomic traits are perceived more easily than the physical movements that give evidence of them" (Cognition 189). He dismisses several attempts at explaining physiognomic perception, including inference: "Do we first see the movements, gestures, and expressions themselves, and then infer their meaning on the basis of past occasions when we saw similar ones? Surely not. The process is too quick and automatic..." (189). He then offers his own view of normal perception in explanation:

> The simplest approach to this problem is to suppose that physiognomic perception is not different from any other kind. It requires a preparatory schema, ready to pick up information and to direct explorations that will pick up still more. As in ordinary perceiving, this information specifies something that really exists. In the physiognomic case, that "something" is another person's emotion or feeling.
>
> I am assuming, then, that we all have schemata for physiognomic perception. [...] Schemata for action and for physiognomic perception undergo as much development as any other cognitive structures; indeed, they are particularly dependent on social experience. (Cognition 191)

If we think about the phenomenon of physiognomic perception, we see that it actually illustrates that there is a mode other than by the intentional actions he takes that the perceiver affects his environment. That is, we will see that, in cases of social interaction involving two perceivers, the processes involved in perceiving include effects due to involuntary physiological responses of the perceiver as he employs dynamic anticipatory schemata in the process of perceiving. Neisser's explanation has a different point: his explanation is meant to show that physiognomic perception can be subsumed under normal perception, rather than to illustrate how perception of other perceivers with whom one is interacting differs from perception of inanimate objects. But we can develop his account to display the difference between perceiving other perceivers and perceiving non-perceivers.

What sort of schemata are schemata for perceiving other people? Neisser's account is not developed beyond treating them on a par with schemata for perceiving inanimate objects. So the most one could draw out of Neisser's own (undeveloped) account of perception here is that a perceiver adds something to the environment because of the change he undergoes in perceiving. That is, Neisser says that a perceiver is changed as a result of perceiving. If the perceiver's emotions or feelings are changed in perceiving, then (based on what Neisser says about physiognomic perception) these changes to his emotions or feelings can be picked up by other perceivers, if there are any around. We want to develop the account of schemata for cases of social interaction further than Neisser does, however. For, in cases where the perception involves interaction with another perceiver, schemata for perceiving others do differ in some ways from schemata for perceiving inanimate objects. This is because, on our account, perceiving involves interactive dynamic anticipatory schemata that direct information pickup and are revised as new information is picked up, and so, in the process of perceiving another perceiver, the anticipations and expectancies that the one doing the perceiving has of the one perceived, may in turn become evident to the one perceived during the process.

It hardly needs to be argued for that there are social situations in which staring at someone will affect that person. Here the feeling one has upon being stared at is instructive: if you are stared at, you may feel as though you want to look back aggressively and say "What is it? What do you want?" Or, you may want to say: "Stop it!" as though you were being assaulted, or make a point of ignoring the one staring, as though you were being panhandled. It's true that being stared at is not a normal social interaction; staring is often taken as a sign that someone is emotionally disturbed or mentally deficient in some way. (One might also feel like saying: "What's the matter with you?") However, although not all cases in which one is perceived are as irritating or feel so demanding, the point is that interacting in a social situation is, figuratively speaking, somewhat like a (not unnecessarily unpleasant) contact sport. Being aware that someone has overheard you may be disconcerting; the phenomenon of someone's look piercing you like a dagger, or, alternatively, melting away your anger, is a different phenomenon. This is not to deny that there may be situations in which staring will not affect the person stared at; for example, the situation of an audience member staring at a lecturer in a large lecture hall. Probably any perceiver, though, including most animals, will be affected by being looked at at close range.

But the point goes deeper than simply respecting social and cultural norms or species-specific innate dispositions about personal space. That one is affected by the presence of perceivers is part of our everyday knowledge. For example, many people have strong preferences about having intimate friends or family in the audience when they are to perform difficult tasks, such as public speaking. Some strongly prefer the presence of supportive friends in the audience, others strongly prefer that even their most positive supporters not be in the audience. But, while studies consistently find that the presence of an observer does in fact affect a person's skilled performance, the effect can vary depending upon whether the task is easy or difficult, and on whether the perceiver is neutral, hostile, or supportive. A recent laboratory experiment on the effect of the presence of "supportive" observers on skilled performances (Butler & Baumeister) concluded that the presence of "supportive" observers impaired performance on a difficult task

of skill, while the presence of "hostile" observers actually improved it (as compared to being left alone or being watched by a "neutral" observer). "Supportive" observers were either friends brought along by the performer, or observers with whom the performer was not previously acquainted, but who stood to gain cash rewards for the performer's successful performances. "Hostile" observers were observers with whom the performer was not previously acquainted but who stood to gain or lose cash rewards according as the performer failed or succeeded at the task. "Neutral" observers were observers who were neither "supportive" nor "hostile". The effect did not appear to be due to increased self-awareness. The authors of the study suggest that "people actually face a difficult trade-off when preparing for an upcoming performance under demanding conditions. To choose between a supportive and a neutral or adversarial audience is to choose between feeling better versus doing better." The details of this trade-off are not important here; I am appealing to the fact that the phenomenon is part of common sense psychology and is confirmed by experimental studies in support of the more general point I wish to appeal to: that the social or personal interactions in which perceiving others takes place can give rise to effects on the one perceived.

Thus, we have established that the one being perceived may sometimes be able to sense attitudes towards him, and expectancies of his behavior, particularly of his immediately impending moves. But there is more than this involved when we perceive others; we have schemata for picking up information we take to be indicative of gender, age, friendliness, intelligence, trustworthiness, and so on. As we employ them, these schemata, which may be thought of as incorporating the stereotypes, or "images" we hold of others, are exhibited. Here I am making use of the point made earlier that, on Neisser's account, an image is a plan for picking up information.

Turning now to our example of perceiving a person leaning against a door, the person leaning against the door will certainly be able to pick up on the differences between perceivers who

expect to see the owner of the establishment, those who expect a valet, and those who expect a loiterer. The person standing in the doorway will often be able to sense the difference before the one perceiving him says a word to him or makes a move. The perceiver cannot perceive without anticipating something, and those anticipations, however general or specific, will show in his demeanor. So, for instance, if the perceiver expects the figure leaning against the door to be the owner of the establishment, the anticipations he develops as he employs his schemata for perceiving the supposed owner will affect his expression as he goes about picking up information about the person's face. The sorts of anticipations for this case might be that he expects the supposed owner to be glad to see him, he expects the supposed owner to greet him, and perhaps expects the supposed owner to open the door and take over the conversation, sort of acting as host during the interaction. In contrast, if the perceiver is expecting a valet, he may look at the person in the doorway in a more formal or businesslike manner, expecting the supposed valet to be ready to accept his request graciously, or, if there is no request, to make himself inconspicuous. The anticipations of the perceiver will be different again for the case in which he expects the person in the doorway to be a loiterer. Since the perceived person in the doorway is a perceiver as well, he will have his own expectations of the kind of person the approaching stranger is, or, at least, of the kind of interaction he expects to have with him. Further, and more importantly, he will respond (with both involuntary and voluntary responses) to the expectations of the one perceiving him. If we were not socially trained, or if we could override our training, we might be able not to react; to train oneself not to react and respond to others' expectations on a regular basis, however, would mean rendering oneself sociopathic. Social interaction requires awareness of, and responding to, other's expectations. So responses to the approaching stranger's expectations will arise in the one perceived.

The significance of this kind of case of perceiving a perceiver is this: <u>It will make a difference to the one being perceived whether the approaching stranger looks at him as though he were the owner of the place or whether he looks at him as though he were a valet.</u> Since, on our view,

one must anticipate in order to perceive, if the approaching person is to perceive the person leaning against the wall, he must have some anticipation of the one being perceived. The one being perceived will be affected by the anticipations of the perceiver as the perceiver picks up information about him.

Neisser does not say this[11]; my claim here is based on Neisser's view as supplemented (in Chapter III.B.2 above) by the points made about physiological anticipations cited by James as common knowledge in psychology in his discussion on attention.

2. Beliefs in Social Interaction -- Effects on Those Interacting

It is no coincidence that the examples just offered to illustrate that the one perceived can be affected by the processes employed in perception were cases of social interaction. For, social stereotypes employed in social interaction are efficacious in ways that schemata for balls or trees generally are not efficacious when employed in situations such as bouncing a ball or perceiving a tree. As mentioned earlier, by "social stereotype" here I mean something like the concept of stereotype in cognitive science; for example, there are stereotypes for balls, cups and trees (a ball is round and will hold its shape when thrown or caught, a cup has handles and will hold fluid; a tree has green leaves and bark); when woven into one's schemata, these cognitive stereotypes aid in perception by directing information pickup, cause anticipatory physiological responses, and enable us to draw conclusions and form intentions. Social stereotypes are woven into an

---

11 Neisser does not discuss exceptions to his point that, in general, perception does not change what is perceived. I am at a loss to explain why, given his statements about physiognomic perception, Neisser says at other places that another person's schemata "are locked inscrutably within his skull where we cannot see them" (Cognition and Reality 186). He makes the latter point in the context of arguing that it is the things everyone's schemata must have in common that enable social prediction. The statement makes more sense if we think of cases where the person whose schemata we "cannot see" is not actually employing anticipatory schemata.

individual's schemata just as schemata for trees or balls are. Just as anticipatory schemata enable perception and interaction with inanimate objects, so social stereotypes (broadly construed) enable social interaction. And, just as there are expectancy effects associated with other anticipatory schemata, so there are expectancy effects associated with the employment of social stereotypes.

An example of the kind of expectancy effect I am thinking of here is the phenomenon that some teachers unwittingly perceive overweight children as less intelligent than other children to whom they are similar in other respects. The social stereotype of the dull overweight child is efficacious in the ways we've already said anticipatory schemata are: i.e., the person holding the stereotype is less likely to recognize signs of intelligence in such children, as compared with the children of whom she expects intelligent behavior. As many have already remarked, one way in which the social stereotype the teacher holds is efficacious is via intentional action: acting on her estimation of the child, which is in part a product of the stereotype she holds, she may decide not to give the child the opportunities reserved for children she deems more intellectually advanced. The child may actually learn less about a subject as a result of that deprivation.

A well-known example of the way in which a social stereotype incorporated in someone's anticipatory schemata may enable social interaction was given by William James. Although he was concerned to explain the rationality of religious belief, he gave an example of social interaction to make his point:

> [Turn now to] a certain class of questions of fact, questions concerning personal relations, states of mind between one man and another. *Do you like me or not?* -- for example. Whether you do or not depends, in countless instances, on whether I meet you half-way, am willing to assume that you must like me, and show you trust and expectation. The previous faith on my part in your liking's

> existence is in such cases what makes your liking come. But if I stand aloof, and refuse to budge an inch until I have objective evidence, until you shall have done something apt, . . . ten to one your liking never comes. ("Will to Believe" 730)

The social stereotype here might be described as that of a congenial acquaintance, and James is pointing out that employing it in social interaction is not a matter of seeking sufficient evidence before determining whether someone is or is not a congenial acquaintance. For, whether or not someone is to become a congenial acquaintance of yours depends on the nature of the interactions you have in the process of becoming acquainted. The nature of the interactions you have, in turn, may depend on the expectations you have of how congenial your interactions are going to be. Thus, to rephrase James, James is arguing that we can be (pragmatically) justified in expecting things we don't know will hold, for the expectancy itself may be causal in bringing about the expected state of affairs. The points that there are expectancy effects associated with the employment of social stereotypes, and that the employment of stereotypes enables something to occur are just the analogues in social psychology of the points made earlier about schemata in cognitive psychology. But, as mentioned earlier, social stereotypes are also efficacious in ways not reflected by either of the modes of causation by which schemata of inanimate objects are efficacious (i.e., the mode of enabling the perceiver to perceive, or the mode of being part of a reason for intentional action).

The people who are involved in a social interaction are affected by the interaction: if two people are to interact, each one's schema for that social interaction must somehow accommodate the other's; if the two people's schemata are not complementary, there is a sense in which one interaction has been aborted, and an interaction of another kind occurs. Consider the contrast between interacting socially and interacting with an object, say, for example, bouncing a ball. A person's schema for interacting with a ball accommodates the ball's features; if the ball is a bit more deflated than expected, or is a superball, the person's anticipatory schema for the ball's

response to his actions is soon revised so as to match the ball's behavior. But, the ball is not required to make any accommodation: it neither resists nor cooperates in the process. Thus it is a joke to complain that the basketball is not cooperating with your attempts to throw it through the hoop, but it is not a joke to complain that your teammate is not cooperating with your attempts to effect a joint action to score points in a basketball game.

In contrast to interacting with a physical object such as a ball, those involved in social interaction may resist or cooperate with each other. Thus, in the example given above of one person taking someone standing in a doorway to be a valet, the approaching person cannot generally unilaterally treat the one standing in the doorway as a valet: As the approaching stranger glances at him in a way that shows he expects to be served, the one standing in the doorway can make a point of looking back at the approaching stranger defiantly, or use body language (such as crossed arms) that shows he does not regard the approaching stranger as someone he expects to wait upon. Similarly, in the case of the child whose intelligence has been underestimated, there are various ways in which the child may be affected. He may be complacent about the underestimation of his intelligence, acting contented when treated condescendingly, doing merely what is expected of him and never challenging those expectations. In so doing, his schemata for interacting with others, or at least with some teachers, may change, while the teacher's schema for interacting with overweight children is unaffected (or perhaps even strengthened). The child's responses will not all be categorizable as intentional actions; they may be subtle responses, such as bodily movements or postures that reflect his compliant attitude, or calm, vacant, facial expressions.

Alternatively, the child might not comply with the teacher's expectations of his performances in the classroom; if the teacher's schema remains unaffected, the child could then be displaying intelligent behaviors that the teacher does not recognize as such. The child might then learn how to progress in his learning sans teacher reinforcement or, alternatively, his progress might suffer

due to the lack of encouragement his own responses receive. In either case, his schema for social interaction with a teacher, at least in a classroom context, will be affected by the teacher's failure to recognize his intelligent responses: it will not incorporate an expectancy of the teacher's recognition of his intelligent responses. A happier scenario might be that the child persists in expecting the teacher to recognize his responses as intelligent, and succeeds in getting the teacher to revise her schema for interacting with him. Revisions to the teacher's schemata may or may not include revisions to her schemata for overweight children in general[12], but the point is that the teacher can be affected differently by the child's refusal to complement her social stereotype for interacting with him than she is by his compliance with her expectations.

I have so far been appealing to common experience, rather than research in social psychology. This is in part because there is so little research involving the efficacy of stereotypes on the person being stereotyped in the context of personal interactions. This is so despite the fact that the literature on stereotypes has burgeoned since the 1950's (Fiske 357, Oakes, Haslam & Turner , Bodenhausen & Wyer 6). For, in social psychology as well as in philosophy, investigations of the role of social stereotypes in social interaction tend to focus on the perceiver rather than the one perceived. Some of the research on social stigma did attend to the experience of the person being stereotyped but, even then, the research was disproportionately focused on the stigmatizer rather than the stigmatized (Crocker, et al. 504). As to the more general topic of stereotypes, one early, dominant view about stereotypes among social psychologists was that stereotypes serve a cognitive function, e.g., are actually useful in making inferences, make efficient use of one's cognitive resources in helping one organize one's environment, and so on (Oakes et al.). In the voluminous haystack of research papers in social

---

[12] Here I am thinking of a high school calculus teacher who persisted in declaring as a universal exceptionless rule that girls can not understand calculus, in spite of explicitly recognizing the individual students in his class who could solve difficult calculus examination questions and happened to be girls as students who "really understand" calculus.

psychology, however, there are a few recently-arrived wisps in which the subject of the experiment is the one being judged.

One such piece of work investigates a phenomenon identified as "stereotype threat" by the social psychologists Steven Spencer, Claude Steele, and Diane Quinn. Perhaps it will be helpful in clarifying my own view if I identify what I think are the strengths of their work, as well as explaining how it still falls short of investigating the expectancy effects of social interaction I have been discussing. They describe the phenomenon as follows: "Being the potential target of a negative group stereotype ... creates a specific predicament: in any situation where the stereotype applies, behaviors and features of the individual that fit the stereotype make it plausible as an explanation of one's performance" (Spencer, et al. 21). What I find valuable about this work is that it breaks away from the "in-group"/"out-group" characterizations of participants in social interaction, and instead recognizes that specific social interactions will involve specific stereotypes.

> Consider the aging grandfather who has misplaced his keys. Prevailing stereotypes about the elderly . . . establish a context where his actions that fit the group stereotype, such as losing keys, make it a plausible explanation of his action. Stereotype threat. . . is conceptualized as a situational predicament --- felt in situations where one can be judged by, treated in terms of, or self-fulfill negative stereotypes....It can be experienced by the members of any group about whom negative stereotypes exist --- generation "X", the elderly, white males, etc. ... it is situationally specific --- experienced in situations where the critical negative stereotype applies, but not necessarily in others. (6)

Spencer, Steele and Quinn investigated the hypothesis that, when women are in situations in which their mathematics skills are exposed to judgment, stereotypes that women have less ability

in higher mathematics may make them feel extra pressure to perform, and that the extra pressure does in fact interfere with their performance on difficult mathematical tasks. In keeping with their view that stereotype threat is situationally specific, Spencer et al. conducted a controlled study in which they manipulated the relevance of the stereotype. That is, they tried to create both situations in which women did not feel that the stereotype was being applied, and situations in which they did feel that the stereotype was being applied. (Many others have attributed the gender disparity observed in statistical studies of standardized mathematical tests to gender socialization. Stereotype threat is a different mechanism from gender-role socialization. Gender socialization is the process whereby women are given less encouragement to participate in mathematical endeavors than men in comparable situations are.) The situation in which people were tested in Spencer et al.'s experimental set-up was as follows: a standardized mathematics test was administered to a mixed-gender group of individuals, all of whom had done well in mathematics standardized tests and college courses, and all of whom had responded in screenings both that they were good at math and that it was important to them that they were good at math. Several different experimental set-ups were used in the study.

On the set-up in which only the difficulty of the test was varied (i.e., the experimenter made no comments as to whether the test had shown gender differences in the past), women (as a group) did as well as men (as a group) on the easy test, but less well than men on the difficult test. In the set-up in which the stereotype relevance was manipulated, the experimenter either told the (mixed-gender) group of participants that the test they were to take had yielded gender differences in the past, or told them that it had not. They remark: "We assumed that telling participants that there were gender differences would lead them to believe that men did better than women. Of course, this conclusion is not inevitable, but all participants in this condition when asked informally reported this to be their interpretation" (11). When stereotype relevance was manipulated in this manner, women in the situation in which everyone was told that the test had yielded gender differences did far less well than comparably prepared men; women in the

situation in which everyone was told that the test had not yielded gender differences in the past did as well as comparably prepared men. Thus, the relative underperformance of women as compared to men of similar education and training that has been widely reported did not show up at all when the experimenter administering the test described it to the participants as not being a gender-sensitive test.

Thus, it was something about the test-taking situation in the latter case that was responsible for the difference. Further tests were made in which the participants were asked questions meant to measure each participant's evaluation apprehension, his or her sense of self-efficacy (mathematical ability) , and his or her anxiety (i.e., feeling nervous, jittery, worried, or indecisive); these questions were asked after the test had been described as either gender-sensitive or gender-insensitive, but prior to taking the test. The results suggested that, of these three, only anxiety was plausibly a mediator of the effect of stereotype threat on women's test performance. So, the negative effect on women's performance that was observed when the test was described as being gender-sensitive was not attributed to the mediating effects of apprehension (of being evaluated by an imagined audience), or a lowered evaluation of their own mathematical ability. Thus, the authors concluded that "stereotype threat has effects on performance that go beyond any effect it has on [lowering participants'] expectations and that it was these extra expectation effects that mediated the present results" (23).

I think these researchers are on to something important. It is not only refreshing that they have managed to avoid considering only the kind of condescending explanations of women's performance on difficult mathematical tasks that are usually given as the only alternatives to biologically-based explanations, e.g., that women have been "socialized to fail", or that they have "internalized" something that keeps them from learning mathematics or succeeding on mathematical tasks. These latter types of explanations imply that women have been somehow intellectually damaged or deformed by society in some way and that they really are less

intellectually able than men. The possibility Spencer et al. have presented is that it is a specific kind of predicament women are often put in when performing mathematical tasks that degrades their performance, a predicament that men in the same room are not put in. The predicament is what they call stereotype threat. Putting a group of people in the same room and administering the same test to them by the same person in the same way with the same instructions is not necessarily gender-neutral. If the stereotype that difficult mathematical or scientific skills will separate the men from the women is felt to be in play, the situation each is in really is different depending upon one's gender. We could say that in some sense, they are in different social situations; this is a respectable application of the concept of "social reality." The significance to social psychologists of this stunning finding is that "Predicaments are circumstantial and should be easier to change than internalized characteristics", which is of great practical import. The gender-sensitivity of demanding mathematical and scientific tests is practically taken for granted by many in educational research institutions; the fact that, without changing the education or any previous experiences of participants, Spencer et al. were able to obtain results with standardized test questions that did not exhibit gender differences is a convincing demonstration of the practical importance of their hypothesis.

As for philosophical significance, it is the possibility they have identified that concerns us. The significance to the question in this dissertation about the efficacy of belief is this: If Spencer et al. are right about the causes of gender differences on standardized tests, then they have identified a case in which stereotypes are efficacious in a mode that need not be mediated by intentional action. It is extremely important to realize that the effect of manipulating stereotype relevance here cannot be an artefact of a difference in "interpretation" by a contest judge, or of a difference in the opportunities to answer the test administrator or judge. What's being measured is quite objective and independent of a grader's discretion or interpretation: the answers the subjects gave on a mathematics multiple choice exam. Spencer et al. identify the determining factor to be the specific situation a person is put in, in which the person's performance is affected because he

or she feels a stereotype may be applied to that performance. Rather than use the term "stereotype threat", though, I would emphasize that the testing situation involved other perceivers, and describe the significant difference in the test situation in which the stereotype was made relevant as a perceiver's employment of a stereotype (or leading the subject of the experiment to believe that the experimenter was employing a stereotype) with a certain kind of feature: the feature of expecting the subject to underperform on difficult math tasks, for whatever reason. This is to say that the effect of the test administrator's employment (or feigned employment) of a stereotype on the subject's performance is pretty directly the result of feeling one is perceived to be less able than one is. On the account being proposed in this dissertation stereotypes, like beliefs, are efficacious via being incorporated into someone's schemata for perception and action. Although we will regard stereotypes as beliefs, the term "stereotype" is often used to indicate something like a compacted version of a cluster of interrelated beliefs concerning a certain kind of object of perception. These distinctions are not crucial for the points being made here.

As I've noted above, I would place more importance on the experimenter's involuntary responses during the interaction with the test subject than the researchers did. The experimenter administering the test may or may not have actually held stereotypes about gender disparities in mathematical abilities; what was varied in the experimental set-ups was what he or she told the participants about the gender-sensitivity of the exam. On my suggestion that the perceiver's employment of the stereotype is efficacious in ways other than intentional action, the stereotype actually held by the person addressing the participants in the experiment is a factor worth paying attention to. Thus, rather than simply manipulating the statements the test experimenter makes to the participants, perhaps the experimenter, or, better yet, his or her beliefs about the gender-sensitivity of the exam, ought to be varied as well. One way in which a set-up employing the same experimenter (i.e., test administrator) whose beliefs vary in different parts of the experiment might come about is as follows: an experimenter uninformed about the hypothesis being tested might

start out convinced women have less ability due to unfortunate gender socialization effects, and, unenlightened about the suspicion that stereotype threat is the cause of the gender differences that have been observed on mathematics tests, administer the part of the experiment in which the participants are told the test is gender-sensitive. Later, if that experimenter is informed about the effects of stereotype threat and becomes convinced that women really are able to do as well as men, he or she could act as a test administrator for the test wherein participants are told that the test does not yield gender differences. In both cases, the experimenter would be truthfully stating the beliefs he holds about the gender sensitivity of the test, so whatever involuntary effects of holding the beliefs about the gender sensitivity were relevant would be consistent with the statements the test administrator made.

However, the experiments actually performed are still significant in identifying a distinctively different mechanism by which stereotypes can be efficacious. This way in which beliefs are efficacious --- by the perceived person's performance being affected by the perceived threat that a stereotype may be applied to her (i.e., that her actions might be perceived by another person according to a stereotype she resists) -- is qualitatively different from the more common approach to modelling human interactions in game-theoretic approaches. Here I am thinking of the approaches taken by game theorists and complex adaptive systems theorists such as Thomas Schelling, Robert Axelrod and John Holland. A complex adaptive system model models individual agents that interact with each other and change their characteristics as a result of these interactions. The purpose of building various models of agents and systems, and running simulations using them, is to see what kinds of agent-level and system-level behaviors and patterns can evolve. Axelrod and Holland both model the mechanism by which stereotypes are

efficacious as a matter of each agent bearing a "label" by which others who hold stereotypes for agents bearing labels recognize them. [13]

To see the how different the mechanism dubbed "stereotype threat" is from such approaches, consider Axelrod's discussion in his influential The Evolution of Cooperation, in which he describes a mechanism by which a belief about how another agent behaves is self-fulfilling. Each player has labels; a label is defined as "a fixed characteristic of a player that can be observed by other players when the interaction begins" (147). A label is one of four factors that gives rise to social structure (the others are reputation, regulation, and territoriality). Labels of one particular player allow other players to make choices about how to interact with that player based on something other than previous interactions with him or her. Axelrod begins his explanation of stereotypes using a cognitive conception of stereotypes. Observable features such as gender, age, race and dress, he says

> ... allow a player to begin an interaction with a stranger with an expectation that the stranger will behave like others who share these same observable characteristics. ... the observed characteristics allow an individual to be labeled by others as a member of a group with similar characteristics [which ...] in turn allows the inferences about how that individual will behave.

---

[13] These theorists offer the usual qualifications that their models do not of course capture the richness of human social interaction, but are meant to illustrate the structure of that interaction, and are meant to be useful in exploring how behaviors of an aggregate of individual agents are related to the behavior of those individual agents. In asking what may have been left out by taking such a label-bearing model of stereotypes, I do not mean here to disparage those enterprises. Rather, I regard them as having gone some distance in illuminating that relationship and, in fact, enabling further questions.

> The expectations associated with a given label need not be learned from direct personal experience. The expectations could also be formed by secondhand experiences through the process of sharing anecdotes. (146-147)

Axelrod then shows that the use of such labels can lead to self-confirming stereotypes. In fact, stable stereotypes that are not based on <u>any</u> objective differences can arise. These can result in undesirable social arrangements, such as an inability to take advantage of cooperative strategies that would have made everyone better off, and status hierarchies in which those worst off cannot refuse to participate without doing even more poorly. Our interest here, however, is in the way in which beliefs are efficacious. On Axelrod's account, it is the actions the players take based on the inferences they draw about others based on their labels that lead to self-confirming stereotypes. He supposes that ". . . everyone has either a Blue label or a Green label" and that "both groups are nice to members of their own group and mean to members of the other group. " Then,

> The Blues believe that the Greens are mean, and whenever they meet a Green, they have their beliefs confirmed. The Greens think that only other Greens will reciprocate cooperation, and they have their beliefs confirmed. If you try to break out of the system, you will find that your own payoff falls and your hopes will be dashed. So if you become a deviant, you are likely to return, sooner or later, to the role that is expected of you. (148)

Schelling has used a similar mechanism to study the "macrobehavior" of a group of agents, in terms of their "micromotives." That is, an agent's response to the presence of others is conditional (perhaps among other things) upon which of two labels he bears. John Holland uses a similar notion, which he calls tags. Holland's tags can have several parts, but tags are like labels in that they are observable and that they are the basis for selective interaction among agents. Each agent's interaction with another agent --- e.g., cooperation, exploitation, avoidance--- is

determined by the tags borne by those interacting. In models with tags and labels, the possibilities open to an agent depend upon labels or tags -- his own and those of the other agents with whom he can interact. These studies are really thought experiments that answer questions about the kind of aggregate behavior that would result from certain kinds of interactions between individual agents.

Undoubtedly the mechanism for self-fulfilling stereotypes simulated by these models using labels is one mechanism by which self-fulfilling stereotypes develop. But, I want to argue, it is not the only mechanism leading to self-fulfilling stereotypes. There are two distinct senses in which the mechanism Axelrod described leads to self-fulfilling stereotypes. The senses in which these stereotypes are self-fulfilling when considered as labels on agents in game-theoretic models are (i) the stereotypes, though initially without basis, are confirmed by the person holding them in his or her interactions with the person about whom they are held, and (ii) the person about whom the stereotype is held actually comes to change his or her decision rules in such a way that the stereotype, though initially without basis, is true. What's important, and probably very surprising to those who have a game-theoretic notion of interaction in mind, is not only that there is another mode by which stereotypic beliefs are efficacious (i.e., so-called "stereotype threat"), but that this latter mode of efficacy gives rise to self-fulfilling stereotypes that are self-fulfilling only in the sense of (i). That is, they will tend to be confirmed by the person who employs them in the context of an interaction with another person about whom they are held, but it is not true that the person about whom the stereotype is held actually comes to change in such a way that the stereotype is true. That is, if we compare Axelrod's example of the Greens and Blues to the example of stereotype threat in women's performance on mathematical tests, it is true that, in interactions with women, someone who lets on that he or she holds a negative stereotype of the person taking the test will find the stereotype confirmed in this sense: on average, the women he or she administers difficult math tests to will probably score less well than men of comparable education and confidence. Yet, it is not true that, outside such interactions, those very same women are less

able to perform mathematical tasks; in interactions in which there is felt to be no danger of a negative stereotype being applied, such as if its relevance to the current situation is disavowed by the person perceiving them, a group of women will score as high as a group of men of similar background and confidence.

The authors of that study were interested in the practical significance of their findings: the elimination of this undesirable effect of stereotypes in educational and testing contexts. My interest in citing the phenomenon they studied is that it illustrates the point that, in the context of specific interactions, social stereotypes may well have effects on those interacting, in ways that are not accounted for solely in terms of their role in causing intentional actions. Inasmuch as features of the stereotypes one holds are beliefs, then, beliefs are efficacious in social interaction in a manner that is not covered by intentional action.[14]

To go farther in our investigation of *how* beliefs are involved in social interaction, we can ask: what about the role of beliefs not only in employing stereotypes in social interaction, but in *creating* and *maintaining* social institutions? The kind of answer sought here is not a blanket generalization meant to explain the creation and maintenance of *all* social institutions, but rather some insight into the way in which beliefs and social stereotypes are involved in the social institutions that are especially dependent upon people's expectations for their existence and maintenance --- for example, traffic manners, the fame enjoyed by a certain singer, or the prestige accorded the owning of diamond jewelry.

---

[14] Although the simulations and illustrations given by those sociologists and game theorists described above (i.e., Axelrod and Holland) do seem aimed at mimicking intentional action, I do not see anything in my suggestion to consider modes of efficacy of stereotypes that are self-fulfilling only in the sense of (i) (and not in the sense of (ii)) that is contradictory with their models. Thus the suggestion I am making could be seen as supplementing that work.

B. How Beliefs Make A Difference to the Existence of Social Institutions

Besides their effects on the individuals who are involved in social interactions, beliefs are also efficacious in establishing and maintaining social institutions.

In examining how beliefs are causal in establishing and maintaining social institutions, we are interested in what difference it makes to the existence of social institutions that a person believes one thing rather than another. The notion of cause meant here is a commonsense one. The general idea behind the claim that beliefs are efficacious in the establishment and maintenance of social institutions is that we can affect the existence of the social institution by changing people's beliefs.

The impetus for investigating how beliefs are efficacious in the existence of social institutions is that there seem to be cases in which it is because individuals in a social arrangement have the beliefs they do that something is the case. The standard examples in philosophy are that the practice of contract-making exists within a certain social group, or that something is prestigious. It is at least not immediately obvious that these states of affairs can be seen as simply the cumulative effect of the intentional actions of an aggregate of individuals. I will argue, as I did above for the case of social interaction between individuals in specific social contexts, that beliefs are efficacious in the establishment and maintenance of social institutions in ways that are not fully accounted for via their role in causing intentional action. And, in this second half of the chapter, I will show how, on the view of beliefs proposed, we have a mechanism of sorts --- the employment of anticipatory schemata --- by which the belief is involved in a mode of causation in addition to its role in causing the intentional action of individuals.

1. Three Different Claims: Coordinating, Forming, and Employing Beliefs

There are a variety of relations between higher-order beliefs held by individuals and social institutions one might appeal to in explaining how beliefs might be efficacious in establishing and maintaining social institutions, and, correspondingly, there are several different kinds of claims one might make:

(a) The existence of social institutions can be accounted for in terms of coordinated individual actions wherein the only mode of a belief's efficacy is via the intentional actions of individuals in which each individual's beliefs figure as premises in determining that individual's actions. This view is usually qualified (as in Lewis, Convention 141) so that it is not required that the agent actually follow a reasoning process involving the belief in deciding on an action; it is instead sufficient if the agent's actions could be justified in terms of a reason in which the belief is a part.

(b) In order for an individual to form certain higher-order beliefs (i.e., beliefs about other individuals' beliefs) responsible for the existence and maintenance of the social institution, it is necessary that the individual be part of a social arrangement (or, at least, be aware of the existence of that social arrangement, perhaps through secondhand knowledge).

(c) In order for an individual to employ the kind of beliefs responsible for social institutions, something other than higher-order beliefs about each other that could be cited as part of the reason justifying one's actions is involved; the employment of the kinds of beliefs involved here requires interaction with others who are in a certain social situation.

My task here is to show that beliefs are efficacious not only via bringing about intentional actions of individuals: I will argue for (c), and furthermore, that the kind of social interaction required for employing the kind of beliefs required for the existence and maintenance of social institutions involves modes of efficacy other than via intentional action.

My reasoning will proceed as follows. First, we will examine David Lewis' account of the existence and maintenance of conventions in terms of mutual concordant expectations as a candidate for an explanation of the more general case of how beliefs are involved in the existence and maintenance of social institutions. We shall see that, although it does not appear prominently in the definition of convention, in many cases the notion of a social role is actually essential to conceiving a situation in the manner required to participate in a convention. This point will be made clearer by appealing to David Gauthier's explanation of how salience works in achieving coordination when it does work to do so, and then to his explanations of participation in a joint strategy. By examining a variety of joint actions, we shall see not only that the notion of a social role is essential, but that participation in the joint action often requires interaction in order to determine what doing one's part requires of one. On the proposed account of belief, beliefs are efficacious by being incorporated into dynamic anticipatory schemata employed in perception and action. We'll see that having the kind of beliefs associated with the existence and maintenance of conventions requires understanding roles, and that these require the employment of beliefs that require interaction with others in order to be efficacious in enabling perception and action. (The distinction I will try to draw here is akin to the difference between a closed form solution in terms of input variables, and a method employing an interactive process in which a response at some point in the process may depend upon an outcome or response of something else that must be sampled or measured.) Further, I will show why the mode of efficacy of belief in the kinds of interactions involved cannot always be accounted for by intentional action alone.

In making these points, I will also examine what it means to anticipate the actions of others with whom we interact. We shall see that, to the extent to which it is true that people participating in a convention or a joint action together rely on holding similar beliefs, background knowledge, and standards of inference, the notion of 'similar' involved does not imply what we might think it does: that we are able to simulate what others will do or think in a situation. The substantial point is that a notion of role is required that is very like the notion of schema that I have developed --- I can anticipate where a ball will go next, without being able to do so by considering what I would do next were I a ball -- except that for social interaction the other person's cognitive and psychological features matter. That is, it is not generally true that when we interact successfully with others we do it by understanding our complementary roles along the lines of being able to figure out the next move of someone in another role by "putting ourselves in their shoes".

Another important notion we will introduce in our path to showing that beliefs are efficacious in social institutions in ways not accounted for by intentional action alone is the notion of acting on a joint strategy. We will elaborate on David Gauthier's discussions of how the notion of acting on a joint strategy creates new opportunities; we will complicate the notion of "doing one's part" a bit, to allow for accommodating others' idiosyncrasies, aberrations, and failures when achieving a desired joint outcome. This will further emphasize the interactiveness required in participating in joint actions. The point of interest in our case will, unlike for Gauthier, not be one of ethics, but of the causal efficacy of belief: if we conceive of participation in terms of a joint strategy, anticipatory schemata for interaction with others will include references not only to others' anticipations, but to the outcome desired. Here we will be appealing to examples such as singing in a choral group where others are singing off-key, canoeing with another paddler in rough waters in which split-second decisions are made, and so on. Here people will want to rely on whatever resources they have available to inform them of others' anticipations; again, the mode of efficacy of belief in these kinds of interactions cannot always be accounted for by intentional actions alone.

A picture will emerge on which the proposed account of belief in terms of anticipatory schemata is seen to be well-suited to account for the phenomena we want to account for in explaining how beliefs are involved in the existence and maintenance of social institutions.

2. Getting Together: Reconceiving Situations of Common Interest

David Lewis' well-known Convention proposes an account of convention that, following Schelling, involves a system of people holding higher-order beliefs, or expectations, about each other. That is, each person has expectations of other people's expectations, and expectations of other people's expectations of their own expectations, and so on. However, as we shall see, on the formal definition of convention he ends up with, the only mode of efficacy of the expectations responsible for the existence and maintenance of conventions is via intentional actions of individuals. Lewis' account involves beliefs in the existence and maintenance of social arrangements, in the first two ways identified in B.1 above: i.e., on his account, beliefs are involved in generating each agent's coordinated individual actions ((a) above) and, he suggested, the beliefs involved in a convention may be of the sort that could only be acquired in the context of interacting with others in a social arrangement ((b) above). As for (c) above --- whether employment of the requisite beliefs required social interaction --- Lewis took it to be a virtue of his account that he was able to dispense with the need for social interaction between individuals during the process in which they determine how to act in coordination problems. Although he acknowledges that the beliefs one used in the process will have been acquired in a social setting: "By our interaction in the world we acquire various higher-order expectations that can serve us as premises" (32), he is explicit about rejecting the view stated in (c) above (in B. 1). Lewis writes: "Note that replication is not an interaction back and forth between people. It is a process in which one person works out the consequences of his beliefs about the world --- a world he believes to include other people who are working out the consequences of their beliefs, including their belief in other people who . . ." On Lewis' account, one does take account of the

other individuals with whom one is dealing in a coordination problem. However, in the process that leads to an individual's action of conforming to a convention, "we are windowless monads doing our best to mirror each other, mirror each other mirroring each other, and so on" (32).

Lewis is interested in the expectations one person has of another only insofar as these expectations are involved in the genesis of the other person's reasons for choosing one action rather than another. He is upfront about this, saying something stronger, in fact: that if one person knew what actions the other was to take, he would have no interest in the other's expectations. Lewis is often precise about it, too, speaking of "expectations of the other's choice", rather than just of "expectations". However, the focus on actions may be an inessential concomitant of the fact that he is using the tools of game theory to analyze convention. Lewis describes the relationship of his approach in Convention to Schelling's work as follows:

> My theory of convention had its source in the theory of games of pure coordination --- a neglected branch of the general theory of games of von Neumann and Morgenstern, very different in method and content from their successful and better known theory of games of pure conflict . . . Yet, in the end, the theory of games is scaffolding. I can restate my analysis of convention without it. (Convention 3)

Thus, we shall have to be careful in prematurely concluding that specific features of Lewis' analysis are essential to his account of convention; this comment indicates that some of the features of his account may be general to game-theoretic approaches but inessential to his account of convention. As sorting out all the features of his approach that are essential to his analysis could take us far afield, my approach here will be primarily to focus on drawing out the insights he is trying to capture, and only secondarily to point out where I find the specifics of the account unnecessarily limiting.

Lewis' project is to explicate David Hume's view that "speech and words and language are fixed by human convention or agreement" (Morals, Appendix III 95). Hume made the remark in the context of examining the origin of justice. He rejects the view that justice arises from actual promises made, but not that justice arises from "a sense of common interest". In explicating a meaning of convention on which it is correct to say that justice arises from convention, Hume uses language and money as examples of other things that arise from convention:

> Thus two men pull the oars of a boat by common convention, for common interest, without any promise or contract: Thus gold and silver are made the measures of exhcange; thus speech and words and language are fixed by human convention and agreement. Whatever is advantageous to two or more persons, if all perform their part; but what loses all advantage, if only one perform, can arise from no other principle. There would otherwise be no motive for any one of them to enter into that scheme of conduct. (95)

As game theory is the discipline that treats interaction of two or more individuals formally, it is only natural that Lewis would look to it for a language with which to explicate Hume's view. Schelling, who was investigating how game theory could be used to inform policies to be employed in various real contexts, urged that, rather than classifying games into zero-sum and non-zero sum, we ought to classify them into games of pure conflict, games of pure coordination, and mixed games.[15] Games of pure conflict are like zero-sum games, in that the players' interests are

---

15 Another way of classifying games that has proven to be of major significance is John Nash's classification of games into cooperative and non-cooperative games. In cooperative games, the structure of the game is such as to allow free communication and enforceable contracts. The use of the terminology is not consistent among later commentators, however: Harsanyi, for instance, takes the defining criterion of cooperative games to be only the enforceability of contracts as a feature that is built into the structure of the game. Schelling sometimes treats communication as the defining criterion of cooperative games. And Harsanyi allows non-cooperative games to

opposed to each other: each stands to gain if the other loses, and to lose if the other gains. Games of pure coordination are just the opposite; here the players' interests are the same: each stands to gain if the other does, and to lose if the other does. Mixed games are conceived of as somewhere along a continuum between these two extremes; some combinations of the players' actions may involve one player gaining at the other's expense, and other combinations may involve both players standing to gain or lose together. Lewis' approach is to use Schelling's notion of pure coordination to show the sense in which Hume's insight about mutual advantage explains language as a convention.

Schelling emphasized the significance of the interdependence of expectations in coordination:

> It is to be stressed that the pure-coordination game is a game of strategy in the strict technical sense. It is a behavior situation in which each player's best choice of action depends on the action he expects the other to take, which he knows depends, in turn, on the other's expectations of his own. This interdependence of expectations is precisely what distinguishes a game of strategy from a game of chance or skill. In the pure-coordination game the interests are convergent; in the pure-conflict game the interests are divergent; but in neither case can a choice of action be made wisely without regard to the dependence of the outcome on the mutual expectations of the players. (emphasis added) (Strategy of Conflict 86)

Lewis' account is to be an account of convention in general. He redescribes convention as the solution to a coordination game. The solutions to, or outcomes of, such games are joint actions, of which each individual does his part. On this formalization, joint actions are just combinations of

---

include the ability to make enforceable agreements when explicitly stated. Thus one must be careful in comparing general statements about cooperative and non-cooperative games made by different game theorists.

actions by individual players. As the game theoretic approach dictates, Lewis takes preferences of the players as primitives; in coordination games, the preferences taken as primitive are preferences each player has over joint actions, i.e., outcomes of the game. Thus, it is in the nature of coordination problems that each agent's choice is dependent in some way on his or her expectation of the others' behavior; Lewis says that "each [agent] chooses according to his expectation of the other's choice." There is nothing in the nature of coordination games, however, that dictates that each agent has the same choice of actions to perform. In fact, in a joint action, the agents often do perform different actions. However, Lewis argues that, by appropriate redescription of the action, any joint action can be put in the form of a joint action in which all the agents do the same action: "Any combination [of actions] . . . is a combination of actions of a same kind (a kind that excludes all the agents' alternative actions)" (11).

An example Lewis uses to illustrate this claim is the coordination of actions when a phone call is interrupted and the conversants want to re-establish the connection; coordination is achieved when one and only one of the two conversants calls the other back. Lewis points out that the replacement of a choice set of two individual actions described as {"calling back", "not calling back"}, by a set of two individual actions described as {"calling back if and only if one is the original caller", "calling back if and only if one is not the original caller"} replaces a joint action in which each individual does something different by a joint action in which each individual does the same thing. Using the first set of action descriptions, coordination is achieved by one person calling back and the other not calling back; using the second set of action descriptions, coordination is achieved by each doing the same action, e.g., calling back if and only if one is the original caller. Lewis then proceeds with his analysis of convention, ultimately arriving at a characterization of "the phenomenon of convention" in which the decision faced by each person party to the convention looks exactly the same: to conform or not to conform to the convention. The significance of being able to describe the action in terms of conforming versus not conforming is that it is by recognizing other people's actions as acts of conformance to a convention that expectations of

conformance arise: "As long as uniform conformity is a coordination equilibrium, so that each wants to conform conditionally upon conformity by the others, conforming action produces expectation of conforming action and expectation of conforming action produces conforming action" (42).

The path of reasoning by which Lewis arrives at this point is as follows: conventions are regularities in behavior, and they represent solutions to coordination problems by arbitrating between competing equilibria of a coordination problem such as coordinating where to meet. In such problems, our interest is in meeting at the same place, and we care less about where we actually meet. One way to achieve this kind of coordination is of course by agreeing on where to meet. But then the problem is solved.[16] Lewis shows how people faced with a familiar coordination problem can, without communicating with each other, let alone formulating an agreement, achieve coordination: by means of their mutual expectations based on their acquaintance with past solved instances of their present coordination problems. People can be guided by analogy to do what worked before; coordination of expectations is thus based on our tendency to draw the same points of similarity:

---

16 The relationship between convention and agreement depends on the kind of agreement, and on whether or not the agreement still influences people's decisions. Lewis treats the topic in Convention, in Section 1 ("Agreement") of Chapter III ("Convention Contrasted"). The main points are that an agreement can give rise to a convention, by the process of "a growing causal chain of expectations, actions, expectations, actions, and so on. The direct influence fades away in days, years, or lifetimes." For some agreements, the convention begun by agreement only becomes a convention, he says, when "the direct influence of the agreement has had time to fade." There is a type of agreement that results in a regularity that fits his definition of convention from the start, though: "If . . . we agreed by exchanging conditional promises binding us to conform to R only if others did, or by exchanging noncommittal declarations of intent, the resulting regularity would be a perfectly good convention at once." (Lewis 84)

> Every coordination equilibrium in our new problem (every other combination, too) corresponds uniquely to what we did before under some analogy, shares some distinctive description with it alone. Fortunately, most of the analogies are artificial. We ignore them; we do not tend to let them guide our choice, nor do we expect each other to have any such tendency, nor do we expect each other to expect each other to, and so on. And fortunately we have learned that all of us will mostly notice the same analogies. That is why precedents can be unambiguous in practice, and often are. ... We are not in trouble unless conflicting analogies force themselves on our attention. [. . .]
>
> It does not matter why coordination was achieved at analogous equilibria in the previous cases. Even if it happened by luck, we could still follow the precedent set. (38-39)

We can see what Lewis has achieved here: consistently with his view of people interacting with each other as though they were "windowless monads", he has explained how coordination can be achieved without any interaction between the people achieving it. However, his explanation of convention is not given solely in terms of the formal structure of the game-theoretic description of it (which, he showed, could carry us only part-way to the solution, i.e., to the point of identifying multiple equilibria). Lewis' explanation appeals not only to the formal properties of the game, but to a rich social context in which the participants have similar experiences, similar ways of conceiving those experiences, and similar conceptions of each other, including each other's conceptions of each other. They might also have shared experiences, or perhaps even legends and stories, on which they may draw. [17]

---

[17] It seems to me that the fact that Lewis clearly appreciates that the surrounding social context is essential to his account of convention is often not recognized. What is true is that Lewis cordons off the effect of the social context so that it is involved only in generating concordant mutual expectations from a basis for common knowledge --- after that, the expectations straightforwardly produce decisions to conform, without any involvement of social context --- but Lewis certainly shows an appreciation of the subtleties of how the expectations are generated. I suspect it is because his definition of common knowledge has been of such significance among game

Surely the rich social context in which the participants live is the right thing to appeal to. Yet Lewis' explanation, which puts all the contribution from social context into expectation formation, goes so fast here it is not very clear how the social context is called upon in solving coordination problems. Lewis ends up giving an account of convention in which this rich surrounding social context is more uncoupled from the game-theoretic formulation of the game than I think is warranted for social interaction in general, so we will want to take a closer look at how people use salience in particular, and resources in the social context in general, to solve coordination problems.

To explicate what is involved in employing salience, I will compare Thomas Schelling's and David Gauthier's discussions of an example in which salience is used to achieve coordination. We shall see that there is a relationship between Gauthier's explanation of what is involved in reconceiving the situation so that salience can help and Lewis' remarks about reconceiving a situation in terms of conformance so that the expectations needed for a convention to take hold can form. The place where I think Lewis' account goes a bit fast is in assimilating participation in a joint action to performance of an individual action of a certain kind (characterized as an act of conformance). Lewis treats the process as one wherein agents use some bits of knowledge about the situation --- often culturally-based features of the situation that are not part of the formalized game-theoretic description --- to pick out one (though not necessarily the best) equilibrium from a number of coordination equilibria. The idea is that there are coordination problems that require people to choose a unique equilibrium (a joint action in which each one's action is a best response to the other's) from the possible equilibria: for example, to choose the same place to meet, from among the many suitable meeting places accessible to both people who wish to meet. What I'll argue is that it is no accident that the notion of role is involved, and that the proposed account of beliefs as

---

theorists that the discussion motivating the definition is less well known than the formal definition is.

anticipatory schemata is well-suited to explaining how beliefs are involved in joint actions that involve people in role-mediated dynamic interaction.

I should explain why I am mentioning Gauthier's discussion of coordination at this point, as I will be discussing his account of cooperation later. At one time, I thought that the key idea I was seeking about the interactiveness of social interaction was reflected in Gauthier's explanation of how cooperative interactions differ from non-cooperative interactions. He is concerned with the difference between what a rational agent should do in "non-co-operative interaction of nature and the market" as opposed to what a rational agent should do in "co-operative interaction"; the first calls for utility maximization, the latter calls for seeking an optimal outcome. The first is coordination; the latter, cooperation. However, I came to see that, for the question of what is distinctive about the cognitive structures required to effect social interaction, this is not the distinction I sought. The cognitive structures that are efficacious in social interaction are nascent in the account he gives of coordination, as well. Schelling and Lewis only hint at the cultural resources and cognitive structures agents assume of each other in settling on one of the multiple equilibria available by example; they appeal to the fact that salience can work sometimes, though they do not explain how it works when it does. In the account Gauthier gives of how salience works, the agents have to reconceive the situation in a way that they know is available to both; doing so requires some primitive version of the notion of people in different roles doing what's required to cooperate to achieve some outcome they are both interested in. The reader who is not interested in the details here can probably skip to the last paragraph of this section without losing the main thread of the discussion.

The use of salience to achieve coordination in a novel (i.e., unfamiliar) coordination problem is an extreme, specific case of the general case of using salience in a coordination problem to achieve coordination. Thomas Schelling performed experiments in which two or more people were given the novel coordination problem of choosing the same numeral as other players with whom they

were unable to communicate and about whom they know nothing except that all the players were informed of the structure of the task. In such experiments, coordination is often achieved by each player choosing a uniquely situated number: e.g., the lowest numeral, the number one. In contrast, when simply asked to choose a numeral, people generally do not choose the number one. (Strategy 94) Any number would do to achieve coordination, so long as every player chose it. It is the knowledge that no-one has anything but the structure of the game to go on, and that all know it, that gives rise to the strategy of making the choice on the basis of a feature that all are bound to notice. We can see the common sense in this reasoning. Lewis appeals to Schelling's explanation of how salience works to solve coordination problems, but remarks that "salience of an equilibrium is not a very strong indication that agents will tend to choose it" (57) and so the expectations that are generated due to noticing salience are in general weaker than those generated by an agreement.[18]

What is the proper way to generalize how salience has helped the participants achieve coordination? There is no analogue of "the smallest number" when, for instance, choosing between places to meet. Gauthier provides an analysis of Schelling's commonsense account of what is going on when salience is successfully used to coordinate a place to meet. Whereas Schelling appealed to asymmetries from outside the game to influence players by an unspecified mechanism, Gauthier gives an explanation in which salience is not a supplement to utility maximization, but actually works by utility maximization. This will become clearer by looking at an example. Schelling discusses the problem of two people coordinating where to meet, using the following pairs of payoff structures; it is assumed that each knows the other's payoffs as well as his own payoffs:

---

[18] Lewis discusses Schelling's experiments on using salience as a method of reaching a coordination equilibrium on pg. 35 of Convention.

FIGURE 1. ---Neither Cares Where They Meet Nor Where They End Up If They Don't

(after Schelling, Strategy of Conflict, Fig. 32)

	Strategy I per player C	Strategy II per player C
Strategy i per player R	R's payoff=10 C's payoff=10	R's payoff =0 C's payoff =0
Strategy ii per player R	R's payoff =0 C's payoff =0	R's payoff=10 C's payoff=10

Here the players have no way of choosing between their strategies without additional clues; the best they can do is randomize between the two strategies each can use. With clues, they can choose one of the two equilibria, provided each knows the other will recognize the same clue.

FIGURE 2. Neither Cares Where They Meet But Each Cares Where They End Up If They Don't

(after Schelling, Strategy of Conflict, Fig. 33)

	Strategy I per player C	Strategy II per player C
Strategy i per player R	R's payoff=10 C's payoff=10	R's payoff =0 C's payoff =5
Strategy ii per player R	R's payoff =5 C's payoff =0	R's payoff=10 C's payoff=10

Here the players do have a way of choosing between their strategies. R should choose Strategy ii, and C should choose Strategy II. However (Schelling says) there are two kinds of reasoning in support of this combination: utility maximization, and salience. Which is really at work?

Although utility maximization would yield the choice of (Strategy ii, Strategy II), Schelling thinks that maximization of utility is not the correct explanation of how the outcome is achieved. One reason to be suspicious of utility maximization as the explanation for selecting between equally good equilibria is that utility maximization seems unable to explain why (Strategy ii, Strategy II) could ever be chosen for the following payoff structure:

FIGURE 3. Each Cares A Bit Where They Meet But Not Where They End Up If They Don't

(after Schelling, Strategy of Conflict, Fig. 30)

	Strategy I per player C	Strategy II per player C	Strategy III per player C
Strategy i per player R	R's payoff=10 C's payoff=10	R's payoff =0 C's payoff =0	R's payoff =0 C's payoff =0
Strategy ii per player R	R's payoff =0 C's payoff =0	R's payoff=9 C's payoff=9	R's payoff =0 C's payoff =0
Strategy iii per player R	R's payoff =0 C's payoff =0	R's payoff =0 C's payoff =0	R's payoff=10 C's payoff=10

Here the players do have a way of choosing between their strategies, if they use salience. Coordination could be achieved by choosing the strategies yielding a unique payoff (although this would go against maximization of utility, according to the payoff structure shown).

Lewis' explanation of how salience can work to achieve coordination problems seems to be referring to this example of Schelling's. Lewis writes that the people trying to achieve coordination

> "try for a coordination equilibrium that is somehow salient: one that stands out from the rest by its uniqueness in some conspicuous respect. It does not have to be uniquely good; indeed it could be uniquely *bad*. It merely has to be unique in some way the subjects will notice, expect each other to notice, and so on. . . .

> . . . Their first- and higher-order expectations of a tendency to pick the salient as a last resort would be a system of concordant expectations capable of producing coordination at the salient equilibrium. (37)

Lewis' explanation follows Schelling's account of how salience works when it does work, i.e., as a method of generating mutually concordant expectations. Schelling had said that it is not because they prefer a payoff of 5 to a payoff of 0 that the people in the situation represented in Figure 2 who are trying to coordinate their strategies so that they meet at the same place choose strategies II and ii, respectively. What they are trying to do is meet, and the knowledge of where each prefers to be if they don't meet serves as a clue to help them coordinate their actions so that they do in fact meet. "It is useful to the players --- and each recognizes that the other recognizes that it is useful -- to take note of where the fives are, but only as a step in the process of coordinating intentions." He seems to be according the payoff structure some sort of psychological effect on the players' decision processes, for he says: "The tendency for the matrix in [Figure 2 above] to "converge" on (ii, II) is in principle the same as if the printed matrix had arrows pointing toward the lower-right hand corner, arrows with no logical role or authority other than the power of suggestion and hence the ability to coordinate expectations" (298).

Gauthier treats examples with the same structure as the ones Schelling discusses, but explains the success of salience in achieving coordination by means of a completely different kind of mechanism than Schelling does. In his "Coordination", he treats examples with payoff structures similar to those of Figures 1 and 2 above:

FIGURE 4. ---Neither Cares Where They Meet Nor Where They End Up If They Don't

(after Gauthier, "Coordination")

	Action I per player C	Action II per player C
Action i per player R	R's payoff=5 C's payoff=5	R's payoff =0 C's payoff =0
Action ii per player R	R's payoff =0 C's payoff =0	R's payoff=5 C's payoff=5

This situation has a "bedeviling symmetry".

(Salience can help if there is a clue each knows the other will notice.)

FIGURE 5. Neither Cares Where They Meet But Each Cares A Bit Where They End Up If They Don't

(after Gauthier, "Coordination")

	Action I per player C	Action II per player C
Action i per player R	R's payoff=5 C's payoff=5	R's payoff =0 C's payoff =1
Action ii per player R	R's payoff =1 C's payoff =0	R's payoff=5 C's payoff=5

"We could expect to coordinate on [... (ii, II) ..] ;
this would be the salient outcome. Why is this so?"

In answer to the question of why the two agents R and C could expect to coordinate in this case, Gauthier says that salience works to enable coordination by yielding a reconception of the situation in such a way that the new, reconceptualized payoff matrix has a unique best equilibrium. In the reconceptualized situation the choice is between "seeking salience" and "ignoring salience". "Seeking salience" is choosing the action that is doing one's part of the joint action that is indicated as salient; "ignoring salience" amounts to not favoring any particular action

over any other, but randomizing between the choice of actions in the initial situation. It is important to note that, unlike Lewis' example of reconceiving the actions of "call back"/"don't call back" as "call back iff one is the original caller"/"call back iff one is not the original caller", Gauthier's suggested reconceived situation yields a matrix with a unique equilibrium. Recall that the purpose of Lewis' reconceptualization was to yield equilibria in which the actions of each person participating in the joint action were identical, but Lewis' reconceptualization did not in general yield a matrix with a unique equilibrium. Gauthier's reconceptualization accomplishes both: the actions of each person in the solution to the coordination problem are the same ("seeking salience") for each person involved in the problem, and the reconceptualized situation has a unique equilibrium.

Of course the kind of reconceptualization Gauthier suggests is only possible if there is in fact some salient feature each will notice and expect the other to notice. Gauthier's explanation works as an explanation of how salience works to achieve coordination when it does, though, not only for the payoff structure in Figure 2 but for the ones in Figure 1 and Figure 3 as well. (I should point out that what's more important to Gauthier, however, are the situations in which salience will not work, in which mutual interest alone is not sufficient to yield a solution to an interaction problem.[19]) Thus, for the examples in Figures 4 and 5, if the participants can pick up on some salient feature that serves as a clue to pick out (ii, II) as the salient outcome, the decision facing them is whether to seek salience or to ignore salience. The resulting payoff structure for the reconceptualized situation is:

---

[19] The theme of Gauthier's "Coordination" is showing how coordination differs from cooperation, and that, inasmuch as the institution of promise-making can be a solution of a coordination problem, there is no reason it cannot be given a utilitarian justification.

FIGURE 6 **Gauthier's Account of How Salience Works**

	C Seeks Salience	C Ignores Salience
R Seeks Salience	R & C will meet for sure	R & C have a 50% chance of meeting
R Ignores Salience	R & C have a 50% chance of meeting	R & C have a 50% chance of meeting

In terms of utilities:

FIGURE 7 **Gauthier's Account of How Salience Works for FIG. 4**

	C Seeks Salience	C Ignores Salience
R Seeks Salience	R's payoff=5 C's payoff=5	R's payoff=2 1/2 C's payoff=2 1/2
R Ignores Salience	R's payoff=2 1/2 C's payoff=2 1/2	R's payoff=2 1/2 C's payoff=2 1/2

FIGURE 8 **Gauthier's Account of How Salience Works for FIG. 5**

	C Seeks Salience	C Ignores Salience
R Seeks Salience	R's payoff=5 C's payoff=5	R's payoff=3 C's payoff=2 1/2
R Ignores Salience	R's payoff=2 1/2 C's payoff=3	R's payoff=2 1/2 C's payoff=2 1/2

In these two examples, salience functions to allow a person to substitute a conception of the situation in which one of the several best equilibria in the original conception emerges as the unique best equilibrium. A similar reconceptualization of the situation in Figure 3 yields the following representation of the situation (Figure 9 below), which has a unique best equilibrium. However, notice that the new best equilibrium (R chooses action or strategy identified as ii in Figure 3, and C chooses action or strategy identified as II in Figure 3) is not among the several best equilibria in the original representation shown in Figure 3. The reconceptualized payoff structure is:

FIGURE 9 **Gauthier's Account Applied to FIG. 3**

	C Seeks Salience	C Ignores Salience
R Seeks Salience	R's payoff=9	R's payoff= 3
	C's payoff=9	C's payoff= 3
R Ignores Salience	R's payoff= 3	R's payoff= 29/9
	C's payoff= 3	C's payoff= 29/9

Thus Gauthier has shown that the role of salience in selecting an equilibrium can be explained in terms of utility maximization, once one realizes that where salience comes in is in how the situation is conceived. What's needed to reconstruct a coordination problem in which there is no solution is to find a distinguishing feature of one of the joint actions. Then, if the people trying to coordinate their actions or strategies know that they will both notice the distinguishing feature, and if each knows that the other will recognize that to solve the problem they will conceive of the situation as one of seeking salience, the coordination problem is solved. Notice that, although this solution appeals to the social context which exists outside the game, the extra-mathematical features are not absent from the representation of the payoff structure; it is in the choice of the conception of the situation that salience is involved.

This highlights just where I think Lewis' discussion warrants a more careful look. Lewis' use of redescription of combinations of actions to yield a description of a combination of actions (i.e., a joint action) in which everyone does the same thing can be related to Gauthier's explanation of how salience can be used to reconceive the choice one faces. Here is how Gauthier's suggestion would work for Lewis' example of a convention about which of two conversants is to call back to re-establish a telephone connection that has been interrupted. Recall that Gauthier defines "ignoring salience" as randomizing among the choices available to one in the original description of the situation. Thus "ignoring salience" is not the same strategy or action as the one Lewis called "not conforming", a term Lewis used to describe those actions that do not fit the

description of "conforming" actions. That is, using Lewis' example of replacing the choice between "call back" and "don't call back" with the choice between "call back if and only if one is the original caller" and "call back if and only if one is not the original caller", Gauthier's dichotomy in the restructured situation would be "call back if and only if one is the original caller" and "ignore the rule of calling back if and only if one is the original caller", which would mean randomizing between calling back and not calling back. "Ignoring salience" would thus include calling back when one is the original caller as well as calling back when one is not the original caller. It should by now be clear that the difference between these two suggestions for reconceiving the situation is in how negation is used to delineate a dichotomy of actions; both include reconceiving the situation in terms of a privileged action or strategy, but differ in the scope of the "not" used to create a dichotomy of actions or strategies.

The advantage of Gauthier's formulation is that the reformulation yields a unique equilibrium in cases where salience solves the coordination problem, whereas Lewis' reformulation does not yield a unique equilibrium. All Lewis argued for was a reformulation in which all agents did the same action in the target joint action. Gauthier's reformulation does more: not only do all agents do the same action in the target joint action, but when salience can be used to pick out a unique equilibrium in a coordination problem with multiple equilibria, the restructured conception of the situation will yield a payoff matrix with a unique equilibrium. I find Gauthier's way of applying the negation seems more natural if the choice one is pondering is whether to use the methodology of salience or not to try to use salience, or if one is wavering between choosing to conform or to be indifferent to conforming; Lewis' way of applying the negation seems more natural if one deliberately wants to break with convention, or if the choice one is faced with is between two equally good equilibria -- on his reconception, each person choosing "don't conform" works just as well as if each person chooses "conform". So we need Gauthier's kind of reconception to get a matrix with a unique equilibrium. Gauthier is concerned to explain that it is utility maximization that is at work when salience is successful; but we could just as well use his explanation to point

out how beliefs are involved in achieving coordination. Not, as Schelling suggested, by the extra-game knowledge exerting some sort of psychological "push" that tilts all the players towards one of several possible equilibria, but by their role in producing intentional action (by showing how to reconceive the payoff matrix). Once one has constructed Gauthier's reconception to reconceive the situation so that it has only one equilibrium, one can map the action or strategy of "seeking salience" to Lewis' "conforming" by a suitable change in scope of the "not", and obtain the conception of the situation Lewis presents. However, Lewis' reconception in terms of "conform"/"don't conform" doesn't explain how a coordination problem is solved by salience in the first place, only how a convention can form once people do manage to use salience to solve a coordination problem at some point or other.

I have presented Gauthier's explanation of salience to help us examine what is involved in the reconception of a situation through which salience operates. The fact that the original situation has multiple equilibria reflects that there is more than one way for the outcome to be achieved. In order to come up with the reconception of the situation Gauthier describes, each person has to have the notion of an outcome to be achieved jointly, and recognize that the outcome may involve others differently from himself or herself. Gauthier is concerned to point out that using salience to settle on one of the equilibria does not require anything more of an agent in terms of motivation than what is required by utility maximization; I am interested in asking what is required of the agent in terms of cognitive resources over and above what is required in the original (game-theoretic description of the) conception of the situation. The move to "seeking salience" is a move from choosing among individual actions whose consequences to oneself (payoffs) vary with the actions of others (as they might vary with the weather), to choosing a joint action or strategy based on some feature of the situation that one has discerned has some significance to others in the situation, and reconceiving the choice in terms of doing one's part of that joint action or strategy. In very simple cases, performing one's part of a joint action can be performing a particular action one could conceive of without reference to the others' actions; the significance

of the move in seeking salience is not readily seen in these cases. But, in the general case, it will be crucial that the terms in which the actions comprising the joint actions are described appeal to roles. The people involved must either have, or be able to develop, such notions. I will develop this point in the sections that follow, and what I will then argue is that social interaction in which roles help to structure participation in joint action requires of the participants cognitive resources that can be described as dynamic anticipatory schemata .

This line of thought develops some suggestive remarks Lewis makes on the way to settling on his definition of convention. Lewis makes the point that agents draw upon past experiences in a rich surrounding social context when conceptualizing a situation in order to coordinate actions with someone else. He mentions not only salience, but past precedent and even fictional events, as means by which people might conceptualize the decision facing them in a specific situation so as to choose from among many possibilities in the same way that others will choose. Lewis' statement that "Any combination, equilibrium or not, is a combination of actions of a same kind (a kind that excludes all the agents' alternative actions)" implies that any combination of agents' actions can be redescribed as conforming, in such a way that all the other actions are non-conforming. He remarks that this is less informative than one might think, for the description of the actions the different people party to the convention perform might not strike us as a natural one: "Whether [the given combination of actions] can be called a combination in which every agent does the same action depends merely on the naturalness of that classification" (11). It seems to me Lewis lets slip through his fingers here a notion that does distinguish descriptions of combinations of actions that are arbitrary from ones that are natural. This is the notion of a combination of actions being structured by roles that agents actually have incorporated into their schemata, or which they find quite natural to incorporate into the schemata they have already developed in the process of being socialized and acculturated to their surroundings. Lewis cites the example of the convention about which person is to call back to re-establish an interrupted phone call as one of the conventions for which we would not consider it natural to call

conformance to the convention the same action for the two people involved, remarking that the description is unnaturally complex because roles of the participants may be involved in reconceptualizing the situation in order to achieve coordination (43). He does not seem to regard the need for the concept of roles in describing an action as anything more than additional complexity of an action description. He seems to think it a bit deflationary of his point that, when we have to specify "what we would naturally call different actions for agents involved in situation S in different roles", we may have conforming actions that "do not share any common natural description" and thus, we may not be able to avoid using an action-description that is "unnaturally complex" (44).

The important feature of these reconceptions for our purpose, though, is not whether or not there is a formal description for any given combination of actions, but what kind of beliefs are required of an agent in order for her or him to be able to reconceive the situation as required to establish and participate in a convention. In the original conception of the situation the choices one has to decide between are: (i) to call back, and (ii) to stay off the line and wait for the other person to call back. It is not an essential part of speaking on the phone that one knows who initiated the call; for instance, people whose work requires that they speak to each other many times a day and who often leave phone messages for each to call the other back may not have any reason to pay attention to who initiated a particular phone connection. When several people are in the same room, or when someone has a co-worker who occasionally receives and places calls for her or him, the phone might get passed to someone who doesn't know which party initiated the call. However, for the convention Lewis describes to work, everyone party to the convention must be able to figure out when he or she is the original caller, in addition to knowing the convention and how to perform the actions required (i.e., how to call back, and about how long to wait for someone to call back). For the convention Lewis describes to get under way people must either already have, or be able to come up with, the relevant roles of caller and callee. The issue for us here is not how natural or unnatural it would be to regard all the acts of

conformance as the same action, but the beliefs and capabilities the agents must have in order to develop and participate in conventions.

Now, in the example Lewis gave, this determination is relatively straightforward: being the original caller is not difficult to determine if one makes a point of it; once one knows that identifying the original caller is relevant when speaking on the phone, one's schema for engaging in phone conversation includes picking up the information as to whether the end of the line one is on is the end from which the call was initiated or not, and this will probably come to be done routinely, with little effort. The reconception of the situation involves both the recognition of the roles each person occupies (caller or callee), and what the strategy chosen (seeking salience on Gauthier's account, conforming on Lewis' account) requires of each person. Each needs to know what the strategy chosen requires of the roles he or she expects to occupy; he or she will have some expectations of those in other roles, but in general need not know all that is required in order to occupy these other roles. (If, say, a young child is expected to answer the phone, but not permitted to place phone calls, the child could do her or his part even if she or he does not know how to place phone calls. In this case, being party to the convention would mean the child has developed the appropriate expectations about who is to call back, and who is to stay by the phone and keep the line free for the other person to call back.) These role-mediated interactions involve using schemata for interacting with others: e.g., a schema might direct our information pickup (such as who initiated the call), generate anticipations (listening for the phone call to be returned if one did not initiate the interrupted call), and direct our actions (to return the call if we were the one who initiated the interrupted call). It is quite natural to say that the existence and maintenance of the convention depends upon whether or not people in a social convention have the appropriate beliefs, and that having the appropriate beliefs includes having developed such anticipatory schemata.

It seems to me that Schelling, too, could have developed this point: he hints at it by way of example, without explicitly mentioning it. In arguing that two people named A and B who are given the task of agreeing on how to split a sum of money or forfeit any of the sum might coordinate by any clue they happen to light upon, he offers as an example that they see a sign on a blackboard or a note on a bulletin board referring to two people named A and B (Strategy 282). Although he does not discuss what is involved in making such a clue work, we can see that each would have to either conceive of himself as A and the other as B, or vice versa. Each has to see that the note refers to some joint action they can do by dividing up the joint action into parts such that the joint action is accomplished if and only if each does his part, and that in order for the note to be a clue, each one of them has to have a conception in which each takes the clue to apply to himself in the same way that the other person takes it to apply to him. That is, they have to conceive of their actions (their own and the other's) in terms of a role that makes sense only in light of how the roles fit together into a coherent way to accomplish a joint action in which they have a common interest.

In order to use the clue, they have to already have some idea of people in such roles -- in Schelling's example, some role in which one shares a certain sum of money in terms of an arbitrary proportion. In this case, one feature of the role is that you allow someone else to have a certain percentage of a sum of money in order to ensure you get some of it. One such role is common: that of an employee who submits to having taxes or union dues taken out of his or her salary as a condition of receiving the paycheck. Another might be the different percentages of ticket sales people in a touring show receive; one takes a lower proportion than the crowd-attracting star gets as a condition of being part of a troupe with that crowd-attracting star. The schemata for this role then might be combined with a slightly modified version of a role of being a bidder at an auction, where one's refusal to part with a little more money than one would like means the whole transaction is off. At any rate, the point is that in order to use the clue to figure out what to ask for in Schelling's experiment, one needs to identify roles. And, this involves drawing on one's experience of the kinds of roles in which one has observed people. Schelling hints at the kind of

clues people might use to arrive at mutually concordant expectations without discussing roles, but the examples he offers to argue for appreciating asymmetries in game theory seem to illustrate the point. The symmetrical case is then a special case in which one does not identify more than one role.

Once Lewis has established that there are ways in which mutually concordant expectations can be generated, he does not need to delve farther into what is involved (from the agents' points of view) in achieving a representation of a situation that meets the form of the coordination problem he discusses, i. e., one in which the action description for all the agents participating in the joint action is the same. His formalization of convention appeals only to the fact that a representation of the sort he calls the standard coordination problem must always exist, and that there is empirical evidence (such as Schelling's experiments on how people use salience to achieve coordination) that people do manage to achieve such common representations when they need to. Then Lewis proceeds to show how the standard coordination problem can be solved by the employment of a system of mutually concordant expectations; I will describe his formal definition of convention in the next section. What I want to flag at this point is that although his solution does recognize the role of belief and social stereotypes in <u>generating</u> mutually concordant expectations, he does not discuss the employment of notions of social roles or of social stereotypes <u>during</u> social interaction; in participating in a convention, each person is supposed to be able to work out what the others will do, as a "windowless monad". The kind of roles this allows are ones in which no interaction is required in order to establish or determine a role: being the person who initiated a telephone connection, or being a driver on a road in a certain country are examples of such roles.

3. David Lewis' Definition of Convention

In Lewis' definition of convention, neither the behavior alone, nor the fact that the behavior is based on expecting others to conform determines whether something is a convention or not. The regularity must be due to higher-order beliefs and conditional preferences of the sort that characterize a convention. Lewis' analysis of convention takes as its starting point the solution of coordination problems, conceived of as problems of interdependent decision by a group of individual agents whose decisions are to be made on the basis of their preferences and beliefs, according to some standard of rationality.[20] Higher-order beliefs and conditional preferences (i.e., preferences of one person to conform conditional on some others' conforming) are invoked in accounting for the qualitative difference between a convention and a mere regularity in behavior. Here is his final definition of convention:

> A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that in any instance of S among members of P,
>
> (1) almost everyone conforms to R;
>
> (2) almost everyone expects almost everyone else to conform to R; and
>
> (3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
>
> (4) almost everyone prefers that any one more conform to R, on condition that almost everyone conform to R;
>
> (5) almost everyone would prefer that any one more conform to R', on condition that almost everyone conform to R',

---

[20] The full description of coordination problems is: "situations of interdependent decision by two or more agents in which coincidence of interest predominates and in which there are two or more proper coordination equilibria "(Lewis 24).

> where R' is some possible regularity in the behavior of members of P in S, such that almost no one in almost any instance of S among members of P could conform both to R' and to R. (78)

The common knowledge provision excludes cases wherein each person acts according to his expectation that others will conform to a regularity, but does not believe that others are acting on their expectations of his behavior (e.g., where they may be acting out of habit). Common knowledge is defined as follows:

> Let us say that it is common knowledge in a population P that _____ if and only if some state of affairs A holds such that:
>
> (1) Everyone in P has reason to believe that A holds.
>
> (2) A indicates to everyone in P that everyone in P has reason to believe that A holds.
>
> (3) A indicates to everyone in P that ____.
>
> We can call any such state of affairs A a *basis* for common knowledge in P that ____. (56)

That the state of affairs A will indicate the appropriate higher-order beliefs to a person, Lewis adds, relies on that person having certain inductive standards and background information; that the state of affairs A will serve as a basis for common knowledge requires the mutual ascription of them, as well as of rationality.

Lewis' example of a case excluded by the common knowledge requirement is that of a roadful of drivers who drive on the right because each expects the others to, but each of whom thinks all the others drive on the right only by habit, and that all these others would drive on the right no matter

what they expected others to do. Since these drivers would not have the appropriate higher-order beliefs required by the definition of convention, this case of a roadful of drivers driving on the right hand side of the road would not count as a convention. Examples of a basis for common knowledge are: an agreement (say, to meet in a certain place at a certain time), salience (a uniquely conspicuous solution, say, a place everyone notices, and would expect others to notice), or precedence (where the people involved met last time). It is only due to the common knowledge requirements in the definition of convention that the higher order beliefs are part of what it is for a regularity to be a convention.

Lewis' analysis is a sort of reduction: it explains the existence of a characteristic of a social group (one of its conventions) in terms of psychological features common to all of the individuals in that group. The psychological features of the individuals in the group are: how an individual's preferences are ordered, what beliefs each individual holds about the others' preferences, and the associated higher-order beliefs each individual has (expectations of others' expectations). This reduction requires that the relevant psychology, beliefs and expectations --- and therefore the decision reached and (under some description) the action taken ---- are the same for each individual participating in the joint action that solves the coordination problem.

4. Rationality, Equilibrium, and Compatibility of Expectations

The requirement that items (1) through (5) in the definition of convention are to be common knowledge gives rise to two features: First, the feature that the expectations of the people involved in the convention are compatible with each other. Second, the feature that, under some

description, the relevant psychological attributes (such as certain expectations, preferences, background knowledge, and methods of reasoning) of the people involved are similar.[21]

The requirement that expectations are to be similar means that roles are not a matter of different clusters of expectations had by different people, but that the notion of role is built into complex descriptions of expectations. On this way of treating roles, we would get a rather peculiar account of the difference between citizens and non-citizens: even if I am a citizen of a certain country, my expectation about being able to vote is to be considered the same expectation had by someone who is not a citizen of that country: i.e., for all of us, the expectation is described as: "If I am a citizen of this country, I'll be permitted to vote in the upcoming election", rather than, say, some of us having the expectation of being let into the polling booth, and others of us having the expectation of being turned away from the polling booth. On Lewis' account of conforming actions, the notion of role shows up in complex descriptions of actions.

We can separate the two notions, i.e., the notion of the compatibility of expectations of those involved in a joint action and the notion of a common psychology that the agents involved in a joint action must share. Schelling gives an example that illustrates how these notions can be separated in his essay "For the Abandonment of Symmetry in Game Theory". In that essay, the

---

21 Here I am referring to the fact that items (1) through (5) include statements not only about regularities in behavior, but about expectations ("almost everyone expects almost everyone else to conform to R") and preferences ("almost everyone has approximately the same preferences regarding all possible combinations of actions") (Lewis 78). Of course the context implies that the expectations and preferences referred to here are those associated with the coordination problem that the convention under consideration solves. Also, the way common knowledge works is supposed to be by a common psychology, such that everyone forms the same expectations from the state of affairs that is the basis for the item of common knowledge: "mutual ascription of some common inductive standards and background information, rationality, mutual ascription of rationality, and so on" (Lewis 56-57).

notion of symmetry amounts to the notion that the agents are not distinguished with respect to their psychology; the reference is to John Nash's "symmetry axiom" in his 1953 "Two-Person Cooperative Games". The formal statement of Nash's Symmetry Axiom is "The solution does not depend on which player is called player one. In other words, it is a symmetrical function of the game." Nash explained the symmetry axiom informally as a statement that "the only significant (in determining the value of the game) differences between the players are those which are included in the mathematical description of the game, which includes their different sets of strategies and utility functions" (137). Whereas Nash stated what symmetry meant in terms of mathematical features of the game and its solution, Harsanyi related it to the psychology of the players, in what he called a "Symmetry Postulate": "The bargaining parties follow identical (symmetric) rules of behavior (whether because they follow the same principles of rational behaviour or because they are subject to the same psychological laws)" (149) . Harsanyi argues for the symmetry postulate by appealing to the intuition that "a rational bargainer will not expect a rational opponent to grant him larger concessions than he would make himself under similar conditions" (quoted in Schelling "Abandonment" 219). Against Harsanyi's claim that it should be regarded as axiomatic that symmetry be a necessary concomitant of rationality of the agents involved, Schelling points out that the intuition Harsanyi appeals to presumes that "the only basis for [one party's] expectation of what he would concede if he were in the other position is his perception of symmetry" (219). Whereas, says Schelling, rationality does not require this: "Both players, being rational, must recognize that the only kind of 'rational' expectation they can have is a fully shared expectation of an outcome" (219). He offers the following example in illustration of the point:

> Specifically, suppose that two players may have $100 to divide as soon as they agree explicitly on how to divide it; and they quite readily agree that A shall have $80 and B shall have $20; and we know that dollar amounts in this particular case are proportionate to utilities, and the players do too: can we demonstrate that the players have been irrational?
>
> We must be careful not to make symmetry part of the definition of rationality...

> Specifically, where is the "error" in B's concession of $80 to A? He expected -- he may tell us, and suppose that we have means to check his veracity .... --- that A would "demand" $80; he expected A to expect to get $80; he knew that A knew that he, B expected to yield $80, knew that B was psychologically ready because he, B, knew that A confidently expected B to be ready, and so on. That is, they both knew --- they tell us --- and both knew that both knew, that the outcome would ineluctably be $80 for A and $20 for B. Both were correct in every expectation; the expectations of each were internally consistent and consistent with the other's. We may be mystified about how they reached such expectations; but the feat claims admiration as much as contempt. ....we cannot, on the evidence, declare [the hypothesis of the rationality of A and B] to be false.
>
> ... What we have at best is a single necessary condition for the rationality of both players jointly; we have no sufficient condition, and no necessary condition that can be applied to a single player.
>
> Nor can we catch them up if we ask them how they arrived at their expectations. Any grounds that are consistent would do, since any grounds that each expects the other confidently to adopt are grounds that he cannot rationally eschew. (220)

Thus rationality does not require that all individuals have a common psychology; what rationality requires of the individual with respect to his expectations of others is that those expectations be as accurate as possible. The requirement need not constrain the expectations two people interacting may have to a unique set of expectations; rationality constraints on the expectations of an outcome to be determined by two people jointly require only the compatibility of the expectations of those involved.

But, the fact that two people have reached an agreement on how to divide the reward indicates that, whatever their expectations, they have coordinated on an equilibrium. What about the constraints imposed by equilibrium, then? Does it follow from the assumptions under which an

equilibrium is defined that players adopt the same strategy when in the same situation, including when they switch places with each other? The answer here depends on what is meant by equilibrium. What Lewis calls an equilibrium[22] is a Nash equilibrium: a set of strategies in which each person's strategy maximizes his or her own utility, given the strategies the others have chosen; i.e., each one's strategy is a best response to the strategy of the other (Nash, "Non-Cooperative Games"). That is, looking back after the outcome of the game, the person would say "It turns out I used the best strategy (given what the others' strategies were)." It is true that the notion of a Nash equilibrium is defined within a mathematical theory in which symmetry is part of the definition of the situation to which the concept of equilibrium applies. Thus, someone could argue that, if we want to make use of the concept of a coordination equilibrium, we must accept symmetry, and hence a common psychology among the participants.

However, although this line of reasoning holds for the concept of a Nash equilibrium, there is an alternate notion of equilibrium in which compatibility of expectations, but not a common psychology, is required. This is the notion of an expected outcome such that, if each of the agents expects a certain common outcome, that outcome results. The details of how this would be formulated for different kinds of assumptions are not important here: the concept that this alternate notion of equilibrium captures is the requirement that neither would revise his or her expectations upon discovering the outcome. That is, each would look back on the expectations he or she had during the interaction, and say, "It turns out I was right about that." As it captures the notion that each person's expectation is one he or she would not change in light of the other person's expectations, it has been called the expectational concept of equilibrium (Phelps 225).

---

[22] Lewis defined equilibria as "combinations in which each agent has done as well as he can given the actions of the other agents. In an equilibrium combination, no one agent could have produced an outcome more to his liking by acting differently, unless some of the others' actions also had been different. No one regrets his choice after he learns how the others chose. No one has lost through lack of foreknowledge" (Lewis 8).

Now, if those faced with the coordination problem do not have available any more knowledge than is in the formal mathematical description, neither can communicate with the other, and each knows that the other has nothing else to go on, the players may well perceive the Nash equilibrium as the expected outcome, if it is unique. In such a situation, as there would be no asymmetry in the situation, there is nothing to distinguish the two players. Then it is hard to argue that there is any other perceived outcome that would trump the Nash equilibrium, that is, in Nash's terminology, the maximization of the players' utilities -- or, in the terms Lewis uses in his definition of convention, satisfying everyone's preferences. On this point, most everyone agrees.

It is just beyond this point that Lewis' account of convention takes up: what Lewis provides is a general account (based on Schelling's insight about exploiting asymmetries) of what needs to be added to the psychology of the agents in order for there to be a solution to a coordination problem that has multiple Nash equilibria. Lewis generalizes Schelling's insight that agents would exploit any asymmetry they can light upon, as an alternative to requiring asymmetry in the agent's psychologies or communication between them as they interact in the situation S. Lewis' answer in his definition of convention is that what a convention involves over and above the normal game-theoretic assumptions is common behavior and common psychology among the agents and the common knowledge of them (items (1) through (5) in the definition of convention given above). Lewis explains how it is that, faced with a coordination problem with equally possible and attractive Nash equilibria, all do their part to effect one and only one of the joint actions among the several that are Nash equilibria. But do we need to use the concept of a Nash equilibrium? Lewis said himself that game theory was "mere scaffolding" and that he could restate his analysis of convention without it (3). Could Lewis use the alternate notion of equilibrium in a restatement? And, if so, could he jettison the requirement of a common psychology in his restatement?

To explain this alternate notion of equilibrium --- the expectational notion of equilibrium -- let us return to the basic coordination problem. If the situation facing the people trying to coordinate their actions or strategies contains anything on the basis of which they could form an expectation of the outcome, there is no reason they might not successfully use that, rather than only the structure of the game, to coordinate their actions or strategies. The expected outcome they both perceive may or may not be a Nash equilibrium. But what must be true is that their expectations are compatible and that they both expect the same outcome, and that the fact that they hold such expectations leads to the outcome that they expect. In technical terminology, this is the requirement that the function mapping the set of the players' expectations onto the outcome map the outcome the players expect onto itself. That is, if the players expect a 50-50 split as an outcome, the function mapping their expectations onto the outcome will map the expected outcome of a 50-50 split onto an actual outcome of a 50-50 split. Let us use the notation of an "X-Y split" to indicate that player A gets X% of the portion to be split and player B gets Y% of the portion. If the players have a compatible set of expectations including that they expect a 20-80 split as an outcome, the function would map their expectations of an outcome onto an outcome of a 20-80 split. Thus both these cases satisfy the requirement and are equilibria according to the expectational concept of an equilibrium.

Notice that this is not the claim that there is no restriction on their expectations; we do not get a mapping of expected outcome onto actual outcome for the case of one person expecting a 30-70 split and the other expecting a 40-60 split --- here, the outcome would be zero, for they would not have settled on a solution that divides up the total sum. For these cases, they would look back on their expectations about each other's expectations and say: "Oh, I was wrong about that." On this more general notion of equilibrium, the requirement can be stated in terms of the mapping f between sets of expectations of the outcome and outcomes; the requirement for equilibrium would be that f({ joint-o, joint-o}) = joint-o, where joint-o denotes the joint outcome, and the ordered pair within the set brackets ({joint-o, joint-o}) indicates the state of affairs the participants

expect to be the joint outcome. The players must expect the same joint outcome, but they need not expect the same individual outcome. In short, although their expectations must be identical with respect to the joint outcome, they need not be identical in every respect: they do need to be compatible in every respect relevant to producing the outcome, but there need be no justification for why there are differences between them as to what each expects for himself or herself.

The reason we might want to separate off the notion of the compatibility of expectations from the notion of common psychological features is that it seems an important feature of many joint actions that different agents participating in the joint action perform very different actions. Lewis recognized this in a way; as I mentioned above, he allowed that, on any natural description of action, the required conception of a coordination problem sometimes involved roles such that each agent does something different. Our question here is whether the difference in role between participants can be consolidated into action-descriptions such that the agents can all have identical psychologies, that is, the same expectations and strategies. The example of convention Lewis treats in detail throughout his exposition in Convention is of driving on the right hand side of the road, for which the question of which role someone party to the convention takes on does not arise (the only roles are driver and non-driver). He lists more complicated conventions, but tends to treat them that way, too: for instance, example (8) of a convention is of people cooperating in a hunt ("Rousseau's stag hunters"). Lewis' description does not include any structure within the hunting party; i.e., that the enterprise of participating in a hunting party necessarily requires that some of the participants will be leaders, some followers: "Each must choose whether to stay with the stag hunt or desert according to his expectations about the others, staying if and only if no one else will desert" (7). As the example is described, the only action to be chosen is to stay or not to stay in the hunting party.

However, if we think about the situation realistically, the relevant action-description for joining the hunt may involve far more than is at first apparent. For example, suppose there is specialization of

functions within the party. Certainly whether or not the hunting party is formed is an important part of the description of the joint outcome on which all must be consistent. But this does not mean the joint action description here must be whether a sufficient number to form a hunting party joins or not. The point can be stated in terms of the example of the game of splitting some reward if two can agree on how to split it. Here, the point would be that, although the description of the outcome in an equilibrium certainly includes the fact that the reward is split between them, it has to include more detail as well, namely, the proportion each is to get.

The choice one is faced with in Rousseau's stag hunting example could perhaps still be described as Lewis describes it if everyone is indifferent to the position he or she holds in the hunting party, or if there is very little division of labor in the party. But if the risks vary enough between different positions within the hunting party (e.g., being the scout, being the one to distract the animal versus being the one to shoot an arrow from a safe distance), one person's decision to join the party or not may depend on whether one feels he or she could do well in the role joining the party is expected to involve. Since Lewis, in treating other examples, allowed that the specification of conforming or not conforming could involve roles, if it is clear which role one would be taking on in joining the hunt (this could happen if a person's role in the hunt is determined prior to the hunting party being formed, such as by one's position in society or by precedents from previous hunts) the description of the act of conforming could be made sufficiently complex to include the relevant details on which one bases the decision. But we can see how the complexity of the decision then begins to spread: one's preference to join might also depend upon which of the others are to take on which roles (e.g., who is to be the leader of the hunting party, who is to be the one shooting the arrow in a timely manner so that the one distracting the animal is not attacked). Thus, in an actual situation, one's conditional preferences might not be as structurally simple as a conditional preference to join the hunt if and only if all the others do. The point is that even if we grant the possibility of formally describing actions in the same way for all participants in the convention, the structure associated with the interrelation of roles then becomes significant.

We cannot formalize the situation in such a way that there are no distinctions between participants; differences in their psychologies show up somewhere or other.

David Gauthier's discussion, which allow people's preferences to enter agreements to vary with who is offering the deal, is a step in the right direction towards delving into the question of the complexity a person's preferences to participate in a convention might have if that person wants to take into account the differences in how different people will carry out their part of the joint activity. Along the way to providing a normative account of the basis for morals in his Morals By Agreement, he gives an analysis of the social institution of contract making.[23]

In Gauthier's analysis, the practice of contract-making occurs in a society composed of a sufficient proportion of what he calls constrained maximizers. Constrained maximizers are contrasted with straightforward maximizers; straightforward maximizers are characterized as always acting to maximize their utility. A constrained maximizer is a maximizer in that he enters into contracts which maximize his expected utility, but he is constrained in that he refrains from reneging on his part of the contract, even if reneging maximizes his expected utility. Gauthier is then concerned to show how, and under what conditions, it is rational for a straightforward maximizer to convert to constrained maximization. My interest here is in how beliefs are involved in the emergence and maintenance of the institution of contract-making, on his account.

Gauthier emphasizes the difference between employing an individual strategy and acting on a joint strategy. A joint strategy, he points out, genuinely differs from the coincidence of individual

---

[23] Gauthier's account is a piece of rational reconstruction, that is, it is meant to give an account of what it would be rational to do, and was not meant to serve the purpose for which I'm using it---of investigating a question in philosophy of mind about the role of belief. I use it here for its helpfulness in finding a description of the role of belief in social institutions, not to judge its value in answering a question it wasn't meant to answer.

strategies: ". . .a set of strategies, one for each person, may always be represented by a joint strategy, but not conversely. . .; thus, "In extending the range of strategies from individual to joint, we have implicitly altered our previous characterization of co-operative interaction, as having a set of strategies, one for each person, as the object of agreement." And so, ". . .this new characterization broadens our initial conception of co-operation" (Morals By Agreement 120).

Informally put, the difference is that, in agreeing on a joint strategy, what those involved in the agreement choose is an outcome that is to be achieved by them as a group, and this outcome in turn determines each person's choice of strategy. He offers examples:

> We may think of participation in a co-operative activity, such as a hunt, in which each huntsman has his particular role co-ordinated with that of the others, as the implementation of a single joint strategy. We may also extend the notion to include participation in a practice, such as the making and keeping of promises, where each person's behavior is predicated on the conformity of others to the practice. (166)

Lurking in this paragraph, in the "We may also extend this notion", is an important idea: that the practice of making and keeping promises is analogous to the team-like activity of hunting.

Let's take the analogy slowly, with an intermediate step. One could first make the analogy between participation in the cooperative activity of the hunt, in which each participant fulfilled a role---agreeing (implicitly or explicitly) to adopt a certain course of action and then following it---and participation in carrying out a specific promise or contract, in which each party to the promise or contract agrees to perform a role; think of a landlord-tenant contract, or a musician's recording contract. Each party's role ---how the actions he is to take are related to the other party's actions--- is spelled out in the contract.

These relations to others are, in one sense, relations to specific individuals, and in another sense, are relations to others only in virtue of the roles they occupy. One's relation to the other parties involved in the hunt or contract is in terms of each one's role: a huntsman is to respond to the scout's report even if the scout position is a rotating one; the tenant pays the rent to the new landlord if the building is sold. However, the relations to others are to specific individuals inasmuch as one chooses who to join in the joint activity with, or (if explicit agreement is not involved) whom he chooses to regard as a holder of that role. Hunters will join in only with others they regard as sufficiently competent and trustworthy; parties to a contract will consider whether the specific person they are making a contract with is worthy of the role---whether tenant, landlord, recording company, musician, etc.

Now, a further analogy can be made: the one between participation in a cooperative activity, such as a hunt, and participation in a practice such as promise-making or contract-making. Here the role a participant takes on is, rather than that of a party to a particular promise or contract, the role of someone who can be party to one. To make this clearer, think of the difference between being a participant in a particular conversation and the more general role of being a speaker of the language (a particular language, of course). The role of being a member of a social group in which the practice of contract-making exists is the role of being trustworthy. If one is not part of a social arrangement where such a practice (contract-making) exists, then responding to others as a trustworthy person would respond, inasmuch as it's even possible, wouldn't be being trustworthy. It might even be acting looney, Don Quixote-style.

As I said, Gauthier is concerned to explain how a rational person is led to become a constrained maximizer. Without going into the details of his quantitative analysis, one of his results is that a person finds it rational to become a constrained maximizer only if a sufficient proportion of the population are constrained maximizers. The analysis takes into account that a person cannot perfectly judge whether another person is a constrained maximizer or not.

So the cognitive figures in this account of a person's conversion to constrained maximization in that his beliefs about, or perceptions of, others' trustworthiness are involved. Gauthier's story goes something like this: A person sees the benefits of cooperating with trustworthy partners in fair and optimal enterprises. Since he can only be admitted to such deals if others count him among the trustworthy, he decides to develop the "disposition" to cooperate with other individuals he judges (i.e., believes, or perceives) to be trustworthy, when offered the chance to join in with them on fair and optimal enterprises. However, he realizes that the benefits of developing such a disposition depend on his ability to detect whether others are constrained maximizers or not. If his ability is perfect, there is no question about it: being a constrained maximizer who cooperates only with other constrained maximizers is the best way to go. But, if he has less than a perfect ability, then he will lose every time he errs and mistakes a straightforward maximizer for a constrained maximizer. If he isn't very good at telling constrained maximizers from straightforward maximizers, he can still do well if the proportion of straightforward maximizers around is small, but will do poorly if the proportion of straightforward maximizers around him is high.

So, the story continues, he should convert to constrained maximization only if the proportion of constrained maximizers in the group is sufficiently high.[24] Thus, his decision as to whether or not to convert to constrained maximization will depend on his perception of others' "dispositions." Thus, one's perceptions of others are involved in two ways: (i) perceptions of specific individuals are involved in deciding who to engage with in cooperative enterprise and (ii) perceptions of how widespread free-rider tendencies are in his society are involved in deciding whether to become (develop the disposition of) a constrained maximizer at all. Lewis' account of convention includes

---

[24] The proportion that is sufficiently high will be relative to one's ability to judge if others are constrained maximizers or not. And, Gauthier says, a constrained maximizer should develop and hone this ability.

the latter kind of belief, but not the former. Perhaps that is as it should be for the special case of convention; but for the more general case of social interaction we will want to include (i) as well as (ii).

Returning to the analogy of participation in a cooperative venture such as a hunt, becoming a constrained maximizer is deciding to be a cooperator. And, cooperation involves making decisions differently: "A person co-operates with his fellows only if he bases his actions on a joint strategy; to agree to co-operate is to agree to employ a joint rather than an individual strategy (166). " The beliefs, or perceptions, of others' trustworthiness, are involved in deciding whether to cooperate, but are not part of what it is to base one's actions on a joint strategy, as Gauthier sees it. For him, to base one's action on a joint strategy is to choose to do what the joint strategy requires him to do, and this conception of basing one's action on a joint strategy need not include reference to expectations about how others will act. For, he says, "Normally, of course, one bases one's action on a joint strategy only if one expects those with whom one interacts to do so as well, so that one expects actually to act on that strategy. But we need not import such an expectation into the conception of basing one's action on a joint strategy" (166). [25]

On his account, the institution can exist only when agents can see each other as cooperators (being a co-operator entails being trustworthy). Since to be a cooperator is to be capable of acting on a joint strategy, this is an advance of sorts over acting on an individual strategy. The advance seems to be this: being a cooperator is a matter of employing an optimal joint strategy, versus employing a utility-maximizing individual strategy based on looking at what one predicts others will do. Then, the difference is that new situations are possible to people who respond this way, or, more precisely, who have developed the disposition to respond this way. I think the move to

---

[25] Gauthier sees the conversion to constrained maximization as a choice of type of rationality, but I will bypass this topic, as my interest has been in the existence and maintenance of a social institution such as contract-making.

acting on a joint strategy is important, but, as things stand in Gauthier's analysis, that an action is performed in "doing one's part" of a joint strategy does not say anything about the agent being aware of, or responsive to, the agent's actions or expectations during the interaction. I want to extend the notion of role to incorporate roles in which expectations are formed and revised during the process of carrying out the joint action.

We will shortly turn to the question of the kind of cognitive structures the agents must have in order to show this kind of complexity in their interactions. But first I want to point out something Lewis cites as empirical fact in support of the plausibility of his claim that we can always assimilate different individual agents' psychology to one common sort of psychology: that we figure out what actions other agents will choose by "putting ourselves in the other fellow's shoes" (27). We shall see that the proposed account of the role of belief in sustaining social institutions differs in an illuminating way from Lewis' account about how beliefs about others are employed. Lewis says:

> We may achieve coordination by acting on our concordant expectations about each other's actions. And we may acquire those expectations, or correct or corroborate whatever expectations we already have, by putting ourselves in the other fellow's shoes, to the best of our ability. If I know what you believe about the matters of fact that determine the likely effects of your alternative actions, and if I know your preferences among possible outcomes and I know that you possess a modicum of practical rationality, then I can replicate your practical reasoning to figure out what you will probably do, so that I can act appropriately. (27)

This point is of fundamental importance to Lewis' appeal to how coordination is modelled between agents in developing his account of convention, for, based on it, he then goes on to reason that

replicating reasoning of others involves replicating the other person's replication of your own reasoning, and so on.

It is no doubt plausible that we can and do "replicate" the reasoning of others with whom we interact in order to predict their actions, for cases in which our roles and knowledge are similar to those of the person whose action we are trying to predict. But what about predicting actions of people in roles I am unfamiliar with, such as experts in fields in which I have no understanding? Recall that, although Lewis said that the (perhaps unnaturally complex) action-description would be the same for the agents participating in the convention regarding interrupted phone calls, he said (rightly) that they would have different roles. Now, I may often have expectations about the actions someone in a different role will take, without imagining the reasoning process that person uses. These expectations I have may have been formed by experience, but I may not have any illusions at all that I can replicate the reasoning the other person used.

Based on our past interactions, I may glumly expect that the vet will prescribe yet another X-ray, without having a clue as to the reasoning she goes through in deciding to prescribe X-rays. It is not just with experts, though, that we expect without replicating reasoning: it is likely to be true whenever we interact with others whose roles are very different from any we have taken on. So, for instance, I have no idea how waiters and waitresses do their jobs; their organizational abilities are often a source of wonder to me. Yet, I generally have little problem anticipating when and what to expect of them when I go to a restaurant. But, I do not do this in a "windowless monad" fashion, any more than I interact with a ball I am bouncing by asking myself where I would go next if I were a ball. As in bouncing a particular ball, I may revise and refine my schema for interactions with a particular waiter depending upon how our interactions go. The difference is not that I can put myself in the waiter's place and cannot put myself in the ball's, but that the waiter has psychological attributes that figure in our interactions, and the ball does not. I may ascertain that he expects me to keep my fork without having a clue why he expects me to.

Now, Lewis does not rule out that there are other ways we may have of obtaining expectations about each other's actions besides replication of reasoning, so my observations here do not contradict anything he says. The point is that he presents an account for conventional coordinated action, on which the people involved are able to replicate each other's reasoning. In contrast, the account of the more general case of social interaction using the view of beliefs proposed in this dissertation makes no such presumption. Rather, beliefs are efficacious in social interaction by being incorporated into people's anticipatory schemata; social stereotypes and social roles incorporated into one's anticipatory schemata allow one to anticipate what the person one is interacting with may do next, without requiring that we be able to "put ourselves in the other's shoes". On my account of the more general case of social interaction, we have to make up for the fact that we are not able to predict what others will do in a "windowless monad" fashion. We do this in part by developing anticipatory schemata that incorporate knowledge from past social interactions (as Lewis allows), but in the general case the anticipatory schemata we employ are dynamic and we employ and revise them in conjunction with actually paying attention to the living, thinking beings with whom we interact as we interact with them.

5. "Windowless Monads" and Joint Actions

I've been saying that the "windowless monad" picture of interaction featured in Lewis' account of convention is an obstacle to regarding his account of how beliefs are involved in the existence and maintenance of a convention as an account of the more general case of how beliefs are involved in the existence and maintenance of social institutions.

What we do want to retain from Lewis' account is the recognition of the significance of a network of mutual expectations. Recall that Lewis was concerned to distinguish a convention from a mere regularity in behavior, and what he said distinguished them was that the behavioral regularity of a

convention is caused in a certain way: by similar conditional preferences to conform, along with a system of mutual expectations coordinated by a state of affairs that ensures everyone has the appropriate first and higher-level expectations. The state of affairs -- such as an announcement --- ensures not only that people expect others to behave in conformance with the regularity, but that the others know that each one of them expects every other to behave in conformance with the regularity, and that all know this.

Thus, the kind of interaction involved in this account of convention is a kind of interaction in which expectations are static throughout the interaction. It would be unfair to attribute this to Lewis as an account of social interaction in general; the statements were made in the context of presenting his account of convention. We can at least say that, in describing a kind of interaction in which the agents are 'windowless monads", he has picked out a class of interactions which we can distinguish from interactions in which agents acquire and revise expectations of each other during the interaction. Thus the question is not whether Lewis is right or wrong, but, rather: what kinds of interactions and social institutions does Lewis' analysis apply to, and what kind of interactions and institutions require dynamic expectations during the interaction itself?

Lewis gives eleven examples of conventions. We have already discussed the convention of calling back to re-establish a cutoff phone conversation if and only if you are the one who originated the call. The act of conformance is described as "calling back if and only if you are the original caller", which does refer to agent roles. This case is not representative of action descriptions that refer to an agent's role, however, for the role is so simple in this case that it is determined by an action that can be described without reference to the others' actions in the interaction. One occupies the role simply by being the one who initiated the call, in the literal sense of being the one who established the telephone connection, and conforms to convention when in that role simply by re-establishing the connection when it is cut off by the telephone system. Here the interactions are interactions with telephone system hardware, whereas, in

general, a role will involve the kind of actions one does in response to, or in coordination with, the actions of another person.

Another example of convention Lewis cites is one David Hume used in both the Treatise and his later Enquiry to illustrate that conventions need not arise from promises: "Two men, who pull the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other" (Hume, Treatise 490). Hume is concerned to show that a general sense of common interest can account for much, including a respect for another's possessions:

> When this common sense of interest is mutually express'd, and is known to both, it produces a suitable resolution and behaviour. And this may properly enough be called a convention or agreement betwixt us, tho' without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are perform'd upon the supposition, that something is to be perform'd on the other part. (Treatise 490)

Our interest in the example is a bit different from Hume's; we are interested in how the expectations are involved in producing the joint action of successfully propelling the boat, rather than in the fact that common interest can give rise to regulation of one's conduct. Hume does not really go into much more detail than to say that the actions of each have a reference to those of the other, that they are based on a "common sense of interest that is known to both. . ." and that they are "perform'd upon the supposition, that something is to be perform'd on the other part." Hume's concern here is not with the metaphysics of joint action, and he does not address how one would go about analyzing or describing the actions the individual agents perform in effecting the joint action.

Lewis, on the other hand, discusses the example in detail at several points in Convention. First, he says that each of the two rowers "constantly adjusts his rate to match the rate he expects the other to maintain" (5-6). This description seems to allow that each rower could be continuously keying off the motions of the other, and that the rate (and perhaps direction, style, and force) of the strokes can be continuously changing. However, in later discussions of the example of two rowers, Lewis seems not to allow this. For, he describes doing one's part in the joint action of rowing in rhythm as rowing in a particular rhythm: "A regularity in their behavior --- their rowing in that particular rhythm --- persists because they expect it to be continued and they want to match their rhythms of rowing" (44). Lewis' remark in that discussion that "we could easily catch on to" the regularity, though we might find it hard to describe, further indicates that he is thinking of a certain rhythm of rowing that each recognizes and tries to follow, rather than the earlier description whereby the rowing rhythm could be continuously changing, requiring each rower to anticipate the changes in the stroke that his partner will be making next, and those that are expected of him. In a later discussion, in which he uses the example to make the point that we have and use knowledge we cannot express verbally, it is clear that Lewis regards the example of the rowers to be such that it is a particular rhythm that each rower expects of the other (rather than a continuously changing anticipation of a continuously changing stroke):

> If I am one of the rowers who row in a certain rhythm by a tacit and temporary convention, I have evidence that we have a convention to row in that rhythm. . . . I cannot say how we are rowing --- say, one stroke every 2.3 seconds --- but I can keep on rowing that way; I can tell whether you keep on rowing that way; later, I could probably demonstrate to somebody what rhythm it was; I would be surprised if you began to row differently; and so on. Now there is a description that can identify the way we are rowing. We take 1.4 $\pm$ .05 seconds for the stroke and .9 $\pm$ .1 for the return, exerting a peak force of 70 $\pm$ 10 pounds near the beginning of each stroke, moving the oars from 32 $\pm$ 6 forward to 29 $\pm$ 4 back, and

> so on, in as much detail as you please. But, as we row, we have no use for this sort of description. (63-64)

This understanding of how rowers coordinate their actions -- by picking up on the specific rhythm they are to effect --- shows up in the only other place Lewis discusses Hume's paradigm of a convention achieved without agreement or promise: Lewis says that the two rowers "can perfectly well agree to row thus, specifying a rhythm of rowing by demonstrating it" (86-87). This treatment of Hume's example of two people rowing is consistent with Lewis' claim that, in participating in a convention, we are all "windowless monads". That is, on Lewis's account, the two rowers do not need to interact with each other in order to perform their part of the joint action. The rowers' interactions are coordinated in terms of a specification of an intermediate entity: in this case, by exhibiting a specific way of rowing.

Likewise, we find in Gauthier a notion of acting on a joint strategy that might be taken to mean "doing one's part" in an enterprise in which the individual actions called for by the joint strategy can be delineated up front; I say this because Gauthier's concept of constrained maximization implies that the individual's "part" of a joint outcome he has promised to carry out can be executed without regard to whether or how others do theirs. Certainly this does make sense for some enterprises, but I think we also need a less reductive version for some others.

In contrast, on the account of how beliefs are efficacious in social interaction (and thus in establishing and maintaining social institutions) I am proposing, people's interactions are guided by anticipations and expectations that are constantly being revised and updated, and each person is perceiving anticipations and expectations of the person with whom they are interacting. Thus, on the account I am proposing, the rowers could be constantly changing the pace, direction, and style of their strokes. During social interaction, including cases of joint agency, one's actions result from being drawn into participation in an interdependent and interactive

process. What joint action requires, in addition to a common interest among the participants, is that, as Hume put it, each one's actions "have a reference to the other". If one of the rowers' actions indicates that he is trying to make the boat slow down or bear right, the other rower might revise his strokes as appropriate to achieve the joint outcome of the boat going more slowly or bearing right. Or, perhaps these actions of the first rower are taken by the second rower as slip-ups on the part of the first rower; the second rower might then instead compensate for the first's mistakes in order to achieve the joint outcome of staying on the course they had been going on before the aberrations. Sometimes the rower's duties and perogatives depend on which seat he occupies in the craft.

To illustrate the significance of interactiveness in joint action, let's examine a variety of cases, with an eye to asking whether the participants could determine the action expected of them without actually being drawn into an interactive process. (Where not obvious, the joint action is stated.)

(i) speaking your lines in a performance for which there is a script

(ii) performing an ad-lib skit in which you play a character type (e.g., playing the straight man in a comedy duo)

(iii) carrying on a conversation in which you are to be an empathetic listener

(iv) co-paddling a canoe in rough waters

(v) dancing a part in a dance in which some of the movements involve some of the dancers lifting, or leaning on, others.

(vi) playing in an orchestra

(vii) recognizing someone's fame (the group action here would be to confer the honor of recognition on someone).

Example (i), speaking your lines in a performance for which there is a script, is the most literal interpretation of a role. This kind of role, in which what the role requires of one is merely speaking

scripted lines in a play, is at one end of a spectrum; at the other end is a kind of role in which what one does next depends not only upon what those with whom one is interacting has done, but also upon one's anticipations of what they are to do next, and, inasmuch as you can discern them, the anticipations others have of your next actions. (Thus, what you do depends on your anticipations of their anticipations of your next action.)

Performances in which each person's lines are scripted in advance can perhaps admit of a reduction of the joint action to individual actions: one can delineate what is required of one and do one's part all by himself (say his lines) although one cannot perform the joint action (putting on the performance) by oneself. This way of looking at a performance of a troupe was illustrated by an avant-garde dance piece performed in the following way: members of the audience, chosen randomly, were each given a tape player, with headphones, containing a tape of individualized instructions. They were lined up, each in his respective starting position, and were instructed to start the tape of instructions simultaneously. A coordinated dance performance followed, stimulating thoughts of The Monadology. I do not know what, if any, point the authors of the piece wished to illustrate, but the point I wish to make is that this dance differed from normal social interaction. This dance was not an instance wherein one is forced to bring in a notion of joint agency; coordinated individual agency is sufficient.

How the notion of "role" should be extended to cover the full spectrum of cases encountered in social interaction is better illustrated by examples (ii) and (iii), the roles taken on by performers of an ad-lib skit, and the role of being the empathetic listener in a conversation. In these examples, the role is a matter of the relation between people's actions: the actions (what's said) themselves may not be specified by the role, but the kind of reaction or response is. And the response is really a response --- it is a response to some specific action by another person. For instance, the person playing the straight man of a comedy duo in an improvised skit has to recognize when his partner has made a joke, and come up with some cleverly chosen remark that shows that he has

not gotten the joke, that he has misinterpreted the statements just made in some way, either through his ignorance, lack of a sense of humor, or some understandable reading of an ambiguous phrase. The appropriateness of an action one takes to "do one's part" in the joint action depends upon the respective roles in a way that depends on the other person's action. When playing the role of an empathetic listener in a conversation, one needs to show in some way, without taking over the conversation, that one recognizes the other person's predicament; this requires responses that key off what one discerns to be the matter of concern to the other person. Example (iv), co-paddling a canoe in rough waters, which is similar to the role of the rowers discussed above in the discussion of rowing, also illustrates this. There, too, the rower had to determine what his role requires of him based on his knowledge of how his actions are to complement and aid his partners to achieve the joint action; their conceptions of their respective roles are what helps them coordinate their actions. The responses are specific actions, although they are not determined by a script beforehand, but by what is appropriate for someone in that position of the craft to do. (This notion of role is not meant to rule out cases where the role is no more specific than being another human being.)

On this extended notion of role, the relation between the actions each person takes in response to the other is a matter of the relation of their respective roles, and how the actions they take are meant to achieve the aimed-for joint outcome. This notion of role differs from the dance performed via taped instructions in that the volunteer drafted from the audience was performing actions the tape instructed him to do in his role as a person whose role was to carry out the actions assigned to whoever was given that tape player; he was not performing actions determined according to what a role required of him. I distinguish this case from acting according to what a role requires because acting according to a role would be recognizing relationships between himself and the others, and recognizing the sort of behavior that would be appropriate in response to the sort of action the other has performed. Even if the participants are given "roles," say by named or numbered parts, if they are not acting from this sort of recognition, they are not acting according to

what a role requires, in the sense of "role" I want to capture here. And, what I'll want to show is that, for joint actions in which this sense of "role" is involved, it does not always make sense to say that one could perform one's part of the joint action outside of a social interaction with another person.

Example (v), dancing a part in a dance in which some of the movements involve some of the dancers lifting, or leaning on, others, could fit or not, depending on how it is looked at. If we look at dance the way the avant-garde dance troupe caricatured it, we could see a dance performance as the result of individuals acting out their parts, all coordinated by timing, say, to the orchestra's score. This would mean we could describe what was going on without resorting to speaking of joint agency. But it gets tricky when we think of movements that require lifting, or leaning on, one another. We could still stubbornly stick to saying that each individual's movement is prescribed, and all the actions are coordinated, but we are getting near the borderline here: although it may be possible to describe the actions involving leaning and lifting others in a "windowless monad" fashion, it seems forced to do so. Rather, it seems much more natural to describe the dance performance in terms of the joint outcome they have achieved, and the roles each performed in achieving it. One can still judge the performance and the way each person performed his or her role separately when it is described in this way.

Playing in an orchestra (example (vi)) is similar to example (v). It may sound plausible to describe an orchestral performance as coordinated individual actions, each done by individuals who base their actions on a joint strategy. It seems especially reasonable to do so since the score delineates each person's part, and there is a conductor coordinating individual action. However, I believe this account of orchestra performances, often used in philosophy as an example of a joint action, is a philosopher's legend: actually, playing in an orchestra involves playing <u>with</u> others in that one must listen to how the others are playing and accommodate the rest as required for harmonizing. I've also been told that singing in a choral group is similar. In fact, singers with

perfect pitch pose a problem: since they tend to sing their part exactly as written, i.e., sing the exact note written for their part, they are often fighting the rest of the group. Orchestras and choral groups are borderline instances of reducing a group action to coordinated individual action, because one could argue that, if humans and instruments were perfectly precise, orchestral and choral performances could be performed solely by each participant sticking to his part, coordinated by a score and/or conductor. But suppose it is granted that we are talking about real musicians and vocalists. Then, if one wants to stick to an account in terms of coordinated individual action, one must add a lot of detail to the description of the individual parts, and these will have to include descriptions of whose actions need to be kept track of in order to figure out what to do, and of how one should respond (this may differ depending upon one's position in the orchestra or chorus). Then the descriptions we come up with tend toward describing the joint action in terms of roles in the expanded sense of "role": the actions taken by those in roles in terms of which participation in a joint action or strategy is defined are a matter of the kind of response that is appropriate to particular actions taken by those in other roles, in various circumstances. If we start to think about joint actions that are not nearly so circumscribed as playing a specific piece of music, we can see how roles (in this expanded sense), rather than roles characterized in terms of the actions to be performed, are really a much more natural way to look at the relation between what a participant "doing his or her part" does, and the joint outcome. We are not faced with a puzzle about how to describe the action associated with the notes coming out of a singer's mouth: we get an action characterized in terms of how her singing relates to the rest of the choral group, rather than a "true pitch" characterization of the notes as they would be recorded on a musical score. The empathetic listener's responses are characterized as they relate to understanding the speaker (i.e., validation, confrontation, supportiveness) in contrast to the sorts of characterizations one would give of the responses of a straight man in a comedy duo whose role is to further the ad-lib skit along by pretending not to recognize the humor in his partner's remarks (e.g, taking a sarcastic comment literally, or responding to a pun by repeating himself on the pretense that he thinks the other person has not understood him the first time).

This account of how actions are characterized in terms of social role is related to the question of how beliefs are involved in social institutions, inasmuch as belief is defined in terms of its role in determining one's own actions, as well as in discerning other's actions.

Thinking of example (vii), recognizing someone as a famous person, as a joint action, highlights the subtleties involved in these different ways of analyzing joint actions.

Perhaps I should first say a few words about why it is appropriate to think of achieving fame as a joint action. To begin with, notice that fame has to be more than just a matter of being someone whom a lot of people happen to know or know of, for then there would be no difference in kind between recognizing someone and recognizing someone as famous. And there is a difference, because recognizing someone as famous includes at least the additional feature of an awareness that the person is well-known. The next step that naturally suggests itself is to give an account in terms of Lewis' "common knowledge" provision. If we try to give an account of what a person's fame depends on analogous to Lewis' reductive account of a convention, the common knowledge provision amounts to something like this: There is a state of affairs A such that (i) everyone has reason to believe that A holds, (ii) A indicates to everyone that everyone else has reason to believe A , and (iii) A indicates that X is famous.[26] An example of the state of affairs A would be the publication in a major newspaper of an article referring to "the famous X." This "common knowledge" feature does get us closer to what is involved in being famous, compared to the characterization of fame as merely recognition by a lot of people, because it includes in the notion of recognizing someone as famous that others believe so, too. However, both a publicity

---

[26] To say that this is the account Lewis would give would be putting words in his mouth. I mean merely to say what an account of fame analogous to his account of convention would yield, with respect to the question of the way in which beliefs are involved in the maintenance and/or existence of someone's being famous.

hoax and the media attention that genuinely reflects public recognition can satisfy the formal constraints of such a common knowledge requirement.

Certainly there are publicity hoaxes in which it seems that fame can be manufactured merely by creating some state of affairs that meets formal requirements like those given on a Lewis-style account of fame. But, when someone says that fame can be manufactured, part of the point is that the fame that can be achieved this way is somehow less genuine than appropriately earned fame, that it is a counterfeit of the kind of fame that is achieved by doing something that is remarked upon because people actually find it remarkable, talked about because people find it of interest, and influential due to some feature other than merely having become well-known. Even if it is granted that there is no correlation between how meritorious someone's achievements are, and how famous they become, we can still recognize a difference between publicity stunts like Carl Laemmle's [27] and one where no stunts are involved, where the media has not acted inappropriately nor been deceived. And, in sticky cases where the fame does seem too much "generated", say, by newsmagazines and newsshows being manipulated by publicity seekers, the notion of fame seems to be a degenerate version of a stronger one.

In Warhol's frequently-quoted statement that "In the future, everyone will be famous for 15 minutes", there is something tongue-in-cheek about the notion of fame in play, and an impishness about the claim. Even taken soberly as a statement about the future, it's a comment on how little it will then take, but also how little it will then mean, to achieve fame. The point is that

---

27 "The star system was born in 1910 when producer Carl Laemmle planted a false story in the St. Louis Dispatch saying that actress Florence Lawrence had lost her life in a trolley car accident. The next day Laemmle was loudly proclaiming in indignant advertisements that the story had been a lie, and that Florence was very much alive and soon to appear in his latest film, The Broken Oath. . . . All this publicity naturally generated a great deal of interest in Miss Lawrence. . . Before Laemmle's publicity coup, the public did not know film actors and actresses by name" (Learning, If This Was Happiness: A Biography of Rita Hayworth 44).

the "common knowledge" characterization gives a formal characterization that a degenerate notion of fame meets as well. And this is likely to be a weakness not only of Lewis-style accounts where it is "common knowledge" that does the work of coordinating individual actions, but of any account where each can do his part (the individual act of recognizing someone as famous) of the joint action (someone becoming a famous person) alone. On an account of fame as just the sum of coordinated individual actions, all that's needed to simulate fame is to somehow coordinate all the individual actions. Probably any such simulation will result in a degenerate notion of fame. My suggestion, given below, should not allow such a trick; any attempt at simulating fame will either fail or will suceed in bringing it about that someone becomes famous in a robust sense.

Employing the suggestion of thinking of social institutions as joint actions in which individual actions are determined by roles, the relevant roles involved in someone's being famous will include the role of the fan/supporter/detractor, and the role of media professionals, at least. The role of the famous person is also involved (although the person in that role may not be able to do much to change the state of affairs of his or her fame), as is the general role of just being a member of a culture or subculture, regardless of whether one has any particularly strong opinions or feelings about the famous person. The specific roles are not important for my point here, since what I am interested in is examining how what I have in mind differs from the "common knowledge" characterization. The preceding is just an example of how one might see the institution of fame as a joint action in which individual participation is involved via interrelated roles that determine the kinds of responses that are appropriate.

To examine the difference, consider what the appropriate actions and responses associated with these roles are. For the general role of being a member of a culture (or subculture), keeping informed of developments in your culture is part of your role. Thus, some people consider knowing what is on the front page of the New York Times before 11 am of the day it's printed part of what it is to be a member of their culture, even if they disagree with the statements made there,

or, even, with the choice of subjects as first-page news. There is the role of media professional, for whom this role means reporting whatever is newsworthy on the "beat" to which he is assigned, even if he feels it means giving media attention to trends he would personally prefer not to encourage. This differs from the role of public relations professional, whose role involves trying to get positive media coverage, avoid negative coverage, and, usually but not always, increase media coverage for his client. The institution of fame requires that these roles be distinct; inasmuch as they are blurred, we feel some sort of inauthenticity has contaminated the resulting fame. The roles of famous person and fan/supporter/detractor are part of the institution as well. The famous person may give people her photograph or signature as if she is doing something gracious, which would be considered at least bizarre if she were not famous. When she wrote to her older sister at college, k. d. lang, who hadn't even begun a performing career, told her sister to keep the piece of paper with her (k.d.'s) signature on it, as it would be valuable someday. The act was a way of saying that she planned on becoming a star, and would have made no sense otherwise, i.e., in the absence of an institution in which there is the practice of asking for autographs.

The difference between the account I'm trying to sketch here and the one in which each can do his part alone is rather subtle, and turns on how we understand the notion of role. For instance, consider the role of the media professional. If, instead of using such complexly interrelated roles, we delineate the media professional's role more simply as his part in making someone famous, say, by performing the task of reporting on that person, and if we delineate other roles similarly, we can see that , on such an account, we could simulate fame by having each do "his part" in making that person famous. But this sort of "playing one's part" is not the same as role-mediated actions on the expanded concept of role I've sketched here. The news media professional's role is to report on things that are genuinely newsworthy, according to some conception of newsworthines, which it is not essential to spell out here. This role will even vary with a particular professional, and might depend on things such as the genre for which he is writing and the beat

he's charged with covering. For example, if he is in charge of the entertainment section of a major newspaper, and a performer is giving shows that audiences find remarkable, it's part of his role to report on that performer. It would not be appropriate to use the space to write about another performer out of loyalty to a blood relative who is trying to promote the other performer. And it is part of his role to ensure that the sources he is counting on to inform him of developments and trends are also performing their roles properly and not to treat them as sources if they are not. For the example of the institution of fame, I am saying that these particular features of the role are important. However, there could be a role of the media professional in a culture or subculture in which the institution of fame is degenerate. The philosophical point does not depend on the claim that the role has these particular features. The point is that, to perform his role, one has to be alert to a myriad of things going on, to know which are relevant to his role and what the appropriate response to them is, to be able to critically evaluate if others are performing their roles properly, and to be aware of how others could be using him to exploit his role. This characterization of the role of media professional is in contrast to a characterization of a person's role obtained by breaking up the joint action of recognizing someone as famous into different "parts" that each person can perform on his own, such as a characterization of media professional as simply the person who writes the entertainment section of the paper, and who is to relate to certain other designated role-holders as sources.

The existence of an institution or of a subculture cognizant of, say, a certain art form, or a certain sport, depends on expectations people in the subculture or who participate in the institution have of each other: the conversations, publications, performances, etc., constitutive of the subculture can only happen if there are people to converse with, and audiences to address, who have kept abreast of the same things. The notion of expectation here is akin to that of anticipation or presumption: I begin my conversation with someone in that subculture using references I count on them getting; my artwork includes visual quotations from other works that I presume my audience will have seen, or at least will be able to make some associations with. But it is

interactive, like a conversation: I judge from their reaction what to say or do next, maybe even picking up something of the reaction in my next piece. It is through the presumptions about and anticipations of others' actions and reactions that beliefs are involved in maintaining the institution or subculture. They are interlaced with our own actions; they aid in perceivng other's actions and they guide our own.

The kind of difference is like that mentioned earlier in the discussion of the example of dancers who have to lean on, or support, each other: if the role is characterized in terms of the individual actions each dancer has to perform, we find that the characterization would have to be extremely detailed, and, more importantly, would have to include very complicated references to the other dancers' actions. As we worked it out, we would find that the intricate interrelatedness of the roles would show up in the descriptions of the individual actions called for by a role. In effect, to do it right would mean we no longer really had roles characterized by individual actions that could be done "by oneself", in that one could not carry out the action, or often, determine the action required by the role, unless others did their "parts," too. A more natural way of characterizing the role would be in terms of kinds of expectations the role allows each to develop as to what the others' actions indicate, and about what is expected of them next: the role is characterized by the kind of responses that are appropriate with respect to other role-holders' actions. To make the point about this contrast, think of the difficulties of spelling out a role in terms of actions for the example of the role of a straight man in an ad-lib skit.

It would not be impossible for a straight man to continue a skit if his partner were suddenly seized off the stage, just as it sometimes happens that a ballet dancer has to continue alone when his or her partner has become injured and leaves the stage during a duet. However, in these cases, the performer is really doing something different from performing as a solo the part he would perform were he still part of a duo or duet; he or she is performing actions in response to implied actions of the other. The action is not seen as a coordinated solo action, but as an "as-if" interactive one.

To make sense of what the abandoned performer is doing, the audience fills in the implied absent participant's action (e.g., holding out her hand, saying the thing that the straight man pretends to be repeating with a question mark at the end) .

The characterization of social interaction involved in the existence and maintenance of social institutions taking shape in this section extends the progressive step we identifed in Gauthier's work that allowed for discrimination between various agents when responding to them. Recall that, in Gauthier's reconstruction of bargaining situations, the appropriate response of a constrained maximizer to someone he regarded as another constrained maximizer differed from the response that was appropriate to make to someone whom he regarded as an unconstrained maximizer. This is a progressive step over a characterization of a bargainer's role on which one in that role is to respond to everyone making the same offer in the same way. The account of role-mediated interaction I've suggested above picks up on this, in contrast to some accounts of social institutions (such as Lewis' account of social conventions) that analyze individual actions in such a way that the appropriate response, though role-mediated, does not differ according to the person holding the role. Thus, we see that a proper account of the beliefs involved in the actions that are responsible for social interactions should include the resources to discern the characters or other psychological features of individuals that enable a person to make discriminations between various role-holders in the course of his social activity.

6. Roles, Stereotypes, and Beliefs

The account of belief developed in the first part of this chapter is thus well-suited to account for how beliefs are responsible for the existence and maintenance of social institutions. Beliefs are causal by being incorporated into anticipatory schemata that are employed as people interact with each other. The kinds of cognitive structures that are incorporated into dynamic anticipatory schemata that are employed in social interaction include roles and stereotypes. Agents conceive

of social situations in terms of roles, and they perceive and interact with other individuals by employing social stereotypes.

I have argued that people structure situations in order to coordinate with each other by using social roles they have developed from past interactions; these are, however, constantly revised, combined and synthesized as one picks up more information about the particulars of the situation in which one finds oneself. I have pointed out that even for the special case of convention, Lewis shows that roles may be involved in solving coordination problems, and have tried to show why I think the point is nascent in Gauthier's explanation of how salience works in coordination problems as well. The notion of social stereotypes (in the broad sense) is a natural extension of stereotypes for inanimate objects; the extension takes into account that people have psychological features, such as expectations and characters, as well as physical ones. It is especially natural to think of social roles and stereotypes in terms of anticipatory schemata: they are used to perceive a social situation by directing the pickup of information, they give rise to involuntary responses[28], and they are used to guide one's actions. As we have already

---

28 How involuntary responses should be distinguished from voluntary ones is a topic deserving of a separate investigation. I mean here to indicate there is a realm of response that falls outside intentional action. There is the issue of whether or not we can gain control over involuntary responses, whether simple ones such as blinking, more complicated physiological ones such as weeping or facial expressions, or cognitive ones such as the stereotypes we use in sizing up and reacting to a situation. I think it is uncontroversial that we do have the ability to gain control over many of our involuntary responses; some might argue that we could gain control over any given one identified, including automatically controlled biological processes, such as heartrate and body temperature. Then there is the other extreme philosophical view, which emphasizes the skill involved in producing many intentional actions, thus developing the proper habitual responses is important in order to do anything voluntarily; here I am thinking of John Dewey's essay "Habits and Will" in Human Nature and Conduct : The Middle Works of John Dewey 1899-1924, Vol. 14. Ed. by Jo Ann Boydston, Carbondale: Southern Illinois University Press, 1983.

established, the ways beliefs are causally efficacious in the employment of anticipatory schemata in general, and in social stereotypes in particular, are not restricted to intentional actions.

The answer to the question of whether the role of beliefs in the existence and maintenance of social institutions involves beliefs in a mode of causation other than via intentional action is thus that it is indeed unnecessarily limiting to treat the efficacy of belief in the establishment and maintenance of social institutions as only possible through intentional action. However, this is not because there is a new kind of agent or agency involved; beliefs are causal in the existence and maintenance of social institutions in the ways in which dynamic anticipatory schemata are, and these include a variety of physiological, psychological and social phenomena. The difference between beliefs being causal in the existence and maintenance of social institutions and beliefs causing physiological and psychological phenomena is that they are incorporated into social roles and social stereotypes, and are employed in social interaction. This is just to say that the account of how beliefs are causal sketched in this dissertation in Chapter III accounts for the kind of causality by which social institutions are created, maintained and revised, as well as for intentional action and nonintentional physiological responses.

CHAPTER V

BELIEFS IN ECONOMIC INTERACTION

A. Expectations and Money

In Chapter IV, I argued that the notion of an anticipatory schema for social interaction was well suited to accounting for the kind of cognitive structures that enable people to create and sustain social institutions. That economies show qualitatively different responses to policy changes depending upon the anticipations of the agents in them augments that view by showing the importance of anticipations when social institutions undergo transitions. The so-called "rational expectations" approach that has recently gained prominence in economics explicitly recognizes the pivotal role of anticipations in implementing institutional change. I'll explain why I think that the approach, though often used along with the (unnecessary) presumption that agents' expectations are efficacious via an agent's "decision rules", makes room for the account I've proposed in this dissertation, on which beliefs and expectations are also efficacious in modes other than rational intentional action.

In addition, we shall see that there is a connection between David Lewis' explication of Hume's suggestion that words get to mean what they do because of the kind of mutual expectations people have developed of each other, and the economist Robert Lucas' explication of Hume's suggestion that a token of currency has the value it does because of the kind of mutual expectations people have developed of each other's future behavior. These mutual expectations, in Hume's words, are expectations of "a confidence of the regularity of [all one's fellows'] future conduct" (Treatise 490) founded on a common interest.

My main goal in this dissertation has been to answer the question of how beliefs are efficacious, while remaining neutral on the question of what a belief is. However, the analogy between Lewis's account of convention and Lucas' notion of rational expectations suggests how to extend the concepts of rationality and belief in such a way that the notion of belief can include anticipations that are efficacious, but not necessarily via being part of a reason someone can give. Later in the chapter, I shall briefly outline the analogy, and point out the significance I see in the extended concept of belief it suggests: that one can preserve the valuable insight behind the claim that beliefs can only be attributed to a being for behavior that can be seen to make sense, without being forced to the view that beliefs can be attributed only on the basis of behavior that can be seen to make sense in terms of the reasons a being can give for acting or believing. Sometimes we can make sense of a person's behavior in terms of common interests and shared anticipations, instead of in terms of their reasons. The really fundamental feature of so-called "rational expectations" supporting this insight is a constraint on expectations: that they be model-consistent expectations.

Informally put, what it means to say that the economic agents interacting in an economic system have model-consistent expectations (so-called "rational expectations") is that the econometrician trying to predict their behavior under a certain postulated institutional change attributes to them (collectively, in a way that can be made mathematically precise) the same expectations that he himself would have of the effect of various changes in the world, were he to possess the information he attributes to them. That is, he attributes to the agents as a group the anticipations of future events or conditions they will experience in the model of the environment in which he is placing them, conditional on the information they would have at that point in time, were they in the position in which he has placed them. Lucas makes the point that the hypothesis of rational expectations is not a hypothesis about how people behave, but a constraint on the construction

of models of economies used to predict their behavior; I don't know that any philosopher has appreciated the point. [29]

1. Expectations and Causality

When David Lewis applied game theoretic methods[30] to answer the question of whether language was conventional, he found that he had vindicated Hume. When the economist Robert Lucas applied modern mathematical methods to answer the question of whether money was neutral (i.e., changes in the amount of money in circulation do not result in more wealth, nor in any effects at all, other than proportional changes in prices), he too found that he had vindicated Hume.[31] Both gave explanations that cited, not only a mere regularity in conduct, but what Hume

---

[29] Martin Hollis, for instance, in the chapter entitled "Rational Expectations" in The Cunning of Reason, writes "Hence agents with fully rational expectations are agents guided by 'the relevant economic theory'. But, unfortunately, this is not only clever but also ambiguous, depending on how one regards the relevant economic theory" (104). I believe Hollis' evaluation of rational expectations would have gone differently had he appreciated Lucas' warning, such as this one, in Models of Business Cycles: "The term 'rational expectations' as Muth used it, refers to a consistency axiom for economic models, so it can be given precise meaning only in the context of specific models. I think this is why attempts to define rational expectations in a model-free way tend to come out either vacuous ('People do the best they can with the information they have') or silly ('People know the true structure of the world they live in')" (13). Hollis does not cite Lucas anywhere in the book.

[30] Lewis said that game theory was "scaffolding" and that he could state his view without it. (4)

[31] There may also be a deeper structural connection in that the existence of Nash equilibria in game theory and the existence of rational expectations equilibria are proven by the use of "fixed point" theorems. A fixed point theorem is a theorem about the conditions under which there is a solution to the equation $f(\underline{x})=\underline{x}$, where f is a mapping with some nice properties, and $\underline{x}$ is some sort of object mapped. Lewis' account of convention is presented in terms of Nash equilibria (sets of strategies) to coordination problems, the existence of which Nash proved using the (Brouwer) fixed point theorem (applied to mappings of pairs of (mixed) game-theoretic strategies); Lucas applied the (Banach) fixed point theorem to mappings of price functions to show the existence of

referred to as "a confidence of the regularity" arising out of a shared interest. Lucas recently observed: "The main finding that emerged from the research of the 1970s is that anticipated changes in money growth have very different effects from unanticipated changes" ("Monetary Neutrality" 262). That unanticipated changes in policy have very different effects than anticipated ones do was not a premise, but a consequence, of the employment of so-called "rational expectations"[32] in economic modelling. Lucas notes "The importance of this distinction between anticipated and unanticipated monetary changes is an implication of every one of the many different models, all using rational expectations, that were developed during the 1970s to account for short-term trade-offs" (262). Whatever the ultimate verdict on the rational expectations approach may turn out to be, I take the significance of anticipations in differentially causing qualitatively different kinds of economic phenomena to be established.

It is a fairly common idea that the social sciences differ from the physical sciences in some fundamental way that has to do with the fact that the way things go in the realm being studied depends upon people's expectations of how things will go; a common way of putting it is that, in the social sciences, unlike in the physical sciences, "expectations are causal".[33] The dynamic nature of interaction often seems to make the social theorist's problem of predicting the effects of people's expectations intractable. The problem is not simply a matter of the large number of interactions economic systems typically involve, but of the dynamic nature of interaction itself: if A

---

equilibrium price functions in proving that an economic model he built exhibited the neutrality of money. (Appendix , "Expectations and the Neutrality of Money")

[32] The use of scare quotes indicates that what are often called rational expectations are more accurately described as "model consistent" expectations. An especially clear explanation of the difference between the terms, and when they do and do not coincide, can be found in Whitley (154).

[33] This particular phrase is used, with emphasis, by John T. Harvey (181); Paul Davidson has promoted the view, as well (182). Martin Hollis wants to urge that "expectation is a creative force quite foreign to the natural world"; he implies that the point is inconsistent with the rational expectations approach (Cunning 98). I discuss Hollis' views in more detail later in this chapter.

notices that B is expecting her to do something, she may respond by not doing it to avoid being taken advantage of; B may pick up on the change in A's plan and revise her expectations accordingly . . . and so on, ad infinitum. It has been shown, however, that this problem is not always insurmountable.

In the latter half of the twentieth century, economists have been especially tenacious in formulating and answering questions about how the expectations of people in some sort of social structure bring about changes to the very parameters those expectations are about. A common example here is the effect of price expectations on the prices of goods produced by members of an economy: if suppliers of a certain good expect the price of that good to rise, they tend to increase production of it; the increase in supply of that good may in turn cause its price to fall. A similar example is the effect of expectations of election results on those election results: if enough people come to believe that a certain highly desirable candidate has no chance of winning the election, enough of them may switch their votes to a less desirable candidate who does have a chance of winning the election over a third (utterly despised) candidate so that in fact they make it true that the highly desirable candidate does not win. The former is an example of a self-defeating prophecy; the latter, of a self-fulfilling prophecy.

The methodology that has come to be known as "rational expectations" was developed in part as a way to make prediction of the effects of policy changes tractable[34], without compromising the recognition of the dependence of those effects on the dynamics of people's expectations. The change was not a matter of fine-tuning an approximate result, but of offering an alternative to modelling methodologies that did not recognize the dependence of the results on the expectations of the agents in them and, as a result, could be wrong in arbitrary ways of arbitrarily

---

[34] Taken as a statement of Lucas' view, this is mild; for a strong mission statement as to why the models used for policy evaluation prior to the advent of rational expectations are useless for that purpose, see Lucas' "Econometric Policy Evaluation: A Critique".

large magnitudes. I have already mentioned that a key factor in determining the effects of a policy being contemplated turns out to be whether the change in policy is anticipated or unanticipated---perhaps it is worth emphasizing that this is a feature of the expectations people have, rather than of the content of the policy. Thus, in a very definite sense, whether or not, say, inflation, results from the implementation of a particular policy may turn on the expectations people have. If they expect the authority to be able to effectively implement the policy it announces and not waver in its resolve, the result will be very different than if they do not know what they can expect of the authority, or expect the authority to be ineffective in implementation of the policy, or expect it to relent. Another case is that of agents' behavior changing in anticipation of the implementation of a policy that is never actually implemented; anticipation of a liquor tax may cause an unseasonable increase in liquor purchases, whether or not such a policy is ever implemented. It is ironic that a number of writers associate the methodology of rational expectations with the view that expectations are not causal (e.g., Harvey 181), for the rational expectations approach arose from endeavoring to take such effects into account.

In The Cunning of Reason, the philosopher Martin Hollis evaluates various attempts economic theorists have made to develop methodologies that incorporate agents' expectations, including rational expectations, and concludes that the rational expectations approach is just one more failed attempt at modelling collective social behavior. He argues that the employment of rational expectations yields models that fail to account for the social phenomena, and so that rational expectations is yet another illustration of the fact that the social sciences have wrongly presumed that "the social world sets no radical further problem for the analysis of belief." I am not sure how much of his argument would still stand had he maintained a clear-cut distinction between rational expectations as a methodology of model-consistent expectations and the methodology of attributing "perfect information" (108) to agents. The target of Hollis' criticism is a view of social science modelled on the natural sciences, in which "it remains crucial that believing something so does not make it so"; in which the social scientist adopts "the ideal of the scientist as observer,

whose theories of probability and evidence are detached aids to prediction" (97-98). Hollis wants to urge instead that "expectation is a creative force quite foreign to the natural world" (98). I disagree with Hollis' conclusions about the rational expectations approach, but I am in sympathy with his goal.[35] Much of what I have said about beliefs so far, especially about beliefs in social interaction, could count as urging that "expectation is a creative force", as Hollis does. However, I would not want to say, with Hollis, that expectation, or belief, is "foreign to the natural world." On the contrary, I have been trying to make a single picture of the world in which we can place studies in social psychology and cognitive psychology along with psychophysical researches, in order to show that expectation and belief are part of the natural world --- or, rather, that a proper view of the natural world includes expectations and beliefs. In contrast to Hollis, I think that the methodology of rational expectations, as developed by Lucas, actually does permit incorporating agents' expectations in a way that recognizes that they can be "a creative force." To counteract the kind of dismissive attitude that regards the methodology of rational expectations as a simple-minded philosophical error or an unwarranted simplifying assumption about people's behavior[36], it will be helpful to review the history of the concept.

---

35 I find Hollis' exposition of adaptive and rational expectations on pages 98 -112 in The Cunning of Reason very clear, until it reaches a certain point. That point is on page 107ff, where the discussion takes a turn at "Observers predict; agents decide." The discussion that follows investigates what a hypothesis of "perfect information" would entail, and dismisses rational expectations on the basis of those consequences. Hollis, incidentally, also rejects Lewis' account of the conventionality of language in terms of game theoretic concepts (32).

36 It is not clear to me how these misconceptions got started. I suspect that some negative misconceptions of rational expectations are due to disapproval of the political ends to which the methodology was put.

2. The Development of the Concept of Rational Expectations

a. Tokens of Meaning and Tokens of Money

Just as the roots of present-day philosophical explanations of convention in terms of mutual expectations can be found in Hume's writings on human nature ("the actions of each of us have a reference to those of the other, and are perform'd upon the supposition, that something is to be perform'd on the other part" (Treatise 490)) , so the roots of present-day explanations of the efficacy of expectations on prices in terms of mutual expectations can be found in Hume's writings on economics. In fact, in his Nobel prize acceptance lecture, Robert Lucas, probably the best-known proponent of the methodology of rational expectations, couched his explanation of the concept as a resolution of two incompatible ideas found in Hume's essays ("Monetary Neutrality" 246). Lucas pointed out "the double standard that characterized Hume's argument" (252) about the effects of sudden increases and decreases in the quantity of money in an economy. Yet, he credited Hume with uncovering a problem that would remain unresolved for over two centuries: "The discovery of the central role of the distinction between anticipated and unanticipated money shocks resulted from the attempts, on the part of many researchers, to formulate mathematically explicit models that were capable of addressing the issues raised by Hume" (262).

The two incompatible ideas Lucas was referring to are found in Hume's essays "Of Money" and "Of Interest". There, Hume is concerned to explicate as a "principle of reason" what is now often called the neutrality of money; in Hume's words: "the quantity of gold and silver is itself altogether indifferent" to the well-being of a nation (Money 289). Hume argues[37] that a nation would be no

---

[37] Lucas describes Hume as arguing from a quantity theory of money to the neutrality of money. But since Hume begins his essay "Of Money" with the statement that "Money is not . . . one of the subjects of commerce; but only the instrument which men have agreed upon to facilitate the exchange of one commodity for another. It is none of the wheels of trade" (281), it seems to me that Hume is taking the neutrality of money to be a principle. It seems to me that Hume states the

better off were it to have more gold and silver; the only difference arising from an increase in a nation's money supply would be proportionally higher prices. For, he argued, "money is nothing but the representation of labour and commodities, and serves only as a method of rating or estimating them" (285). Should there be an increase in the quantity of a nation's holdings of gold and silver, the only effect would be that "a greater quantity of it is required to represent the same quantity of goods; it can have no effect, either good or bad . . . " (285). He compared the difference between needing more rather than fewer coins to the difference between using Roman numerals and Arabic ones: ". . . the greater quanitity of money, like the Roman characters, is rather inconvenient" (285). However, Hume also ascribed the increase in industriousness in Europe to the influx of gold and silver occasioned by the discovery of mines in America. He recognized the problem this empirical phenomenon created for his principle about the neutrality of money, and took on the challenge of explaining the apparent contradiction.

Hume stuck by his principle that, if one considers only the influence of "a greater abundance of coin", the only result is a proportional increase in prices ("obliging every one to pay a greater number of these little yellow or white pieces for every thing he purchases" (286)). Thus, looking only at the consequences of a postulated instantaneous increase in money, one cannot account for the beneficial effects of an increase in a nation's holding of money. Instead of revising the principle contradicted by this empirical phenomenon, Hume explained the beneficial effects of an increase of money as the temporary effects that arise from the fact that it takes some finite amount of time for the added amount of money to be distributed from "the coffers of a few persons" to the wider population, and so takes some finite amount of time for people to perceive the increase in gold and silver:

---

neutrality of money in such a way that it is explicit that his view that money is neutral is analogous to his view that language is conventional. For an explanation of various interrelated hypotheses concerning the quantity theory of money, and a history of debates about it, see Thomas M. Humphrey's "The Quantity Theory of Money: Its Historical Evolution and Role in Policy Debates".

> To account, then, for this phenomenon, we must consider, that though the high price of commodities be a necessary consequence of the encrease of gold and silver, yet it follows not immediately upon that encrease; but some time is required before the money circulates through the whole state, and makes its effect be felt on all ranks of people. At first, no alteration is perceived; by degrees the price rises, first of one commodity, then of another; till the whole at last reaches a just proportion with the new quantity of specie which is in the kingdom. In my opinion, it is only in this interval or intermediate situation, between the acquisition of money and rise of prices, that the encreasing quantity of gold and silver is favourable to industry. (Money 286)

Lucas finds Hume's reasoning inconsistent: if Hume deduced the principle that, in the long run, changes in money are just changes in units as a principle of reason, then shouldn't he explain why that "principle of reason" doesn't apply in the short run as well? Lucas complains: "[c]onsistency surely requires at least an attempt to apply these same principles to the analysis of the initial effects of a monetary expansion or contraction," but blames Hume's inconsistent approach on the intractablity of the problem: "I think the fact is that this is just too difficult a problem for an economist equipped with only verbal methods, even someone of Hume's remarkable powers" ("Nuetrality of Money" 248). But, he thinks, Hume admirably got a lot of things right: not only did Hume deduce the robust conclusion that, as Lucas puts it, "prices respond proportionately to changes in money in the long run", but he recognized the dependence of this result on people's expectations. Lucas observes that in Hume's accounts of monetary expansions and contractions "the motives and expectations of economic actors during the transition are described, even rationalized: The adjustment to a new equilibrium is not seen as a purely mechanical *tatonnement*

process, the character of which is determined by forces apart from the producers and consumers of the system" (253).[38]

b. Making Correct Self-Defeating Predictions

The twentieth-century notion of rational expectations originated in a 1961 paper by John Muth[39]; Muth in turn used a result from Grunberg and Modigliani's 1954 "The Predictability of Social Events". Grunberg and Modigliani addressed the specific problem of the influence of people's expectations of particular future events on those future events. Their investigation has a illuminating twist: they begin with the problem that self-defeating expectations pose for making correct predictions --- here expectations are regarded as troublesome interference; they end up showing that the way to make correct predictions involves taking advantage of the force of self-fulfilling expectations --- here expectations are regarded as forces that help bring the predicted event to fruition. The problem they set out is this: "The fact that human beings react to the expectations of future events seems to create difficulties for the social sciences unknown to the physical sciences; . . . in reacting to the published prediction of a future event, individuals falsify the course of events and thereby falsify the prediction" (465).

---

38 I have not given anything like a history of either the quantity theory of money or of champions of the importance of anticipations of the individuals within an economy to the effects of monetary changes in the economy. For an critical exposition of John Stuart Mill's arguments that the effects of an expansion in the monetary supply were due to the fact that the expansion was unanticipated, see Humphrey's discussion in "Two Views of Monetary Policy: The Mill-Attwood Debate Revisited".

39 Because Muth provided several different characterizations of the hypothesis of rational expectations that are equivalent to each other under certain additional assumptions, various accounts of the historical roots can be given, depending upon which characterization one chooses to focus on. A more common presentation of the development of the concept of rational expectations is the one given by Sheffrin, who focuses on the different significance Muth and Simon attributed to the principle of "certainty equivalence" (Sheffrin 2).

They address only the specific question of whether, given that one is able to make correct "private" (unpublished) social predictions, it is ever possible to make correct public social predictions, and give two kinds of cases in which it is. Their concern is the interactive nature of expectations and the social events those expectations are about:

> The problem of invalidation of a public prediction arises because the public prediction may affect the agents' expectations and thus become a determinant of their behavior. . . .
>
> ... the difficulty encountered here is not merely that of establishing the concrete form of the agents' reaction to a public prediction. As a matter of fact, once private prediction is assumed to be possible, the agents' reaction to a public prediction must also be regarded as knowable. . . .
>
> But this knowledge is not sufficient to overcome the difficulty created by the agents' reaction to public prediction. Let the forecaster be able to predict the agents' reaction to a predictive statement and adjust his prediction accordingly. Upon publication of this adjusted prediction, the agents will act differently from the way they would have acted if the adjusted prediction had not been made public; and so on and on. In short, the public prediction of social events seems to carry within itself the seeds of its own destruction. (467)

Thus, even full knowledge of how agents respond to a public prediction is not enough to deal with such problems of dynamic interaction. Grunberg and Modigliani remark on the distinctive nature of the expectation function in the mathematical formalization of the economic interaction of agents:

> The expectation function . . . offers particular methodological difficulties and is actually quite different from the other functions encountered in economic theory. . . . While the other economic laws assert what people will do, given (1) motivational postulates and (2) certain expectations, the expectation function asserts what people will predict, given certain past and current observations. In the predictive model it predicts what agents will predict. The familiar motivational postulates of economics are here of no use. (471)

What Grunberg and Modigliani show, however, is that correct prediction is possible; they do this using a formal model of a problem of predicting the price of a commodity. Their treatment is abstract in that they do not specify a specific mechanism by which agents' expectations are formed; they use only the fact that there is a function of some sort or other that relates the price agents expect to hold at a future time (t+1) to the actual market price at time t and the publicly predicted price (if there is one). Agents will vary their production of a good depending on the price they expect; in turn, the supply of a good affects the price that obtains. In this formal model, in general, the price agents expect to hold at time (t + 1) will not be the same as the price that does obtain then; in general, publication of the correct private prediction leads to its falsification. However, in order to find out whether it is possible that there is some public prediction of a price that is correct, i.e., that does coincide with the price that obtains, they examine the case with the addition of the simple mathematical constraint that the price publicly predicted to hold at (t + 1) is equal to the price that actually does obtain then (469). A solution to the model with the constraint that the public prediction is correct is then shown to exist: it is the situation in which the suppliers expect the value that is predicted. In fact, in the model they exhibit, that is the only solution to a model with the constraint that the public prediction be correct.

What does the difference between the case where an arbitrary incorrect public prediction is made and the special case where a correct public prediction is made look like from the point of view of

the agents making decisions about how much of a good to produce based on their expectations of the future price of that good? In the former case, the expectation they act on turns out to be wrong: "suppliers hold an unwarranted expectation and, acting upon it, bring to market a supply that results in a price different from the expected price" (468). In the latter case, the fact that a public prediction has been made and that it is the one that would be correct if acted upon gives the suppliers a correct expectation of the price upon which to act, should they all take it as such: "As suppliers fully accept the public prediction --- which turns out to be correct --- they act on the basis of warranted expectation" (469). That is, the expectation function (how the price expected by producers of a good depends upon the current price and the publicly predicted price), the form of which was not specified in the original problem, becomes specified as simply this: the price producers expect is the publicly predicted price.

They go on to show that the conditions under which such a solution exists are quite reasonable and actually do obtain in most real-life situations. Thus, they conclude that "It is false that the reaction of agents --- where it occurs -- necessarily or normally falsifies public prediction" (475). Of course, there is nothing to prevent agents from deliberately trying to falsify predictions; Grunberg and Modigliani's remarks about why this is unlikely have to do with the practical difficulties of carrying out such a project. But, by the same token, in the general case where private and public prediction differ, and so in which the course of events will be affected by the public announcement of the predicted price, the fact that the agents can change their expectations in response to the publication of a price prediction actually increases predictive ability: "The forecaster may, for instance, have knowledge of how agents react to given expectations but little information on the determinants of these expectations. To the extent that expectations are determined by public prediction, such prediction then supplies the forecaster with the missing information" (476). That the forecaster can affect the course of events isn't as scary as it sounds, though: there are, of course, many factors that are not up to the forecaster. Thus the forecaster cannot arbitrarily choose a price to predict; in most cases there will in fact exist only one possible

correct public prediction and so "the forecaster cannot at the same time both manipulate people and make correct public predictions" (476).

Grunberg and Modigliani thus establish that correct public prediction is never ruled out, distinguishing two situations. Either publication of a prediction does not have an effect or it does have an effect. If publication of a prediction does not have an effect, then the case is similar to cases in the non-social sciences: one can make a correct public prediction given that one knows how to make a private prediction. If publication of a prediction does have an effect, then (given some reasonable assumptions that can be shown to hold in most situations), though publication of the prediction will affect the course of events, there will be (at least) one prediction that will be correct if it is published. In the latter case, the predicted value takes into account the agents' reaction to it, i.e., that agents' expectations do alter the course of events. The point is somewhat subtle: for this (usually unique) event, the published prediction is correct, because the agents act to change the course of events (from what the course of events would have been had the prediction not been published) so as to bring about the predicted event.[40] This is not true for just any event one might choose to predict; in fact, in most cases there is only one such event for which this holds.

These insights were not unprecedented; various accounts of the history of the concept of rational expectations cite Alfred Marshall's 1890 Principles of Economics and J. R. Hicks' 1939 Value and Capital as sources of precedents of the concept of attributing model-consistent expectations to the economic agents in the economy one is investigating (Grossman 542). Grossman points out that Marshall described the reasoning a manufacturer would go through in

---

[40] Their proof makes use of Brouwer's fixed-point theorem, the same theorem Nash used to show that there was always a solution to certain game-theoretic formulations of the bargaining problem. Grunberg and Modigliani credit Herb Simon with suggesting the use of the theorem to them (465).

estimating future wages he would have to pay that captured what Grossman calls "the essence" of rational expectations: "agents have some economic model of the economy which relates the exogenous variables to the endogenous variables which he is interested in forecasting" (542), and that Hicks emphasized that "the expectations held by firms about future endogenous variables (such as tommorrow's price of corn) actually help determine the true value of the future endogenous variables. Thus [what Hicks called perfect foresight] is an equilibrium concept rather than a condition of individual rationality" (542).

c. Correct Foresight and Compatibility of Expectations

In his 1937 "Economics and Knowledge", F. A. Hayek had stressed the crucial and distinctive role that assumptions about the acquisition of knowledge and, especially, foresight played in economic theory (33) , and referred to Irving Fisher as a precursor who recognized " the significance of anticipations" (34 fn) in his 1896 Appreciation and Interest. Like Grunberg and Modigliani, Hayek, too, had concluded (though without mathematical formalism) that "Correct foresight is then not, as it has sometimes been understood, a precondition which must exist in order that equilibrium may be arrived at. It is rather the defining characteristic of a state of equilibrium" (42); by "equilibrium" of a society, he meant "that the different plans which the individuals composing [a society] have made for action in time are mutually compatible" (41).

Although Hayek was writing about economics, commerce, and prices, the notion of equilibrium he employs is really the notion of equilibrium described above in Chapter IV as an "expectational equilibrium", rather than the Nash equilibrium, which is defined in terms of optimal strategies. Hayek had given nontechnical arguments for the existence of states of society in which agents had correct foresight about the things needed in order for them to carry out interdependent plans. As Grunberg and Modigliani did in their (1954) discussion, Hayek had distinguished between an economist making forecasts about an economy or society and an individual in the

economy or society about which forecasts were being made. Hayek diagnosed the confusion in economics at the time as turning on a conflation of senses of "data":

> There seems to be no possible doubt that these two concepts of "data," on the one hand, in the sense of the objective real facts, as the observing economist is supposed to know them, and, on the other, in the subjective sense, as things known to the persons whose behavior we try to explain, are really fundamentally different and ought to be carefully distinguished. . . . the question why the data in the subjective sense of the term should ever come to correspond to the objective data is one of the main problems we have to answer. (39)

Hayek's answer to this question is based on the insight that compatible expectations are necessary for successful social and economic interaction among economic actors. Since one person's plans to build a house, for instance, can only be carried out if those plans are compatible with the plans of the people who produce and sell building materials, provide services, and supply financing, their plans must be mutually compatible --- else some will have to revise their expectations. This insight is in conflict with the approach of assuming that every economic actor involved in building the house knows everything relevant to any other economic actor involved in building the house. Hayek proposes that economists accept that the problem they face involves a division of knowledge, just as they accept that it involves a division of labor: the problem is the problem of "how the spontaneous interaction of a number of people, each possessing only bits of knowledge, brings about a state of affairs in which prices correspond to costs, etc." (50-51) Hayek did not have the mathematical tools that Grunberg and Modigliani applied to the problem of the prediction of social events and that John Nash applied to bargaining problems, but his conclusion that correct foresight is not inconsistent with the fact that, in the social sciences, expectations affect the course of affairs is essentially the central insight of Grunberg and Modigliani's paper.

In examining Hayek's investigations into how correct foresight can come about, I am not interested in the details of his reasoning, but in his general approach. One of the questions most crucial to figuring out how people form expectations about the future based on their past experience is which information is relevant to what the future will be like. How can this be done, if not in an objective (i.e., nonsubjective) way? The question has a Gibsonian ring, and Hayek's answer is the economic agent's analogue of "affordances": consistently with arguing for a division of knowledge, he explains that what counts as relevant information to an individual are those things an individual needs to carry out his plan, or, alternatively, those things that would require him to alter his plan. Joint action between economic or social actors provides a way to pick out changes (or, put alternatively, the invariants) in the environment without requiring a notion of change with respect to some absolute state. For, the changes or invariants picked out as the relevant ones are identified according to how they diverge from the expectations on which the actors' plans were based:

> . . . it seems hardly possible to attach any definite meaning to the much used concept of a change in the (objective) data unless we distinguish between external developments in conformity with, and those different from, what has been expected, and define as a "change" any divergence of the actual from the expected development, irrespective of whether it means a "change" in some absolute sense. . . . But all this means that we can speak of a change in data only if equilibrium in the first sense exists, that is, if expectations coincide. (40-41)

In "The Use of Knowledge in Society," Hayek explicitly drew an analogy between two mechanisms: (i) the mechanisms one individual (with access to all the information known by anyone) running the economy would use to coordinate an economic plan and (ii) the mechanisms by which the actions of single individual economic actors are coordinated. I suppose an example

of the latter could be that a builder perceives that a line he's marking for where a board is to go has the same direction as the line made by someone else for where a window is to go, without having to determine what that direction is with respect to anything else. Hayek compared coordination of information within a person to coordination of information within an economic system: "Fundamentally, in a system in which the knowledge of the relevant facts is dispersed among many people, prices can act to coordinate the separate actions of different people in the same way as subjective values help the individual to coordinate the parts of his plan" (85). This explanation of how correct foresight comes about, however, seems to come at the expense of saying that the economist trying to figure out how the economy will behave must accept that his or her task is insuperable. For, Hayek's view would require the econometrician to first formulate the plans of all the individual agents, in order to determine which invariants are important. Though it is an explanation of how what he calls correct foresight can come about for some kinds of events, it does not provide a practical economic methodology.

d. Economists' Predictions and Economic Agents' Expectations

Thus by the time John Muth addressed the problem of predicting price movements in the proposal Lucas later developed into a methodology of economic modelling that was to became known as "the rational expectations revolution", the idea of two kinds of expectations being in play within an economic model (i.e., the economic modeller's prediction and the expectations of the agent in the model) had already been proffered. Muth, however, circumvents the problem of having to identify a wide variety of different interlocking plans of individual economic agents, in much the way that Grunberg and Modigliani did. What Muth suggested in his 1960 "Rational Expectations and the Theory of Price Movements" was this: "In order to explain fairly simply how expectations are formed, we advance the hypothesis that they are essentially the same as the predictions of the relevant economic theory." Muth was addressing the problem of how prices change over time within the context of the activity of making dynamic economic models, though

part of this activity involved comparison of model predictions with data in real economies. He granted that the expectations of economic agents are often in error, yet, he claimed, "dynamic economic models do not assume enough rationality" (4). He gave a more precise statement of his hypothesis in terms of probability distributions: "expectations of firms (or, more generally, the subjective probability distribution of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (or the "objective" probability distributions of outcomes.)" His reformulation of the hypothesis in terms of three assertions includes a claim that uses Grunberg and Modigliani's terminology and shows a sensitivity to something important on Hayek's view[41]: that expectation formation is dependent on the economic system in which it occurs:

> The hypothesis asserts three things: (1) Information is scarce, and the economic system generally does not waste it. (2) The way expectations are formed depends specifically on the structure of the relevant system describing the economy. (3) A "public prediction," in the sense of Grunberg and Modigliani (1954), will have no substantial effect on the operation of the economic system (unless it is based on inside information). This is not quite the same thing as stating that the marginal revenue product of economics is zero, because expectations of a single firm may still be subject to greater error than the theory. (5)

What the third of the assertions in the quote above implies is that the information that individuals have access to in the society and economy they inhabit is sufficient for them to be able to do what the forecaster imagined in Grunberg and Modigliani's paper did: figure out the price p such that, if all the individuals in the economy acted upon the expectation that the price at time (t + 1) would be

---

[41] Muth explicitly mentions Grunberg and Modigliani's paper. As Muth does not mention Hayek, I do not know if he would agree with the connections I've made here; I mean only to be pointing out related themes, not suggesting causal historical claims.

p, they would bring it to pass that the price that actually obtains at time (t + 1) would be p. Muth's statement is a bit more sophisticated, for he allows the expectations that the economic actors have to vary from actor to actor; what he is appealing to is the value of the price about which their expectations are distributed.

e. Conventions and Prices

This suggestion --- that what can be accomplished in an analytic model of an economic system by an economic forecaster making an explicit announcement can, alternatively, be achieved by the individuals in the system without the use of an economic forecaster to coordinate their actions --- is structurally very much like David Lewis' point in Convention. Lewis' driving insight (inspired by Hume) was that the agents party to a convention can coordinate their actions without ever making an explicit agreement. The individuals in Lewis' account can achieve the same joint result -- establishment of a convention in the society they inhabit --- that they would have achieved by explicitly agreeing on a convention, but this is instead achieved by a certain kind of system of mutual expectations upon which they all act. The kind of mutual expectations required can be generated by a public announcement that serves as a basis for common knowledge, or by an agreement (so long as the agreement is to conform conditionally on others' conformance), but they can alternatively be generated by salience or precedent. Notice the analogous point would hold for Lewis: once a convention is established, it does not make a difference if the convention is explicitly announced, whereas it would make a difference otherwise. That is, prior to the establishment of the convention of driving on the right hand side of the road, making such an announcement could change the course of events; after the establishment of the convention, announcing that most everyone is going to drive on the right hand side of the road (so long as others do) would be superfluous.

Muth's view that coordinated action is capable of producing the same price movements that would obtain were prices announced by an economic forecaster is like Lewis' in several other ways as well: Muth emphasizes that his hypothesis "does not assert that the scratch work of entrepreneurs resembles the system of equations in any way" (5). Lewis had pointed out that he was not asserting that individuals actually went through the process of reasoning formalized by the explicit deductions he exhibited to show that a system of mutually concordant expectations could issue in conforming behavior via certain kinds of reasons --- only that their actions could be justified by such deductive reasoning (55). And, later in that work, in speaking of what competence in a certain human language requires, he wrote: "The user of [the language] is a finite being with very limited experience; yet somehow he has acquired an enormous, and enormously varied, repertory of propensities to action, expectation, and preference in a wide variety of situations. That is what it takes for him to be -- as we but not he would say -- habitually truthful in [the language]." The competent language user, Lewis thought, really does have something analogous to a concept of truth, just as a competent bicycle rider has something analogous to the concepts and knowledge of laws of physics, although "unless our bicycle rider happens also to be a physicist, it would be wrong to say that he had those concepts or knew those laws, although his expectations regarding a wide variety of particular situations would work according to those laws" (184). Similarly, on Muth's suggestion, the economic actors' behavior can be justified by the reasoning the economic forecaster would use, although we need not claim that they actually go through the same reasoning process as the forecaster would.

In Chapter IV, I questioned Lewis' approach for trying to do without interaction between agents during dynamic processes in which they interact (his "windowless monad" approach), and for requiring agents' expectations to be symmetrical. Muth's approach, though, does not contain the analogues of the things for which I questioned Lewis' approach. Muth does not say anything that would rule out interaction between agents during the social and economic processes that lead to the joint result; he shows no interest in arguing that individuals can reason as "windowless

monads" about each other's actions. Neither does he require that all agents have the same expectations; he follows the statement of the three assertions asserted by his hypothesis with the clarification: "nor does [the hypothesis] state that predictions of entrepreneurs are perfect or that their expectations are all the same" (5). [42]

Lucas used Muth's suggestion to address the problem evident in Hume's essays on money and interest: the reasoning Hume used to argue for monetary neutrality in the long run wasn't used in (in fact, was contradicted by) his explanations of short-term effects. Lucas suggests that the different conclusions Hume reached for long term and short term changes in the quantity of money were a result of the way the changes in the quantity of money were effected: "there is something a little magical about the way that changes in money come about in Hume's examples [of long term monetary changes]. All the gold in England gets 'annihilated.' Elsewhere he asks us to 'suppose that, by miracle, every man in Great Britain should have five pounds slipt into his pocket in one night.' Money changes in reality do not occur by such means" ("Neutrality of Money" 247) whereas the descriptions of the changes in monetary supply in the examples where he is explaining short term effects of increased industry (output) are more realistic, in that the money increase is localized and takes time to disperse. Lucas applied Muth's suggestion in order to avoid similar kinds of inconsistent reasoning in the economic models that were used in the early 1960s. He describes the state of things in economics at that time:

> . . . models of individual decisions over time necessarily involve expected, future prices. Some microeconomic analyses treated these prices as known; others imputed adaptive forecasting rules to maximizing firms and households. However

---

[42] One finds many authors claiming that the hypothesis of rational expectations entails that all agents know everything, and/or that all agents' knowledge is the same (e.g., Phelps). I am not clear on what these authors mean when they say such things; certainly there are many specific models in which the use of rational expectations is incorporated with these constraints, but it is also clear that these constraints are not part of Muth's hypothesis.

> it was done, though, the "church supper" models assembled from such individual components implied behavior of actual equilibrium prices and incomes that bore no relation to, and were in general grossly inconsistent with, the price expectations that the theory imputed to individual agents.
>
> As intertemporal elements and expectations came to play an increasingly explicit and important role, this modeling inconsistency became more and more glaring. (254-255)

Applying Muth's suggestion that "expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory" (4) resolved the modeling inconsistency. It should be noted that Muth cited general findings from empirical studies of expectations that motivated his suggestion (4), and which his suggestion was intended to explain; thus the motivation for using Muth's suggestion was not simply that it would get rid of a modeling inconsistency, but also that it explained results of empirical studies of expectations. In Lucas' words, Muth "showed how [the inconsistency] could be removed by taking into account the influences of prices, including future prices, on quantities and simultaneously the effects of quantities on equilibrium" (255).

What Lucas did in his 1972 "Expectations and the Neutrality of Money" was apply the constraint Muth suggested, that "expectations of firms (or, more generally, the subjective probability distribution of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (or the 'objective' probability distributions of outcomes)" (Muth 4-5) in a simple model he constructed. In his model, there are two physically separated markets, and a given agent gets information on the current state of real and monetary disturbances only through the prices in the market where he or she happens to be. Lucas noted that the model was an oversimplification: "it has been necessary to adopt a framework simple enough to permit a precise

specification of the information available to each trader at each point in time, to facilitate verification of the rationality of each trader's behavior" (84).

However, though the model was very simple, what was important was that it provided a precise and consistent example in which long-run neutrality of money holds and in which the fact that monetary changes can also have real consequences not only holds but is explained: "monetary changes have real consequences only because agents cannot discriminate perfectly between real and monetary demand shifts" (78). What I mean by "consistent" here is that the model did not exhibit the feature Lucas complained of in the patched-together economic models of the 1960s; i.e., the feature that they "implied behavior of actual equilibrium prices and incomes that bore no relation to, and were in general grossly inconsistent with, the price expectations that the theory imputed to individual agents" (Lucas, "Neutrality of Money"254). The arguments in his 1972 paper also illustrated what Lucas presented as a new concept of equilibrium; he defined equilibrium prices and quantities as functions of possible states of the economy, where "states" of the economy are represented mathematically as vectors of finite dimension. This definition of equilibrium values amounts to the assumption that the state variables of an economy determine the equilibrium price associated with the state of the economy; i.e., that the path by which the state was achieved is not relevant. However, the information that people actually have about the state of the economy is the current price; prices are functions of --- and so can provide information about --- the state of the economy.

Lucas put a great deal of stock in the advantages to be gained by the advance in mathematical formalism he employed in the 1972 paper: "This characterization permits a treatment of the relation of information to expectations which is in some ways much more satisfactory than is possible with conventional adaptive expectations hypotheses" (67). He can't mean that the new characterization allows him to be more explicit about the form of the expectation function as compared to alternative adaptive expectations approaches; the expectation function is not

specified beyond the state variables on which a price expectation depends, and the fact that price expectations are model-consistent. What he can mean to say is that he can be more mathematically precise about how an individual in the model determines the future price from the current price in the sense that he is not claiming a specific approximation that is bound to be model inconsistent. Again, no algorithm for this is provided; rather, the expectation function is a mathematical object about which and with which one can reason. Thus Lucas regarded as a virtue what Simon felt was a shortcoming. In his own Nobel address in 1978, Simon lamented that neoclassical economics was emphasized at the expense of descriptive decision theory. Simon, Muth, and Modigliani had worked together in research in the field of dynamic programming (Simon 1978; Sheffrin 1) but, whereas Simon generalized the result they obtained based on seeing in it an account of rational decision making as "an approximating, satisficing simplification" (505), Muth instead generalized from the result to the hypothesis of rational expectations based on seeing it as evidence that the models they were using did not assume enough rationality on the part of the economic agents. Simon remarked that Muth "would cut the Gordian knot. Instead of dealing with uncertainty by elaborating the model of the decision process, he would once and for all --- if his hypothesis were correct --- make process irrelevant" (505). This is just what Lucas thought an improvement; he noted that agents who think they have a sufficiently good model of the economy on which to base their decision -- as agents modeled using adaptive expectations do -- are too confident.

Rather, Lucas does not specify any particular process by which expectations are formed. The constraint on price expectations in an economic model is that an agent's expectations of a price satisfy :

$$p^* = f(p^*)$$

where f is the function mapping expected prices to future prices and p^* is the specific price that the agent expects to hold at a certain future time. One could see this as what is constitutive of a rational agent capable of forming expectations; i.e., that rational agents will have learnt how to

develop expectations such that, based on the information available to them, the price that will obtain is the one they have expected will obtain. Notice that the fact that the agent's expectations are causal does not raise any special problem. But consider that, analogously, the fact that an agent's action may affect his or her environment does not cause a problem for the more commonly recognized constraint on rationality (i.e., that an agent uses effective strategies). Significantly, that the point for the expectational analogue holds means that we can apply any notion of rationality founded on this self-reflective capability to expectations even if expectations are causal in ways other than via rational intentional action. That is, we can apply such a notion of rationality to an agent who has learned to develop his or her expectations in such a way that, barring unanticipated shocks to his or her environment, it turns out that his or her expectations are correct: the agent forms an expectation of what is to happen at time (t+1), and, at time (t+1), the agent says "Well, I was right about that." The constraint on expectations is the expectational analogue of the more commonly recognized constraint on rational agents that they learn strategies (or how to act, or how to decide) such that, in retrospect, they decide that they took the best action available to them; again, this claim is made with the proviso: "barring any unanticipated shocks."

The reader may wonder here: if what counts as an unanticipated shock can then be defined just as one that the agent didn't expect, and the constraint on expectations of an agent is that the agent's expectations turn out to be correct, barring unanticipated shocks, isn't that circular? The answer is that there is an interrelation between the expectations one has and which shocks are deemed unanticipated, but all the interrelation means is that expectations and unanticipated shocks are co-defined. Thus the distinction between anticipated and unanticipated shocks is relative to a system of expectations. But that makes sense; figuratively speaking, an agent (in an economic model) whose expectations are rational is one whose expectations reflect the layout of the world in which the econometrician has postulated he is located, which may be bumpy and changing in time. Whether the bump is anticipated is determined by whether or not the agent

anticipated it; whether someone should have been able to anticipate that bump is not. Remember that Muth's suggestion was based on equating the expectations of the agents being observed to those of the econometrician observer ("for the same information set"). Though Muth did not explicitly say so, of course it was implied that he meant to be imputing rationality to the econometrician. In explaining that attributing rational expectations to agents does not mean attributing unnatural powers of perfect divination to them, but attributing epistemic abilities comparable to a rational observer, he wrote that the hypothesis "asserts that agents' responses become predictable to outside observers only when there can be some confidence that agents and observers share a common view of the nature of the shocks which must be forecast by both."

The idea that the role of the econometrician attributing rationality to the individuals whose collective behavior he or she is trying to predict is fundamental to the notion of rational expectations can be seen as a collective analogue of a Davidsonian idea: i.e., that attributing rationality to an individual is something properly done only in the context of one rational being making sense of that individual's actions in terms of his or her beliefs. The analogy suggested is that one could extend the view of rationality in such a way that the corresponding notion of belief, i.e., the notion of belief whereby beliefs are only efficacious via being part of a reason, could be extended to a notion of beliefs and expectations that are causal via means other than rational intentional action. That is, we can extend the notion of rationality thus: attributing rationality to individuals is something properly done only in the context of a rational being making sense of a group of interacting individuals' responses in terms of the expectations attributed to them (perhaps collectively). On this notion of rationality, the methodology of rational expectations would then amount to the econometrician regarding the interacting economic agents as rational, independently of whether the agents' behavior could be understood as intentional action following from agents' "decision rules".

3. Rational Expectations and Anticipatory Schemata

I want to explain why I think the account of how beliefs are causally efficacious I've offered in previous chapters is particularly compatible with the rational expectations approach. On the account offered in earlier chapters, beliefs and expectations are efficacious via being part of an individual's dynamic schema, and are formed as the result of a process of an individual agent employing the schema in an interactive cycle of perception, interaction, and expectation formation. I did not specify how schemata change, other than to describe the effects of a schema in directing one's perception, cognition, and action. The view of beliefs (as being incorporated into such dynamic anticipatory schemata) is compatible with views in which rational expectations are constitutive of coordinated interaction. In contrast, the view I have proposed would not be as compatible with some alternative approaches, for example, any approach in which an algorithm by which adaptive expectations are produced is specified. On the latter kind of approach, the problem of how expectations are efficacious in economic interactions is reduced to a problem in descriptive decision theory. This view is held by Herb Simon, and he is explicit not only about holding it, but about urging others to adopt it as well:

> . . . the salient characteristic of the decision tools employed in management science is that they have to be capable of actually making or recommending decisions, taking as their inputs the kinds of empirical data that are available in the real world, and performing only such computations as can reasonably be performed by existing desk calculators or, a little later, electronic computers. . . . Models have to be fashioned with an eye to practical computability, no matter how severe the approximations and simplifications that are thereby imposed on them.
>
> . . . decision makers can satisfice either by finding optimum solutions for a simplified world, or by finding satisfactory solutions for a more realistic world. (498)

Now, it seems to me that the point David Lewis made about a competent speaker of a language having a concept analogous to truth, just as the competent bicycle rider has something analogous to the concepts and laws of physics, has gone unrecognized here. There is a slide from what "decision tools" have to be able to do, to how "[m]odels have to be fashioned." Why should the economic agent in an econometrician's model have imposed upon him or her the constraint of using a decision process that can be formulated in equations chosen with an eye to "practical computability"? The bicycle rider is not so much approximating the laws of physics (whether by applying simplified equations to the real world situation or using a simplified model of the world to which the exact laws of physics are applied) as he or she is doing something else that is as good as using the laws of physics. What the flesh and blood human being riding the bicycle is doing might simultaneously include cognitive activity, physiological response, and unconscious habit. In short, the notion of a schema fits here; the bicycle rider forms expectations of the forces to be imparted to the bicycle, and of the postures he or she should take on in response to them, perhaps the rider forms images of what he'll see and do next, or develops physiological anticipations that aid in picking up more information. The same could be said of the economic agent in a model. If the economist is building a simplified model of a physical environment (e.g., specifying what resources are needed to produce certain goods), as is usually the case, then the expectations imputed to the agent should be consistent with the simplified environment the economist has constructed for that agent. The same would hold for a biophysicist modelling the expectations of a bicycle rider.

I do not mean to imply that Lucas ever meant to be making room for a view such as mine; perhaps not even one economist has ever entertained the idea of an agent's expectations being causally efficacious by anything other than intentional action; economists tend to talk about people's behavior being a function of their "decision rules" and inputs to those rules. But I do mean that a fundamental difference between the rational expectations approach and what might be called a descriptive decision theoretic approach to behavior of economic agents is that the former does

not restrict the ways in which beliefs and expectations can be causally efficacious (within the model) to rational intentional action.

Although Lucas does not emphasize it in the early 1972 paper, it is also true that his approach recognizes that agents' expectations of prices can both determine and be determined by expectations of prices. In this respect, too, the proposed view that beliefs are causally efficacious via dynamic interactive schemata is well suited to the rational expectations approach to economic modeling. For, unlike models using descriptive decision theory to model agents' behavior, models using rational expectations not only do not specify how an agents' expectations are formed, but they do not specify how an agent's expectations affect the formation of another agent's expectations either. I have seen this lack of presumption as a virtue, in that it does not prescribe empirical processes a priori. Hayek, however, points out that this means equilibrium analysis, based as it is on economic interactions, has limits:

> . . . if the tendency toward equilibrium, which on empirical grounds we have reason to believe to exist, is only toward an equilibrium relative to that knowledge which people will acquire in the course of their economic activity, and if any other change of knowledge must be regarded as a "change in the data" in the usual sense of the term, which falls outside the sphere of equilibrium analysis, this would mean that equilbrium analysis can really tell us nothing about the significance of such changes in knowledge . . . (55)

This negative point suggests a positive point: if one could expand one's model from economic activity to a larger realm of activity, more knowledge would fall within the sphere of equilibrium analysis, and so the changes in knowledge would be of more significance. However, it was not Hayek's intent to give such encouragement to theorists. In his critique The New Classical Macroeconomics Kevin Hoover points out that the Austrian view of macroeconomics (typified by

economists such as Hayek and Hicks) is only superficially similar to the rational expectations approach in that the Austrians did not think that macroeconomics was susceptible of a closed form solution (1984, 253-257). Whereas Lucas thought that the difference between his own approach and that of the Austrian economists was due mainly to the mathematical techniques he had and they did not, Hoover thinks the Austrian school held that the problems in economic modelling were intractable in principle due to a richness of detail in real economic systems, not due to lacking effective abstract mathematical tools.[43]

However, as Hoover also points out, Lucas often seems to think that, as the rational expectations approach is further developed, the models it produces will become richer and richer in detail. I don't see why this couldn't be so. Lucas' approach requires identifying which variables are relevant (the state variables) to the behavior of an economic system made up of interacting individuals in some environment (here the environment includes features of a political system such as taxation policy and property laws as well as variables of nature such as the weather); doing this for a system of even moderately interesting and perceptive individuals will always leave out some detail about how the individual's behavior depends upon his and others' expectations and the information available to him, but this doesn't mean that, in the context of answering a specific question, it couldn't catch an arbitrarily large amount of detail.

To see what I mean by catching an arbitrarily large amount of detail, consider what the analogue of the rational expectations approach for a cognitive psychologist/biophysicist modelling a bicyclist's

---

[43] The last chapter of Hoover's The New Classical Macroeconomics, entitled "An Austrian Revival?", is devoted to this comparison. He remarks "Austrian business cycle theorists reject the notion of equilibrium business cycles so dear to new classical hearts not because they lack an appropriate concept of intertemporal equilibrium nor because they lack the concept of rational expectations or the appropriate mathematical techniques. Instead their understanding of the nature of rationality and human action rules out the new classicals' conception of a closed economics and their related analyses of rational expectations and equilibrium" (257).

behavior would be. The cognitive psychologist/biophysicist might not capture all the nuances of how the bicyclist develops expectations and acts on them, but by specifying which features of the environment are being modelled (turns, hills, bumpy patches), there doesn't seem to be anything in principle to prevent such a modeller from capturing any particular detail desired; the modeller should get different responses depending on whether a dip in the road is unanticipated or anticipated, but he need not fill in the details of how the bicyclist actually manages to pick up the information that the dip exists. If the model is then expanded to include not only the physical features of the landscape, but the effect on the bicyclist of road signs, billboards, other bicyclists' hand signals, other bicyclists' smiles, and so on, we can see that the approach yields not just knowledge of the features of the world that the discipline of physics covers, but knowledge of any sort of aspect of the world one might care to include.

On the view I propose --- economic agents' expectations and beliefs affect their behavior via multimodal anticipatory dynamic schemata that may differ from agent to agent, but may have features in common --- I locate the richness of detail in an economic system in the anticipatory schemata of individual agents, as well as in the modes whereby agents' anticipatory schemata are efficacious, i.e., how expectations of one agent affect the expectations and actions of another. I identified these modes in previous chapters: expectations can be causal via their role in causing intentional actions, in producing physiological changes in the person having those expectations (examples cited were one's inner ear muscles tightening so as to resonate to a certain anticipated tone, blushing, becoming nauseated at an unpleasant realization, the placebo effect, and becoming weak-kneed upon realizing one is at risk); they can also be causal via their role in giving rise to images, thoughts, or the formation of other beliefs. I argued, in addition, that when the environment one is perceiving, reasoning about, and interacting with does include other perceivers, expectations can be causal in other ways as well; I have pointed here to the phenomenon of the effect of one's expectations on a perceiver who perceives the thoughts and expectations of another (physiognomic perception) and to the phenomenon of stereotype threat.

I grant that any particular economic model employing rational expectations is going to have to limit the description of the expectation function (e.g., a mapping of current price and other variables to future price) in a general way: it will identify the kinds of information that can be picked up and the kinds of behaviors that can be produced. And, the function is not a function of the expectations of one individual, but of some measure of the expectations of the whole population in the economy collectively. Thus it does not specify the algorithm by which an agents' expectations are formed or how his or her behavior is determined in the sense that one could use the functions to predict agents' expectations or behavior, either individually or collectively. What is assumed is that, like a competent bicycle rider who knows no mathematics, the individual agents in the economic system collectively are competent at forming expectations of future values of variables relevant to their plans. Individual agents develop the anticipations that enable them to pick up further relevant information about the environment in such a way that the behavior of the economic system is as though the information were publicly announced. As discussed above, this does not rule out the existence of unanticipated shocks in the environment; what is true is that the distinction between anticipated changes in the environment (e.g., changes in the seasons, taxation rates) and unanticipated ones (e.g., earthquakes, crop shortages due to unseasonable floods, emergency fuel taxes) are relative to the expectations of the collection of interacting agents.

B. Expectations of Others' Expectations

The formalism Lucas used to talk about economic models allowed him to represent not only the fact that agents' expectations depend upon the information presumed available to the agents by the econometrician who has postulated the features of the environment he inhabits, but also the fact that expectations of future events such as future prices depend upon expectations and beliefs about other individuals' expectations and beliefs. David Lewis found expectations of

others' expectations crucial as well. In his account of convention, Lewis was concerned to distinguish between cases of identical behavior produced by expectations that differed in kind: he wanted to rule out as a case of conforming to the convention of driving on the right hand side of the road the case in which someone drove on the right hand side of the road because of his expectations that all the others would, but thought that everyone else drove on the right hand side of the road out of habit, without regard for what anyone else did. A proper case of conforming to convention, Lewis thought, would include the driver's recognition of a system of mutual expectations between drivers: not only does each one expect each other one to drive on the right hand side, but each expects the other to expect him to drive on the right hand side, and so on.

It is just this feature of coordinated interaction that Lucas focused on in his critique of econometric models used for policy evaluation in the 1960s: he criticised the illegitimate presumption of the constancy of agents' decision rules under changes in the variables that determined the behavior of the economic system as a whole. So, for instance, if it were announced that everyone was to drive in the left lane, the drivers in the cases Lewis included as legitimate cases of convention would be able to assume that (in economists' terminology) the other drivers would change their "decision rules" accordingly; the drivers in the cases Lewis wanted to exclude would not. Lewis did not explicitly discuss how the cases would differ under a change in announced policy, but, as I explained in Chapter IV, he cited this case as the kind of case he meant to be excluding by the common knowledge provision in his definition of convention. The effect that implementation of a new economic policy will have is just the subject Lucas was addressing when proposing that rational expectations be used in the econometric models used to make predictions.

Suppose one wants to investigate how an economy will respond to a change in policy or an injection of money. The problem, as Lucas put it in "Models of Business Cycles", is "to go from non-experimental observations on the past behavior of the economy to inferences about the

future behavior of the economy under alternative assumptions about the way policy is conducted" (7). The econometrician's tool to make predictions is a mathematical model of the economy, consisting of equations that he thinks capture the causal relationships between various variables, including causal relationships that are internal to the economy as well as external to it. As for a model, Lucas says, "we want a model that fits historical data and that can be simulated to give reliable estimates of the effects of various policies on future behavior." He goes on: "But what data? And what do we mean by fit? And when can we expect that particular simulations will be reliable? These are hard questions, harder and more open than is commonly acknowledged" (7).

In building an economic model, the economy is presumed to consist of various kinds of individual decision makers, e.g., firms that produce consumer goods, firms that supply materials to other manufacturing firms, farmers who choose what and how much to grow each year, individuals or households who buy consumer goods and services, and so on. Individual decision makers are presumed to take actions in response to how things are in their environments according to functions generally referred to as "decision rules"; as I have indicated before, despite the term, there is no principled reason to restrict these functions to tractable rules. We can treat these functions describing individual behavior --- there is one such function for each individual agent -- as unspecified. I have suggested that these so-called decision rules be thought of as anticipatory dynamic schemata, directing an agent's acquisition of information, being revised as a result of that information, and directing the agent's actions. However, I will retain the terminology "decision rules" in what follows. An agent's decision rules must take into account relevant features of the environment, including not only the state of the economy, but general rules about how the environment behaves, as well as suspicions as to the kinds of unpredictable things that can happen (seasonal changes are expected, droughts are possible, that the sun be extinguished is neither). This much about economic modelling is mundane. However, once it is recognized that the state of the economy depends on the actions other agents take, and that part of one's environment includes the "decision rules" of other individual agents, it becomes clear that

knowing how other agents' decision rules would change under a change in policy is crucial to any agent's decision-making process. Thus, the model of the environment attributed to each agent must include some specification of which aspects of other agents' behavior will remain the same after a shift in policy and which aspects of other agents' behavior will change in response to shifts in policy. It is this fact that raises a problem for any attempts to model agents' behavior.

Dropping the economic jargon, the point is that an economic model must include not only agents' expectations of other agents' behavior, but expectations of other agents' expectations. That is, if we want to capture how agent #1 behaves, we must include agent #1's expectations of how all the other agents will behave, for she will have to take their behavior into account in order to figure out how the economy will behave. Now, in order to form an expectation of what the others will do under a change in policy (e.g., eliminating all capital gain tax on stock market transactions), she must form some expectation of how agent #2, agent #3, etc. form their expectations of all the other agents' expectations and so-called "decision rules" as well. In turn, their decision rules will have to incorporate expectations of agent #1's behavior and of how her decision rules would change under a change in policy. Clearly, one cannot expect to use descriptive equations gleaned from past policy regimes to model agents' expectation formation.

The hypothesis of rational expectations described above is one way to address the shortcomings of models that attempted to specify how agents form expectations. The rational expectations hypothesis is not the only alternative. Other alternatives are adaptive learning approaches, in which agents progressively update their expectations; some models employing adaptive expectations have been shown to produce progressively updated expectations that converge to the rational expectations values. Yet another approach incorporates optimal learning rules along with rational expectations; these are "forward looking" in that they include some means of generating an expectation of, say, a price at some future time, which is then compared with the

value that actually obtains at the future time (Townsend 546; Slade 262; Lee 367).[44] However, what all these approaches have in common is that they recognize the importance of what amounts to a system of mutual expectations; an agent's environment includes other agents' expectations of her and of her expectations.

Thus, regardless of the status of the rational expectations hypothesis, the points about the crucial role played by agents' anticipations stand. An agent's behavior is conditional not only upon what others do, but upon what everyone expects everyone else to do. The effects of a policy change depend upon the expectations of the agents in the society. That it is now recognized as absolutely crucial to distinguish between anticipated and unanticipated policy actions reflects the efficacy of belief in establishing and maintaining some social institutions. However, this kind of efficacy can be accounted for on the view proposed in this dissertation that beliefs are efficacious via anticipatory dynamic schemata of the individuals interacting in the economic system.

---

[44] Many of these, including Lee and Townsend, use Kalman filtering techniques. Rick Grush (1997) has argued that the Kalman filtering technique involves features that should be considered representational, due to the fact that the method involves a means of generating expectations of how certain variables in the environment will change in response to an intervention.

CHAPTER VI

CONCLUSION AND EPILOGUE

I began with the recognition that there are various modes by which beliefs can be causally efficacious, and proposed an answer that, I argued, does account for how beliefs can be efficacious in all three of the modes I identified: (1) via intentional action (as part of a cause of the actions of an individual having the belief), (2) via social institutions (as responsible for the maintenance and existence of some social institutions and socially-constituted states of affairs), and (3) via unintentional action and physiological changes (as causal in effecting physical changes (including bodily motions) unmediated by intentional actions of a rational agent). The answer I gave was that beliefs and expectations are causally efficacious via being incorporated into dynamic anticipatory schemata that are employed in perception, cognition, and action, including social interaction.

There are many questions one could ask about belief; here I have addressed only this one: How do beliefs make a difference? I think my answer is satisfying in one sense and unsatisfying in another. I think it is satisfying in that the account of how beliefs are causal is comprehensive enough that, besides accommodating an everyday notion of belief, those in a wide variety of fields, from cognitive psychology and decision theory to social psychology and advertising, as well as those in developmental psychology and psychophysiology, will find it provides an account of how beliefs are causal that accomodates the ways in which they know beliefs, as they use the term, are efficacious. Some philosophers may find the account unsatisfying in that it is too all-encompassing; I expect some will feel that I should have recognized some constraints on belief more restrictive than the ability to be incorporated in an anticipatory schema such as a plan for touring a room. Here I would protest that I am not claiming to have provided a concept of belief; I

have provided an account of how beliefs are efficacious that is as neutral as possible about the concept of belief.

However, I think that the proposal in this dissertation that beliefs are causally efficacious via the anticipatory schemata of the individual in which they are incorporated does suggest an expanded notion of belief analogous to the notion of belief favored by those who think that, on principle, beliefs can be efficacious only via being part of a reason. The line I have in mind[45] goes as follows. Consider the mainstream view that something can only count as a belief if it can be had by someone who can give reasons for holding it and can cite it as part of a reason for holding some other belief or for performing some intentional action. Analogously, we might say: on an extended concept of belief, which is meant to include expectations, something can only count as a belief or expectation if the person who has it can employ it to anticipate what sorts of things are going to happen in his or her environment (including other agents' behavior) when, in the judgement of those trying to make sense of his or her behavior, he or she should be able to do so, or can be used to pick up information relevant to developing anticipations of what is going to happen next in his or her environment (including picking up information during interactions with others), or to direct his or her actions in accordance with the expectations of others with whom he or she is trying to coordinate (and develop appropriate expectations of those others in turn). We can apply this extended concept of expectation to one particular individual, or to a system of interacting individuals.

Applying the extended concept of belief as expectation to a particular individual would work something like this: instead of saying that the intentional actions of an individual can be understood in terms of the reasons he or she can give, we could say that an agent's responses can be understood in light of the expectations he or she has. An example here of a response

---

[45] The approach suggested as an alternative is not novel; see, for example, John Dewey's "The Reflex Arc Concept in Psychology" .

which we can make sense of in terms of expectations, but which is not a case of action, would be an involuntary response, such as the tensioning of one's ear muscles so as to pick up the expected final note of a piece of music being played. We can understand this response in terms of the expectations the person has, though his anticipatory physiological response cannot be said to be a case of acting for reasons. Yet it is an intelligible response; he is actively doing something: listening to the music and grasping the musical structure of the piece. We can understand it, even if we cannot verify it by "simulation", because we can understand that response as an attempt to pick up information we can appreciate as worth acquiring. How does this sort of response differ from other involuntary responses we would not want to say we can "make sense of" in terms of expectations, such as a particular person's body manufacturing an enzyme? The difference is that, in the first case, the individual can recognize when he has held an expectation that was wrong; he might say, for instance: "Oh, I was wrong about that -- I wasn't expecting that note." In the latter case, I mean to have picked an example of a kind of response such that there is no corresponding "disappointment" of an associated expectation such that the individual would, in retrospect, say that he had been wrong in what he had expected.[46] The case of becoming weak-kneed upon realizing one is in danger can be understood in terms of the person's expectations, although it is not a case of acting for reasons; it is intelligible as a case of recognizing a threat. Similarly, I have offered the example of a person's impaired performance on a math exam as something that can be understood in terms of a reluctance to take risks, due to the recognition of a threat that her actions will be perceived by someone employing a negative stereotype. Here the contrast case would be a case wherein someone has responses according to a regularity, but we do not find those responses intelligible in terms of his expectations. An example here would be a person's fumbling whenever he sees or thinks of the number thirteen.

---

[46] These limits are not biologically defined; responses that are involuntary at one point in a person's development can later become voluntary with training (such as biofeedback training). Just so, I am not ruling out the possibility that someone might develop an awareness of anticipations of which he or she is not presently aware.

Thus the notion of understanding an individual's behavior in terms of the expectations he or she has is not vacuous.

In applying the extended concept of belief as expectation to understanding the behavior of a collection of interacting individuals in a social or economic arrangement in terms of the expectations that are collectively attributed to them as a distribution, the notion of model-consistent expectations provides a constraint. Of course understanding the collective behavior involves another person (the econometrician) who is trying to make sense of it in order to predict the effects of different kinds of changes that might be made to the economic system. The constraint is a constraint on all the agents' beliefs collectively; since not all agents in an interactive arrangement have exactly the same anticipatory schemata, this notion of expectation applied to a collective of individuals will not pick out the expectations held by any given individual, unless we add restrictions about ways in which the agents must be similar to each other. However, the constraint on how an economy as a whole uses information to revise the expectations held collectively (distributed over its members in some way) does provide constraints of a sort on the individuals interacting in the economy and getting information from those interactions: they cannot all systematically ignore relevant information or all show the same bias in their expectations, for instance.

Whatever promise this line may or may not hold, the account of how beliefs are causal I have offered in this dissertation may, I hope, at least stir a few to think about how the sort of thing they say that belief is might be efficacious in <u>all</u> the ways we know beliefs to be so: as I put it earlier, we know beliefs can affect not only our decisions and our verbal behavior, but also our perception (at times even literally "blind" us to some fact or other), influence where our attention becomes focused, make us weak-kneed, make us blush, and make our hearts race. Beliefs and expectations affect other people with whom we interact, too. Our interactions are structured, for better or for worse, by the stereotypes and roles we employ as anticipatory schemata in social

interaction. When others perceive us perceiving them, they may perceive our beliefs and expectations on our faces and in our demeanor, every bit as much as in the questions we ask and the answers we give. I've shown how these social stereotypes enable interactions and social arrangements that would not be possible otherwise, and argued that the employment of such anticipatory schemata by individuals whose expectations form a system of mutual expectations of a certain sort is responsible for the establishment and maintenance of social institutions. All these are ways that beliefs make a difference.

||||||| The End |||||||

APPENDIX A

NINETEENTH CENTURY PRECEDENTS OF CURRENT ISSUES IN PSYCHOPHYSICS

Numerous questions and proposals that have arisen in the mid-to-late twentieth century regarding philosophy of mind are similar to questions and proposals that arose in the mid-to-late nineteenth century. One finds in these nineteenth century discussions explicit questions about the parallelism of mental and physical processes, mind-body interaction, the identity of physical and mental entities, and the reducibility of mental entities and processes to physical ones, as well as proposals of functionalist and behaviorist approaches to analyzing mental states. So, too, one can find there a forerunner of the kind of alternative to materialism I have sought in this dissertation. For, although the account I have developed derives from some specific twentieth-century accounts of active perception, the attraction of those accounts is that they employ an expanded view of science within which beliefs can be accomodated. And, the attraction of such an expanded account of science derives in part from a desire to avoid the kinds of problems that arise if one has a view of science on which each of physics and psychology has its own causal nexus and domain. Many of these problems were expounded in nineteenth century discussions in psycho-physics, as were suggested solutions.

The nineteenth century discussions I'm referring to arose in the context of the quantitative psychophysical research pioneered by Gustav Fechner and subsequently carried on by many others, including Mach, Hering, and Helmholtz, in various Experimental Philosophy laboratories in European universities. The nineteenth century discussions about the relation between the physical and the psychical were carried out in the service of providing foundations for, and interpreting the results of, experimental research that we would now refer to as psychophysical. But it's also notable that other areas of physics were faced with analogous questions: issues of reductionism and supervenience arose in thermodynamics, for which foundations in terms of

statistical mechanics (of molecules) were being developed, and in electrodynamics, for which foundations in terms of mechanical wave motion in a hypothetical ether were being developed. My interest here, though, is not in giving a historical narrative of these debates, but in identifying some helpful ideas in these nineteenth century analogues of late twentieth century philosophical discussions.

The nineteenth century experimental research into psychophysics gave rise to discussions about scientific foundations for psychology. In "Psychophysics and Mind-Body Relations" Fechner characterized the mental and the material worlds as two different viewpoints on one thing. He appealed to the fact that body and mind can never be observed together, and drew on an ancient analogy comparing the mind-body relation to the concave and convex "sides" of a circle, which, he said, "are basically only two different modes of appearance of the same matter from different standpoints." This, he said, is how "the mental and material worlds" are related (157). "What will appear to you as your mind from the internal standpoint, where you yourself are this mind, will, on the other hand, appear from the outside point of view as the material basis of this mind. [...] one is an inner, the other an outer point of view." Thus one can directly experience his own, but not another's, mental world: "One mind, insofar as it does not coincide with the other, becomes aware only of the other's material manifestations." The methodology employed in various disciplines reflected this difference: "The natural sciences employ consistently the external standpoint in their consideration, the humanities the internal." Fechner wanted to provide a basis for his empirical work, in spite of these limitations, which he regarded as fundamental. He considers Leibniz' analysis of the parallelism[47] of the body and the mind (i.e., that changes in one correspond to changes in the other) in terms of two clocks that keep the same time, and criticizes Leibniz for neglecting what he, Fechner, thinks is the analogue of the explanation of mind-body

---

[47] Michael Heidelberger has pointed out to me that whereas Leibniz used the word "parallelism" to mean occasionalism, Fechner and his contemporaries meant by it a more intimate relation than just having the same behavior.

parallelism : "They can keep time harmoniously --- indeed never differ --- because they are not really two different clocks" (158). From this, he concluded that "The truly basic empirical evidence for the whole of psychophysics can be sought only in the realm of outer psychophysics, inasmuch as it is only this part that is available to immediate experience. " However, he said, "the body's external world is functionally related to the mind only by the mediation of the body's internal world." That is, external stimuli do "not awaken our sensations directly, but only via the awakening of those bodily processes within us that stand in direct relation to sensation." And, we influence the "outer" world only via "the body's activities", which are controlled by our will. How this works causally, though, is unknown: "We thus have implicitly to interpolate everywhere the unknown intermediate link that is necessary to complete the chain of effects." Thus Fechner advocated a kind of monism that might today be labeled dual-aspect monism; his interest was in founding a science of measuring the mental: "Physical measurement yields a psychic measurement." He provided a picture on which his view rested, i.e., "The idea of the sensory side of the mind is actually based on the conception that there exists an exact connection between it and corporeality", though he allowed that that picture was not as yet justified: "Great doubt exists, however, as to whether each specific thought is tied to just as specific a process in the brain..."; and he argued for at least some sort of (what we would now call) supervenience: "If we now assume that the higher mental activities are really exempt from a specific relationship to physical processes, there would still be their general relationship, which may be granted to be real [...] and will, in any case, be subject to general laws..." (162). The analogy he then draws between higher mental states and physical bases for them is to the relationship between melodies and the ratios of vibrations that underlie the simpler sensations of tone, on which --- as some would put it today --- they supervene. The goal of this nineteenth century thinker's view was to provide a foundation for a science of psychology in which the same quantitative methods used in physics could be used to study psychology, or the "mental world". Although Fechner did not consider himself a materialist, many who today consider themselves materialists would, I believe, find the features of his view described above quite amenable.

The history of the views that subsequently arose in trying to provide a foundation for experimental psychology is rich and of interest in its own right. Here I focus on the philosophical views of one subsequent nineteenth century physicist-philosopher, Ernst Mach, to help in illuminating an alternative to twentieth century analogues of some nineteenth century views. Mach had done psychophysical research himself, and, as he explains in The Analysis of Sensations, he had attained a stability in his views only after studying both physics and the physiology of the senses alternately, and this only "after having endeavored in vain to settle the conflict by a physico-psychological monadology." He attributed some of the struggle to the work involved in freeing himself from preconceptions acquired in the study of physics: "With the valuable parts of physical theories we necessarily absorb a good dose of false metaphysics, which it is very difficult to sift out from what deserves to be preserved, especially when those theories have become very familiar to us." So it is not a matter of just relating the concepts in two different disciplines; for, he says, when "physics and psychology meet, the ideas held in the one domain prove to be untenable in the other. " He notes that research into the physiology of the senses had, at the time of his writing, taken on "an almost exclusively physical character." In spite of the many successes of physical science, this is to be resisted, for "physics . . . nevertheless constitutes but a portion of a larger collective body of knowledge . . . it is unable, with its limited intellectual implements, created for limited and special purposes, to exhaust all the subject-matter in question."

The general move I see in Mach is to consider a picture that embraces both the physical and the psychical. This is in fact a desideratum of his investigation: "any one who has in mind the gathering up of the sciences into a single whole, has to look for a conception to which he can hold in every department of science" (312). His complaint against looking to atoms as the means by which the connection between the physical and the psychical will be effected was that atoms and molecules were conceptions limited to applicability in the domains of physics and chemistry. He

did not deny that there were physical processes one might profitably investigate in physiological research.

Mach's picture of nature is meant to replace a picture of two distinct domains with a picture of "a single whole" that includes the physical and the psychical; "man himself is a fragment of nature." The distinction between the physical and the psychical becomes a practical, not a metaphysical, one. Thus there are no separate domains between which there exists a connection that must be explained. What we study is physics, he says, "when in searching into the connexions of the world of sense we leave our own body entirely out of account"(311).[48] Mach referred to the

---

[48] A similar characterization of physics is given in "The Economical Nature of Physics" in his Popular Scientific Lectures: "As long as, neglecting our own body, we employ ourselves with the interdependence of those groups of elements which, including men and animals, make up foreign bodies, we are physicists" (209). The example he immediately gives there as a physical investigation is that "we investigate the change of the red color of a body as produced by a change of illumination." In The Analysis of Sensations, he gave a similar example of how the green of a leaf changes color from green to brown under a change of illumination from the sun to a sodium flame. To Mach, this is the kind of connection investigated by physics (rather than by physiology). I think this is right, too; a bit of reflection shows that the impressive thing about our perception of coloured objects is the constancy of our perception of, say, the green of the leaf, throughout very varied changes in illumination. Thus, when Wilfrid Sellars, in his landmark "Empiricism and the Philosophy of Mind", discusses problems for sense-data theorists arising from a concept of "looks red to S" (as a matter of S's 'inner episodes') that must be parasitic on a concept of "is red", he is not attacking any view Mach held, but is actually cashing out one of Mach's insights. The examples Mach used to illustrate investigations into physiological psychology were not changes in lighting conditions, but ingesting a drug that made things appear differently to him than if he had not taken the drug, the stereoscopic vision we have in virtue of having two eyes ("Why Has Man Two Eyes?"), the fact that colors considered as sensations (i.e., in their psychical connections) disappear when we close our eyes, and phenomena such as the sensation that we are moving that suddenly arises upon staring down at a moving river for a while. Thus the element red, insofar as it is dependent on things like our eyes being open and what chemicals we may have ingested, is a sensation, but the changes of color produced by differences in illumination are physical phenomena. An illustration of spatial sensation Mach gives

pieces that made up this single whole as, simply, "elements", though he wished to avoid any metaphysical commitments, and suggested the concept of element be regarded as provisional.

---

is the fact that children often confuse the symbols "b" and "d", and the symbols "p" and "q", but never confuse the symbols "d" and "q"; hence optical similarity differs from geometric similarity (Analysis of Sensations 110).

Incidentally, the story Sellars gives as to how it could come about that our discourse involves talk of our thoughts ("inner episodes") could be seen as one answer (though not exactly the one I think Mach would give) to Mach's remark: "We read the thoughts of men in their acts and facial expressions without knowing how. Just as we predict the behavior of a magnetic needle placed near a current by imagining Ampere's swimmer in the current, similarly we predict in thought the acts and behavior of men by assuming sensations, feelings, and wills similar to our own connected with their bodies. What we here instinctively perform would appear to us as one of the subtlest achievements of science, . . . were it not that every child unconsciously accomplished it. [...] and here much is to be accomplished. A long sequence of facts is to be disclosed between the physics of expression and movement and feeling and thought".

And, in Mach's remark "We hear the question, 'But how is it possible to explain feeling by the motions of the atoms of the brain?' Certainly this will never be done . . . The problem is not a problem" (Popular Scientific Lectures 208), he is careful to restrict the negative point to a particular subdiscipline of physics, not to physics as a whole. This point is reflected in Sellars' remark that the suggestion that science will develop to the point where all the concepts of behavior theory would be definable in terms of theoretical physics is "either a truism or a mistake" (§61 "Empiricism and the Philosophy of Mind"). It is a mistake, Sellars says, if by "physical theory" we are talking of "theory adequate to explain the observable behavior of physical objects," whereas it is a truism if by "physical theory" we mean "theory adequate to account for the observable behavior of any object (including animals and persons) which has physical properties." Mach's characterization of physics was the study of "the interdependence of those groups of elements which, including men and animals, make up foreign bodies" (neglecting our own bodies), so his restriction of the negative point about the explanation of feelings to explanation "by the motions of atoms and the brain", rather than to the whole of physics, is very much like the distinction Sellars makes between the kind of physics for which it is a mistake to hope that behavior could someday be defined in terms of it, and the kind of physics for which it is a truism to say that the concepts of behavior could be defined in terms of it. I leave further development of this topic to another occasion.

Though he sometimes referred to these "elements" as "sensations", he also said that the word "sensation" was bound to be misleading, as it emphasized the psychical aspect of the elements of nature, whereas they are both physical and psychical. The distinction between the physical and the psychical arises only when distinguishing between the relationships in which they are conceived (i.e., physics or psychology). The point of talking about "elements" at all was more a deflationary move to dislodge allegiances to physical concepts such as matter than it was a constructive program proposing a new metaphysical category. The sensations are not "raw feels", but the result of analyzing the things we perceive. Hence, it is not that we perceive "greenness", say, and build up a notion of a leaf from it along with other raw feels such as "smoothness" and "oval shapedness". Rather, what we do do is perceive a leaf, and, after perceiving other leaves of yellow and red, or different hues of green, we come to develop the concept of green in distinguishing differently coloured bodies.[49] Mach prefaces his explanation that the elements he is talking about are "elements in the sense that no further resolution has as yet been made of them [...and ...] are the simplest materials out of which the physical, and also the psychological, world is built up" with the deflationary remark that "every physical concept means nothing but a certain definite kind of connexion of the sensory elements" (42). He wanted, I think, to put the use of concepts in psychological explanations, e.g., sensations, on a par ontologically with concepts the physicist often appeals to. Thus the principle of the parallelism

---

[49] The same view can be found in William James' The Principles of Psychology: "No one ever had a simple sensation by itself. Consciousness, from our natal day, is of a teeming multiplicity of objects and relations, and what we call simple sensations are results of discriminative attention, pushed often to a very high degree. [...] The only thing which psychology has a right to postulate at the outset is the fact of thinking itself, and that must first be taken up and analyzed." James complains that: "The notion that sensations, being the simplest things, are the first things to take up in psychology is one of these [apparently innocent suppositions that nevertheless contain a flaw]"; but he allows that, although they should not be taken "for granted at the start", one's investigations might show that sensations are "amongst the elements of the thinking." (Principles 219).

of the physical and the psychical is not simply a way of providing means to measure the psychical by the physical:

> The principle of which I am here making use goes further than the widespread general belief that a physical entity corresponds to every psychical entity and vice versa; it is much more specialized. The general belief in question has been proved to be correct in many cases, and may be held to be probably correct in all cases; it constitutes moreover the necessary presupposition of all exact research. At the same time the view here advocated is different from Fechner's conception of the physical and psychical as two different aspects of one and the same reality.[50] In the first place, our view has no metaphysical background, but corresponds only to the generalized expression of experiences. Again, we refuse to distinguish two different aspects of an unknown *tertium quid*. The elements given in experience, whose connexion we are investigating, are always the same, and are of only one nature, though they appear, according to the nature of the connexion, at one moment as physical and at another as psychical elements. (Analysis of Sensations 60 - 61)

I think what Mach is saying here about concepts that can be used throughout science is supposed to be very natural, and is analogous to something like this: When a rock rolls down a hill, it can be considered part of a mechanical system that can be analyzed by pure mechanics. When it is warmed by the sun, and absorbs or gives off heat, it can be considered part of a thermodynamical system. But we don't want to say that there are two rocks, a mechanical one and a thermal one, or, even, that there is a mechanical entity and a thermal entity that are aspects of a

---

50 Michael Heidelberger has suggested to me in conversation that here Mach is probably referring to the earlier works of Fechner, and that Fechner's view in the Elements of Psychophysics (which I have quoted above) is actually quite similar to Mach's view. I think this is quite likely.

third, unobservable, thing. We want a concept of rock that we can use in both narratives. The example he uses for physics and psychology is the green of a leaf; what is needed, he says, is a concept of green that is serviceable in both the physical and the psychological narratives.

This may seem, on the face of it, a rather superficial means of dealing with the issue of psycho-physical parallelism; to see the significance, and substance, of the view, consider some analogous advances in thermodynamics and electrodynamics. Mach explicitly referred the readers of The Analysis of Sensations to his treatment of measuring "states of heat"[51], in explaining his differences from Fechner regarding what Fechner called psychic measurements. In his foundational essays on optics and electrodynamics as well as on thermodynamics, Mach advocated unified views on which otherwise insoluble problems dissolved. Inasmuch as they were correct[52], they are helpful in suggesting what a workable alternative to physicalist views of man could look like.

---

51 Analysis of Sensations, p. 81n. See Michael Heidelberger (1993) for an explanation of Mach's remarks on measurement of "states of heat" and Mach's description of his differences from Fechner.

52 It is not my goal here to argue for the correctness of Mach's anti-reductionist views, but I do think that such arguments can be made. I have never understood the much publicized point that Mach never accepted the reality of atoms, for instance; it is clear that the atoms whose existence is now taken as established are not in fact atoms in the sense of being immutable and unchangeable. Mach criticized the presupposition that the ultimate elements, if there were any, were necessarily three dimensional, for he saw it as an illicit presupposition that the ultimate elements of nature were like physical objects located in three dimensional space (he thought it could turn out, as more became known, that atoms might be best conceived as either in an higher-dimensional space or in a space-time manifold). If anything, I think he has been vindicated on that point. There were other uses of atoms, too, that he objected to based on his psychophysical investigations, but the topic is beyond the scope of this discussion. I mean here only to point to Mach as offering a certain alternative to nineteenth century analogues of twentieth century discussions on the mental and the physical.

Consider the situation in optics and electrodynamics at the time Mach was writing this work on the foundations of psychophysics. Because light had many of the characteristics of mechanical waves, most physicists were attempting to give an account of light and electromagnetic waves as mechanical motion in a special light medium (sometimes called the luminiferous ether). As I explain elsewhere ("Sounds Like Light"), Mach explicitly freed the notion of light wave from the necessity of having a mechanical basis. Much later, Einstein would write in his memoirs that he thought Mach singular in advocating this specific kind of anti-reductionist view: he wrote that, except for Mach, "all physicists of the previous century saw in classical mechanics a firm and definite foundation for all physics, indeed for the whole of natural science, [...] they never grew tired in their attempts to base Maxwell's theory of electromagnetism [...] upon mechanics as well. Even Maxwell and H. Hertz [...] in their conscious thinking consistently held fast to mechanics as the confirmed basis of physics" (Autobiographical Notes 19). What Mach said about light and sound waves was that both had a number of features that we can consider characteristic of wave motion; these features were drawn from our experience with waves. Thus he identified a concept of wave that was serviceable for mechanics, optics, and electrodynamics. One feature was the constancy of the velocity of propagation; the velocity was independent of the motion of the source of the wave. Whereas, for sound, one can show that this feature can be deduced from mechanics of a material medium, Mach simply used this feature of light waves without any commitment to a medium, or any mechanical basis at all. He uses this in a proof that the Doppler effect (change in color of light observed as the source of the light is moving rapidly towards or away from one) is valid for light waves. Mach was, of course, correct in this: there is no mechanical medium for light, yet light has the feature of constant velocity in empty space, regardless of the motion of the light source, nonetheless. Rather than explaining light waves in terms of mechanical entities and forces, Mach expanded the picture of nature to allow for nonmechanical waves; in the framework subsequently developed in physics, an account can be given of both kinds of waves, as well as of how mechanical and electrodynamical forces and entities interact. He simply remained neutral about things for which there he thought there was

no empirical evidence, such as the luminiferous ether. I think it no coincidence that what Mach proposed, in response to the conundrums of psychophysical research, was an alternative to attempts at reducing psychology to physics. Not only is Mach's alternative anti-reductionist, but it has the consequence of extending a picture of nature to embrace both fields, and allows other species of causation than physical causation. Proposing a view of nature in which not all causation need be physical causation is analogous to the move he made when he proposed a view of physics that allowed for species of causation other than mechanical causation.

The issue in thermodynamics is less straightforward. Mach is often identified as an anti-reductionist and thus as an obstacle to the progress and acceptance of the kinetic theory of gases. However, here, too, Mach displayed a unifying approach that had the consequence of expanding the field of applicability of formal principles and concepts beyond simple mechanics, to embrace both thermal and mechanical processes. He complained of attempts to give a basis for the principle of energy from which the principle of the excluded *perpetuum mobile* (which is equivalent to the second law of thermodynamics) could be deduced. Again, Mach's approach was to be noncommittal as to what types of processes and causation could occur, and to look to experience for principles: 'It is only from experience that we can know whether and how thermal processes are connected with mechanical ones." He credits Carnot with being the first "to exclude the *perpetuum mobile* in a wider domain than that of pure mechanics". This extension of a principle was necessary, Mach said, "to enable the modern principle of energy to appear." (Theory of Heat 297-298). The "modern" principle of energy was one that could be used in analyzing all sorts of physical processes, not just mechanical ones. Certainly we can now talk about "thermal energy" and about the "mechanical equivalent of heat", but being able to do so meaningfully does not mean that there are two different causal nexuses, one for thermodynamics and one for mechanics. With the new concept of energy comes a view of physics in which these are located in one single causal nexus.

Now, the example in thermodynamics of conceiving of temperature in terms of mechanical motion of molecules is often cited as an example, not of unification of domains, but of reduction of one branch of physics to another. However, the program of reducing thermodynamics to statistical mechanics has not turned out to be as simple as a reduction. The formulation of the foundations of thermodynamics is still a topic of controversy, and none can disagree that thermodynamics is not a simple consequence of the laws of pure mechanics; the disagreements are over what amendments to mechanics are required. Whether one can regard the amendments as merely special features of the reduction, or whether one thinks such amendments mean the relationship does not really count as a case of reduction, is a matter of what is meant by reduction. A related issue arises when there is a change in concepts occasioned by working out the relationship between the two disciplines. Larry Sklar, writing about reductionist accounts that put forward the reduction of thermodynamics to statistical mechanics as an exemplar, says:

> "Given the radical reinterpretation of thermal phenomena forced upon us by the joint theories of the atomic constitution of matter and statistical mechanics, it is not surprising that we will find much to say about just these issues when we discuss the details of the relationship of thermodynamics to statistical mechanics. So complex is the situation that we will be a bit surprised that people still say things like, 'The identification of mental processes with brain processes is no more mysterious than the identification of temperature as mean kinetic energy of molecules,' as if that latter 'identification' (if that is indeed what it is) were as straightforward as all that (Physics and Chance 340).

But, the most spectacular analogue of Mach's approach to finding "elements" that can be used in both physics and psychology was his requirement that there ought to be a notion of mass that could be used for both inertial mass and gravitational mass. The suggestion that a "truly reasonable" theory would use the same concept of mass in both mechanics and gravitational theories is what Mach is most remembered for in philosophy of physics, and, some even say, it

was this desideratum of Mach's that Einstein used as a guiding principle in developing the general theory of relativity.[53]

Thus Mach's emphasis on "elements" that can be regarded as sensations when considered in psychological relations is not the one-sided approach I think it has often been made out to be. In light of the rather spectacular resolutions of conundrums achieved by employing the anti-reductionist, unifying approaches to physics mentioned above, Mach's approach to nature (of which "man is a fragment") can be appreciated as a more substantial suggestion. It is proposed as a solution to dualism: "Now if we resolve the whole material world into elements which at the same time are also elements of the psychical world and, as such, are commonly called sensations; if, further, we regard it as the sole task of science to inquire into the connexion and combination of these elements, which are of the same nature in all departments, and into their mutual dependence on one another; we may then reasonably expect to build a unified monistic structure upon this conception, and thus to get rid of the distressing confusions of dualism" (312). That this suggestion may have been mistaken by others (e.g., the so-called "sense-datum theorists") to lend support to programs we now know to be blind alleys is no reason to disparage it.

Mach presents his view that the whole world is capable of resolution into elements that are both physical and psychical not only as a prescription for the ills arising from dualism, but as a diagnosis of how the dualism arose in the first place: "Indeed, it is by regarding matter as something absolutely stable and immutable that we actually destroy the connexion between physics and physiology" (312). This should strike many contemporary philosophers of mind as an unusual diagnosis: materialism is often seen as the antidote for dualism, not the cause of it.

---

53 To be fair, it should be noted that not all agree on the importance of the so-called "Mach's principle" to the development of general relativity, e.g., John Earman argues in World Enough and Space-Time (84) that Mach's contribution has been vastly overstated.

But it is not my aim here to add to the already large literature on the possibility of reduction of the mental to the physical. I am refraining in part because I think reductionism is not a yes-or-no question, but requires a that-depends-what-you-mean-by-it answer. But it is also true that I do not see any reason to pry the picture of inter-theoretic reduction from those who feel comfortable framing their investigations in terms of it. If I am right, what investigators aiming to eliminate beliefs in favor of material/physiological concepts of the mind, or to reduce beliefs to physiology, will actually find is that, in doing so, they will be revising their notions of physiology.[54] That a reductionist program might have the consequence of extending the reducing theory beyond the domain it covered when the task of reduction was first conceived, is not by any means unprecedented. Even in the reductionist's favorite example, the reduction of temperature to molecular motion, we see physics change in character as the reduction is actually carried out. For, as statistical mechanics developed, concordant with attempts to understand thermodynamics in terms of statistical mechanics, it also changed. Sklar remarks:

> For the first time, statistical mechanics introduced into physics the idea that the aim of a physical theory could be not to provide an account of what must happen, but of what might happen. [...]
>
> The probabilistic theory differs from the earlier theories, as we have noted in detail, not only in its explanatory aims but in its basic posits. (Physics and Chance 347)

---

54 What I expect is thus something like the inverse of what Paul Churchland says the eliminative materialist expects. For, he says, "the eliminative materialist is also pessimistic about the prospects for reduction, but his reason is that folk psychology is a radically inadequate account of our internal activities, too confused and too defective to win survival through intertheoretic reduction. On his view it will simply be displaced by a better theory of those activities." In "Eliminative Materialism and the Propositional Attitudes", in Mind and Cognition, edited by William G. Lycan (Blackwell: 1990), p. 209ff., originally published in The Journal of Philosophy 78 (1981), pp. 67 - 90.

Thus, I do not think it fruitful to make the possibility or impossibility of reductionism of the mental to the physical the issue. What I expect is that an investigation into the kind of physics to which beliefs can be reduced will, if all the causal roles of belief are considered, force one to expand physics to include a physiology that involves more than most reductionists originally would have thought necessary. I expect attempts at finding a natural science to which the mental can be reduced will result in a concept of natural science that embraces belief. It is not a program of reductionism per se, but, rather, a prior commitment to what the reducing theory must include, that I consider objectionable. This digression into nineteenth century philosophy of science[55] is meant to point out that the presupposition of physicalism is just that: a presupposition. There is nothing particularly scientific about it, nor need we be anti-scientific to relinquish it.

---

[55] Michael Heidelberger has commented that, long before Fodor and Putnam ever did, "Fechner held it theoretically possible that mental states could be realized by systems other than the brain" (1993).

APPENDIX B

THE PLACEBO EFFECT AND CONTEMPORARY PHILOSOPHY OF MIND

1. The Placebo Effect And Intentional Action

The import of requiring that a theory of mind account for the first way for beliefs to be efficacious identified in Chapter II is that it recognize the causal role of belief in causing intentional action. And this requires recognizing that beliefs are related to actions by the operation of rationality -- i.e., in forming and responding to reasons. This much is not novel. But, in addition, a theory of mind should include the third mode of causation: beliefs can cause movements and physical changes not mediated by intentional action.

There is an ambiguity here: In saying that beliefs cause movements and physical changes not mediated by intentional action, are we dealing with the same notion of belief used in saying that beliefs cause intentional action? In the kinds of causal explanations in which belief causes intentional action, it is as an intentionally characterized entity that the belief is involved in the causal explanation. Such explanations appeal to the thematic relation between a belief and the actions it caused. The same is true for at least some instances of this third mode of efficacy for belief (i.e., efficacy that is unmediated by intentional action). Consider instances of the placebo effect in which it is a belief that gives rise to a physiological change. Here is a real-life case: Suppose someone becomes convinced that rubbing salt on warts will cure them. Then it is important for the explanation in which the belief causes the effects it does that the belief is intentionally characterized as being about warts and salt and the curative power of salt on warts. Here we can recognize the same belief figuring in the causal explanation in both modes one (causing intentional action via the operation of rationality) and three (causing physical changes unmediated by intentional action). That is, the very same belief should explain both the person's

intentional action of procuring salt and rubbing it on warts, and (perhaps along with some other beliefs) the physical effect of causing the warts to disappear.[56]

In such explanations of so-called placebo effects, the cause and effect are thematically related: the causal relation is between a belief that is about salt and warts, and the effect that the warts that have been rubbed with salt have shrunk. Thus, accounting for the third way in which beliefs can be efficacious (as causal in effecting physical changes unmediated by intentional actions of a rational agent) requires recognizing a special kind of explanation that appeals to the thematic content of the belief--- to the belief as an intentionally characterized entity --- to explain a thematically related effect. The reason I say it is a special kind of explanation is that the explanation can't be one that appeals to the rationality of the believer, because it's not a case of intentional action, and yet (on the double-nexus picture most contemporary accounts employ) it can't appeal to physical causation, either, because it's the belief as intentionally characterized that figures in the causal explanation; it is relevant to the explanation that the belief be about what it is about and be described as it is, and a physical explanation[57] is blind to such features of belief.

Here's where I think a hitherto unrecognized task for theories of mind arises: having a theory of mind which can account for the same belief figuring in both kinds of causal explanations. There are accounts of belief that take seriously the task of accounting for the operation of rationality, so that intentional behavior is not left as a mysterious phenomenon. But, they do not undertake the

---

[56] This example of efficacious belief is modified from Wesley Salmon's Logic. Salmon mentions a treatment of rubbing onions on warts. The specific treatment is not important; it is well known that almost any arbitrarily chosen treatment a patient can be convinced is efficacious will in fact result in the disappearance of warts. From this it is inferred that it is the patient's belief in the treatment that is efficacious.

[57] I am using 'physical' here in the sense in which we speak of the physical sciences: biology, chemistry, physics, and so on, and I use "physical laws" and "physical causation" to mean whatever materialists intend when they refer to material laws and causation.

task of explaining the thematic connection between belief and physiological effect that occurs in phenomena such as the placebo effect. There are theories of mind that take seriously the task of accounting for the occurrence of physical effects of belief, but do so at the expense of substituting some kind of ersatz rationality for genuine rationality.

There is another position one can hold, which I neither advocate nor contradict: to accept that beliefs are efficacious in such so-called placebo cases, without having to explain why. The considerations I've given above for why the placebo effect poses a challenge for a theory of mind do not rule out such a response; after all, it is only a lacuna in explanation, not a contradiction, that I have identified. Someone who is interested in the constraints rationality puts on behavior and accepts the existence of nonrational behavior but does not seek explanations for it, might comfortably ignore the lacuna. What I have pointed out need not trouble one with this view; I have pointed it out for the sake of those who have different aspirations than such a person: for the sake of those who would be bothered by such a lacuna were they convinced it existed. It does seem to me that many who are devoted to physicalist and materialist accounts of mind often think there is no challenge in filling in the lacuna; the placebo effect is too often set aside by a few remarks meant to convince the reader that there is no worry about placebo effects that can not be easily met.

2. The Background Picture: Two Kinds of Explanations of Mental Causation

I want to clarify the consequences of accepting the task posed by the placebo effect. Many theories of mind are developed with the following general picture in the background: There are two kinds of explanations involved in explaining mental causation: (i) psychological explanations, in which mental entities, which are intentionally characterized, play a causal role (although this is sometimes considered 'as-if' causation), and (ii) physical explanation, in which only physically-specified entities can play a causal role. In psychological explanations, both the cause and effect

figure in the explanation as intentionally characterized entities (recall the discussion on explanations that appeal to the operation of rationality, in which it was important to the explanation that the belief be about what it was about, and that both the belief and the action be described as they are). In explanations of physical cause and effect, though, both the cause and the effect figure in the explanation as physically-specified entities. The explanation given when beliefs have physical effects is that the beliefs are identical with, or correlated with, or instantiated by, material entities; then, the story goes, there is some explanation to be had in which it is as a material entity that the belief figures in the explanation under which the belief produces physical effects according to physical laws. Such an explanation is generally just presumed to be available, given enough research. Another view is that beliefs and other so-called mental processes are merely epiphenomena of material processes; here, the only things deemed causal are the material processes.

Although the picture of two nexuses, one of which is a redescription of (or otherwise supervenient on) the other, has its puzzles and challenges, people still seem to hang on to it, and as long as only cases of intentional action are considered, it still seems servicable to people of various views. For, in cases of rational behavior, we can see how the story, even though it has some problems, could go: Just as an agent uses his knowledge of how things go in the physical world to carry out his intentions, there is an appeal to evolution having designed physiological processes appropriately; here it is Nature that hooks into material processes to carry out its intentions[58] of having the relations between intentionally characterized states operate according to what rationality would require. Or, alternatively, Nature allows some slack to be taken up by learning, and merely provides a trainable organism; then we get a more detailed story as to how the organism's experience and/or socialization is involved. In the case of beliefs having effects unrelated to their intentional content, we can see how the story could go, too: although a belief,

---

[58] Of course I use "intentions" tongue-in-cheek here.

intentionally described, is in the mental nexus, it has some corresponding entity in the material nexus, and this can cause side effects.

Here's what I mean by side effects: given the picture that there is a chain of causation in the physical nexus that begins with the physical entity corresponding to the belief as intentionally characterized and ends in a physical entity corresponding to its intentionally characterized effects, there are some physical effects that the physical entity has to cause[59] in order to bring about the physical effects corresponding to the intentionally characterized effects. But, the physical entity corresponding to the belief may have other physical effects, too, besides those that correspond to the intentionally characterized effects of the belief that result from a rational agent having that belief. But, as the physical entity is not inherently about anything, there's no reason these effects should have anything to do with the belief as intentionally characterized; any thematic connection between these side effects and the belief as intentionally characterized would be unexplained. That is, we would require further explanation for why there was any thematic connection; neither the rationality of the agent nor physical science, nor the supervenience of rational processes on physical ones, accounts for such a connection.

Accounting for rationality is an aim of most theories of mind. In fact, it is hard to overstate the attention philosophers have paid to rationality; it's often treated as what's involved in mental activity. And most of the challenge for theories of mind comes from wanting to simultaneously account for laws of rationality and laws of physics governing organisms. Trying to account for both seems to give rise to the above mentioned apparently-double set of networks of causation.

I said earlier that this framework seems serviceable to most people as long as only intentional behavior is being considered, for thinkers of various persuasions have chosen to treat the cases of beliefs figuring in causal explanations of intentional behavior within such a picture. But, if cases

---

[59] Or, rather, it is in virtue of having such effects that it has been "selected".

such as the placebo effect are also considered, the picture becomes less serviceable. For here, we saw, the effect to be explained is neither a contingent side effect of a material state the belief happens to be or correspond to, nor is the relation between cause and effect governed by the operation of rationality. So, the story given to explain why the order of material processes mirrors, or is mirrored by, rational processes, provides no explanation of the placebo effect case. Someone could say, as I believe Davidson does, that there is no general explanation available of the rational processes in terms of the physical ones, and so the placebo effect case is not qualitatively different from the case of rational intentional action. My remarks here are directed at those who think that there is such a general explanation.

Now, perhaps there is no principled reason why one could not retain the framework of two kinds of causation and incorporate enough complexity to account for such processes as well. But then the picture changes drastically: the causal relations between beliefs and actions were defined in terms of being governed by rationality; there would have to be relations between beliefs and effects in which the beliefs were clearly causal as intentionally characterized entities, yet which were not governed by rationality. It seems the appeal of a theory would diminish as such additions were made, in that, instead of the one principle of the rationality of intentional behavior, we add on other principles to account for phenomena. It's hard to see how the kind of account of the basis for rational behavior someone would want to give would be the kind of account they'd want to give for the placebo effect case.

So, the import of including the third mode of efficacy in addition to the first is to question the wisdom of a basic approach to developing a theory of mind that seems a starting point of current theories: the existence of two causal nexuses.

3. Late Twentieth-Century Accounts of Mind -- Davidson, Dennett, and Dretske

To support and illustrate my earlier claim that such theories of mind won't be able to satisfactorily account for all the different causal roles of belief I've mentioned, we will examine several specific thinkers' works on the efficacy of belief: those of Donald Davidson, Daniel Dennett, and Fred Dretske.

a. Donald Davidson on Reasons as Causes of Actions

In his well-known "Actions, Reasons, and Causes", Davidson is concerned with defending what he calls "the ancient -- and commonsense --- position that rationalization is a species of causal explanation" (3). By rationalization he means an explanation of an action in terms of "the agent's reason for doing what he did." Davidson is responding to the conclusion some philosophers have drawn that "the concept of cause that applies elsewhere cannot apply to the relation between reasons and actions, and that the pattern of justification provides, in the case of reasons, the required explanation." He calls his position anomalous monism, using a notion of cause taken from Hume on which "A caused B" entails only that there exists a causal law instantiated by some true descriptions of A and B (but not that "A caused B" entails some particular law involving the predicates used in the descriptions 'A' and 'B'). Thus he is not appealing to any feature about the mental; what he says about cause here is supposed to be true of most causal explanations, and he is only showing that, on this notion of cause, reasons can be causes. This notion of cause fits, for instance, the weather: "The trouble with predicting the weather is that the descriptions under which events interest us -- 'a cool, cloudy day with rain in the afternoon' --- have only remote connections with the concepts employed by the more precise known laws" (17). Thus, though there will be some causal explanation under which there are lawlike connections, laws connecting reasons and actions are on a par with the causal laws that underlie descriptions of the weather: "The laws whose existence is required if reasons are causes of actions do not, we may be sure,

deal in the concepts in which rationalizations must deal." In fact, even if a true generalization connecting a class of actions and a class of reasons does exist, Davidson says the generalization wouldn't constitute a law connecting reasons and actions, for "If the causes of a class of events (actions) fall in a certain class (reasons) and there is a law to back each singular causal statement, it does not follow that there is any law connecting events classifed as reasons with events classified as actions -- the classifications may even be neurological, chemical, or physical" (17). Thus, Davidson has argued for reasons as causes, but reasons as causes in virtue of laws that do not connect reasons and causes. I.e., reasons are causes in virtue of being identical with some other entity or event: "neurological, chemical, or physical."

This seems to recall the picture Fechner used: the identity of every psychical event with a physical event. However, unlike Fechner, Davidson is not interested in finding psychophysical laws. Recall that Fechner said that even if there was no direct identification of a specific thought with a specific process in the brain, "there would still be [the general relationship between the higher mental activities and physical processes], which may be granted to be real [...] and will, in any case, be subject to general laws...." That there will be such general laws is just what Davidson denies. Davidson is, however, retaining a picture of two "worlds", though he considers them to be redescriptions of one another, rather than two ontologically different domains that happen to be correlated. In later essays, he argues for his view that the causal laws in virtue of which reasons are causes will not be stateable in terms of the concepts used in explanations connecting reasons and actions: In "The Material Mind", he says he is trying to show that "we cannot establish general, precise, and lawlike correlations between physical and psychological descriptions" (255). In "Mental Events," he sums up the situation this way:

> There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and

> conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence. [. . .] nomological slack between the mental and the physical is essential as long as we conceive of man as a rational animal. (222-223)

This point seems to be in accord with Mach's statement that 'when physics and psychology meet, the ideas held in the one domain prove to be untenable in the other." But, whereas Mach seemed to take this as a sign that new concepts were needed in order to reunite physics and psychology, Davidson seems to regard the disunity as inevitable, and such incompatibilities as indicative of the impossibility of that kind of unification. It's crucial to Davidson's view that beliefs and attitudes are a matter of one person attributing them to another based on an evolving understanding of the other individual; "the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory" (223). Thus the theory is not only lacking in generality (as it is developed on the basis of understanding a particular individual), but it is always provisional and subject to change (223). When Davidson concludes that "nomological slack between the mental and the physical is essential as long as we conceive of man as a rational animal" he is implicitly assuming that our understanding of human physiology is not, likewise, evolving in light of our understanding of individuals. Davidson's notion of physiology seems to be rather rigid: "Physical theory promises to provide a comprehensive closed system guaranteed to yield a standardized, unique description of every physical event couched in a vocabulary amenable to law" (223). But human physiology is hardly a subspecialty of such a physical theory. There is an unexamined presupposition here about what kind of science physiology is. Davidson's line of reasoning here seems an illustration of Mach's comment that "it is actually when we consider atoms as immutable that . . . there is a gap between the physical and the psychical." As Mach thought our understanding of Nature would always be provisional, he certainly would

have thought that our knowledge of human physiology would evolve as we understood more about psychology. It seems to me that, if we would allow that an understanding of human physiology is of the same character as Davidson says psychology is -- i.e., it involves interpreting an individual person's behavior (physiologically characterized) in light of his beliefs, moods, desires, attitudes, and so on --- the "nomological slack", as he called it, would no longer be essential.

b. Daniel Dennett on The Intentional Stance

Daniel Dennett's collection of essays in The Intentional Stance might be viewed as giving an account of how things could work in the way Davidson says they must. It was the constitutive ideal of rationality that Davidson said ensured that there could not be psychophysical laws -- laws, that is, connecting reasons and causes. Yet, Davidson said, the explanations of actions in terms of reasons could still be causal explanations, because it could still be the case that a particular mental event that was a reason and a particular mental event that was an action could instantiate a physical law in virtue of physical descriptions of them. Thus, it was important to Davidson's overall picture of things that psychological events and entities could in principle be redescribed in physical terms. Dennett gives a picture of persons in which (he claims) all this is true; it's a positive account in terms of levels of description inspired by the disciplines of control systems analysis and computer science. It follows the tri-level approach David Marr employed in Vision: A Computational Investigation into the Human Representation and Processing of Visual Information; Marr identified the three levels: the level of computation theory, the level of representation and algorithm, and the level of hardware implementation (25).

Dennett applies a similar tri-level approach to the mind: at the top level is a theory of mind that plays the role that folk psychology (the kind of understanding of each other we use in everyday interactions and conversations) should, at its best: intentional system psychology. Intentional

system psychology is supposed to be like folk psychology in that it predicts and explains human behavior in terms such as "belief" and "desire". Dennett defines beliefs in terms of their role in predicting and explaining the behavior of a system: ". . . *all there is* to really and truly believing that *p* (for any proposition *p*) is being an intentional system for which *p* occurs as a belief in the best (most predictive) interpretation" ("True Believers" 29). He describes intentional system psychology as a "sort of holistic logical behaviorism"; belief and desire are attributed only in the context of a theory of a whole system:

> [Intentional system psychology] is a sort of holistic logical behaviorism because it deals with the prediction and explanation from belief-desire profiles of the actions of whole systems (either alone in environments or in interaction with other intentional systems), but it treats the individual realizations of the systems as black boxes. ("Three Kinds of Intentional Psychology" 58)

Being an intentional system is a matter of being regarded as an intentional system. What makes something an intentional system is that we adopt what Dennett calls the "intentional stance" toward it. In his words:

> Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many---but not all---instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do. ("Three Kinds" 17)

The attribution of desires and beliefs to a system involves figuring out normative demands on it: i.e., the desires and beliefs the agent ought to have. Ought to have for what? Dennett's answer is, ought to have to do its best to satisfy its biological needs, of which survive is one (49). The specific answer one might give, however, is not important to the plausibility of such an overall scheme; other answers might work just as well. For, at the level of intentional systems psychology, what beliefs may correspond to in the system's makeup does not constrain the attributions of belief made by someone taking the intentional stance.

What is of interest about beliefs is that they serve as premises in the rational agent's reasoning, and can be derived or deduced from other states of the system ("True Believers" 30). Now, are these entities called beliefs also characteristic of the concrete thing that realizes the system, whatever it is that is inside the "black box" of the system? Well, Dennett says that in one sense they are and in one sense they aren't. They are "objectively there" in the sense that patterns are objectively there for someone who has an interest in noticing them ("Reflections: Real Patterns, Deeper Facts, and Empty Questions" 37). But in another sense they aren't a feature of the concrete realization of the system, because the beliefs attributed by the intentional stance might not correspond to anything that can be concretely characterized. Dennett thinks that in fact not all of them can be concretely characterized, for a person can be said to have literally an infinity of beliefs, and it seems impossible to individuate all the beliefs a person holds, anyway. So, Dennett suggests that most beliefs will be implicitly stored. Some "core elements," by means of which all the beliefs are implicitly represented, will be concretely characterizable for a specific instantiated system, but, he says, there is no reason to think that these will themselves represent any beliefs ("Three Kinds" 55-56).

But this is jumping ahead over the mid-level characterization of the system, to the specific concrete instantiation of the system. Intentional system psychology treats the organism as a system characterized as having beliefs and desires appropriate to its biological needs, and as

acting rationally. The next level down is called sub-personal psychology. It's still psychology because it's about intentionally characterized entities, but it's sub-personal because the subjects processing these entities---recognizing and responding to them---are subsystems of the intentional system. In choosing these entities at the sub-personal level, one doesn't lose sight of the environment, though: "the theorist must always keep glancing outside the system, to see what normally produces the configuration he is describing, what effects the system's responses normally have on the environment, and what benefit normally accrues to the whole system from this activity" (63-64). This sounds very promising; it seems like a way to keep the intentional in view as we descend to more detailed descriptions.

What comes next is "exhibiting how the brain implements the intentionally characterized performance specifications of sub-personal theories." Here's where the promising becomes less so, as it has to, for, as Dennett recognizes, to "get semantics from syntax" is an impossible task. However, the brain could, he says, "be designed to approximate the impossible task, to mimic the behavior of the impossible object (the semantic engine) by capitalizing on close (close enough) fortuitous correspondences between structural regularities---of the environment and of its own internal states and operations---and semantic types" (61). This problem is like that faced by someone trying to design a program to discriminate the meanings of written messages:

> You must put together a bag of tricks and hope nature will be kind enough to let your device get by. Of course some tricks are elegant and appeal to deep principles of organization, but in the end all one can hope to produce (all natural selection can have produced) are systems that seem to discriminate meanings by actually discriminating things (tokens of no doubt wildly disjunctive types) that co-vary reliably with meanings. ("Three Kinds" 62-63)

So the intentional characterization of the function is important in that it is what the concrete realization is supposed to mimic. Dennett provides an example: a creature may need to know when it has reached the goal of receiving nourishment. It might accomplish this with a subsystem that detects instead the mechanical signal "friction-in-the-throat-followed-by-stretched-stomach," which then causes the behavior of ceasing eating for the moment. Now, this answer, though certainly more plausible for states such as receiving nourishment than ones such as recognizing a situation as one in which discretion is called for and which then leads, via material processes, to discretionary behavior, does provide an answer---in principle---to the question of how an organism might behave in ways that reason would prescribe.

But it's at a cost: it's still not rationality that is being instantiated, but a mime of it. It's not the intentionally characterized belief that causes individual behavior, but a materially characterized "core element" which causes, according to physical laws, another materially characterized state of the thing, which in turn causes behavior which can be intentionally characterized. So, as in Davidson's picture, it is not rationality that is causally operative. I want to be clear here about what my complaint is: that a substitute for rationality has been slipped into the explanation of intentional action. And, being a substitute, or imitation, of rationality, it results in different causal connections than those that obtain between reasons and actions. Davidson provided a picture to show that a reason could be a cause, even if only physical laws could be causal laws. Dennett fills in more -- enough to explain how something we regard as rationality could occur. But it is not an explanation of rationality.

It is a commitment to material causation as the only kind of causation that can be countenanced that sets up this tortuous explanation of rational behavior. (Dennett makes no bones about this: "I declare my starting point to be the objective, materialistic, third-person world of the physical sciences" (The Intentional Stance 5). As Dennett explicitly set out to explain rational intentional behavior, it is no surprise that he has an account of the causal role of belief for the case of beliefs

causing intentional behavior via being part of a reason for an action (although it is not an account of rationality being operative). Rather, Dennett has explained how ersatz rationality can be grounded in material causation, and suggests that what we are doing when we take the intentional stance is using a theory of rationality (intentional system psychology) to predict the behavior of creatures who are, in fact, governed by ersatz rationality. It seems to me, though, that Dennett's account amounts to relinquishing a claim one might want to keep: that persons are actually acting rationally in cases of intentional action, that reasons are causes; i.e., that there is such a thing as genuine rationality operating.

As for placebo effects, Dennett does recognize, and wants to account for, the phenomenon that belief can have physiological effects; he mentions blushes, verbal slips, and heart attacks. Here, though, it is not really supposed to be the belief that is efficacious:

> In such an eventuality [we discover that the core elements do not explicitly represent beliefs] what could we say about the causal roles we assign ordinarily to beliefs (e.g., 'Her belief that John knew her secret caused her to blush')? We could say that whatever the core elements were in virtue of which she virtually believed that John knew her secret, they, the core elements, played a direct causal role (somehow) in triggering the blushing response. ("Three Kinds" 56)

Here there is outright rejection that it is the belief that is really causal.[60] The mechanism (between the core element and the blushing) is supposed to be governed by physical laws, and it is just a

---

[60] It is not clear to me that this really is a case of belief causing a physiological effect without involving rationality. Another way to see this case is: the belief results in an emotional reaction; there's a rational connection between the belief that someone knows your secret and the emotion of feeling embarrassed. The blushing might then be a concomitant of the emotional state. I'm discussing it as though the blushing is not rationally connected because that's Dennett's assumption in discussing the phenomenon.

contingency of nature that the material instantiation of the state of the machine in virtue of which the person has this belief has the physical effect of causing blushing. Blushing, verbal slips, and heart attacks are not intentional behaviors governed by rationality; according to Dennett, they just happen to be physiological side effects. These so-called "side" effects, it seems, could differ from individual to individual, since beliefs are (implicitly) represented differently in different individuals.

I don't find this explanation satisfying. To explain my dissatisfaction, consider other phenomena besides blushing in which belief is causal, but not via rational behavior: the placebo effect and other physiological effects due to beliefs.

Let's grant that, at least in some cases of a subject receiving a dummy medication, the belief that a person is receiving something that is physically efficacious is responsible for some physiological change that person undergoes during treatment. Now, for Dennett, it is the concrete, physical state of the machinery---the design of which exploits physical laws to produce rational behavior---that is really causal. In intentional system psychology, the intentional characterization of the belief is connected to other intentionally characterized states of the machine (such as other beliefs and desires, and actions it takes) via rationality. However, at the concrete level, where physical causation is operative, we have the "core elements," which are not intrinsically about one's physiology or about medically efficacious substances or knowledgable physicians. Yet, a contingent side effect of the core elements in virtue of which one has the belief that a certain physiological effect will result is that that particular physiological effect does result. The explanation for the thematic relation between rationally connected states invokes natural selection: since such connections contribute to the organism's survival, the story goes, natural selection accounts for the evolution of organisms with internal processes that provide connections between internal states that mimic rational behavior. But the placebo effect could

hardly be explained by a process that imitates rationality, for the placebo effect is not a case of rational intentional behavior.

Could one, nevertheless, give some other account of these connections in terms of natural selection? One might appeal to the advantage of having connections between core elements and physiological states such that mere suggestion can heal, to argue that humans have evolved some such causal connections as well, in addition to those required to mimic rationality. If this could be done, Dennett's theory of mind would provide an explanation of, rather than merely allow the possibility of, the placebo effect. But it's implausible that this could be done, because it seems there are just as many disadvantageous cases---such as that, as a result of believing he's been harmed, one exhibits symptoms of being harmed. And there's the case I mentioned before of someone making a mistake as a result of being preoccupied with the thought of how disastrous the consequences would be if he made it. Even if these problems could be handled, it means straying from the view that what beliefs, as intentionally characterized entities, *are*, is determined by their role in a system regarded as a rational agent with a belief-desire profile.

Now, it seems to me that this raises a problem for what we might call instrumentalist accounts of belief such as Dennett's. Dennett's view is impervious to the earlier complaint I made, that the only kind of rationality that turns out to be causal is an imitation of rationality, for that is only objectionable should we not want to give up the idea of genuine rationality being operative in intentional action. But I don't think his view is impervious to the complaint arising from the requirement that an account of belief explain all the ways in which beliefs are causal.

Dennett calls beliefs abstracta, after Reichenbach's distinctions between abstracta and illata: abstracta are calculation-bound entities or logical constructs, and illata are the posited theoretical entities. Beliefs, being abstracta, have a reality on a par with the earth's Equator (53). Thus, Dennett (justifiably, I think) writes that "[it] is not particularly to the point to argue against me that

folk psychology is in fact committed to beliefs and desires as distinguishable, causally interacting illata; what must be shown is that it ought to be" (54). Dennett's account is likewise impervious to the complaint that his approach denies that beliefs and thoughts, intentionally characterized, are what's causal in cases such as the placebo effect, for he can argue that beliefs are abstracta in an account of beliefs being efficacious in placebo-style cases as well. However, although he can easily shrug off each of these complaints singly by saying that on his account beliefs are "abstracta", this isn't a valid response for them jointly. For, once it is appreciated that beliefs show up in more than one kind of causal explanation, the basis for saying they are abstracta --- that all there is to being a belief is its role in explaining intentional behavior[61] --- no longer holds. An historical precedent in philosophy of science here is the change in status of the molecule in physical science: Salmon argues that the fact that the molecule figured in explanations of disparate kinds of phenomena, e.g., X-ray diffraction and Brownian motion, lent credence to the molecules being real entities. However, I am not arguing that beliefs will turn out to be identified with something material (though Dennett, consistent with declaring himself a materialist, does seem to think this would follow from beliefs being illata). I do think this kind of realist argument for beliefs, if it can be made, is significant, though. It gives a radically different basis than merely an allegiance to a principle for ascribing reality and causality to beliefs----and it's an argument with no commitment to materialism.

On the view I've proposed in this dissertation, according to which beliefs are efficacious by being incorporated into dynamic anticipatory schemata that involve a thinker physiologically as well as aid in perception, give rise to physiological, emotional and ideational responses, and direct one's action, it is intelligible that a belief would figure in both a causal explanation of intentional action, and a causal explanation of physiological effects that are not mediated by intention; the dynamic

---

[61] (". . .all there is to really and truly believing that p (for any proposition p) is being an intentional system for which p occurs as a belief in the best (most predictive) interpretation" ("True Believers" 29)).

anticipatory schemata a thinker has developed and employs are cognitive structures with physiological as well as cognitive aspects.

c. Dretske on Beliefs as Indicators

Fred Dretske's account of intentional action is similar to Dennett's in some ways. But, whereas Dennett appealed to natural selection in the evolution of species of biological organisms to explain the existence in humans of something that operates like rationality, Dretske appeals to an individual biological organism's ability to learn from its encounters with its environment. Actually, as does Dennett, Dretske's explanation of human rationality describes a general case of which humans are a specific case.

Dretske's prior commitments are similar to Davidson's and Dennett's; he says: "I am a materialist who thinks that we sometimes do things because of what we believe and want. I pretty much have to accept the idea, then, that reasons are causes" ("Reasons and Causes" 1). However, Dretske's characterization of beliefs diverges somewhat from Davidson's and Dennett's: for Davidson, it was the constitutive ideal of rationality that governed our explanations of reasons as causes. Similarly, for Dennett, what constitutes believing p is simply "being an intentional system for which p occurs as a belief in the best (most predictive) interpretation" ("True Believers" 29). What Dretske seeks is just what Davidson says is impossible, and what Dennett felt he had to be satisfied with only approximating. For, Dretske says he wants to show that "what we think and know, what we desire and intend, the content of our psychological states and attitudes, . . . is actually the property of our internal states that explains their distinctive causal efficacy, their effects on behavior" ("Reasons and Causes" 5). One can see the point of dissatisfaction from which his project arises: Dretske notes that "If the semantic properties of reasons ... [are] irrelevant to explaining their causal properties, ... then the fact that they are causes, taken by itself, is or should be, very little solace indeed" (3).

However, in the end, I do not see how Dretske's account can offer anything more than an alternative explanation of ersatz rationality. In a nutshell, his story goes like this: a reason explanation has the form that S does A because C obtained. He then turns this into a general problem in system functional design, and explains how it could be solved by a biological organism. The general problem ("The Design Problem") is to make a system that does A when and only when condition C obtains. The general solution is for S to have some "internal mechanism that is selectively sensitive to the presence or absence of condition C", which he calls the indicator of C. Then, the indicator of C must somehow be forged into a cause of A. The Design Problem is thus "solved", he says, by "making an indicator of C into a switch for A. Deliberate design, biological evolution, and individual learning are all methods of achieving the same result" (8).

This picture doesn't look much different in fundamentals from the Davidson or Dennett pictures. Dretske thinks his story deals with the role of meaning in the explanation of behavior in virtue of the attention he pays to the connections between the so-called "internal states" of S and what he calls "external conditions". He cites a thermostat as a case of a designed artifact wherein an element internal to a system has a causal role in the output of the system because of what it "indicates or means about the circumstances in which that behavior is produced". But, in the case of intentional behavior of organisms, rather than well-designed artifacts, the explanation of how reasons can be causes involves an organism's ability to learn:

> "It is only when we examine the changes occurring during the life history of individual organisms, internal changes that occur when an organism learns to do A in conditions C, that we find a plausible instance of an internal indicator (of C) acquiring its causal efficacy (in the production of A) because of what it means or indicates about external affairs. . . Learning is a process in which The Design Problem, the problem of how to produce A when condition C exists, is solved, ...

> by a process (roughly) of rewarding A when (and only when) it occurs in conditions C. That process, when it works, automatically makes the internal indicator of C into a cause of A and it does so, . . . because it is an indicator of C." (12)

He calls the learning process whereby "internal indicators (of C) " are converted into "behavior switches" , through processes we may not understand, "a little bit of magic that some biological systems are capable of performing." Dretske then regards these "internal states" as "having meaning", and says that "the fact that they have [a particular] meaning is what helps explain why they have [a particular] causal role." And that, he says, "comes intriguingly close" to saying more than that the reasons are also causes in virtue of their meaning: it is saying that "what we think. . . makes a difference" (13).

But I fail to see that Dretske has improved upon Dennett's story in any fundamental way. In the end, he has to hedge his claim (with "comes intriguingly close"). Whereas Dennett identified such a "comes close" hedge earlier in his story, when he said that the system would "seem to discriminate meanings by actually discriminating things (tokens of no doubt wildly disjunctive types) that co-vary reliably with meanings". Dretske's talk of "internal states" that "have meaning" makes it difficult to compare some of the consequences of the two accounts with respect to the question I raised about the role of belief in two very different kinds of causal explanations.

On Dennett's and Davidson's accounts, it was clear that beliefs were defined in terms of their role in causal explanations of intentional action, and so straightforward to see that their accounts did not satisfactorily address the role of belief in causal explanations of physiological changes such as the placebo effect. I am not so sure what one can say about Dretske's account, for the examples he gives of beliefs also happen to be indicators of things: "a hawk indicator in the chicken, a color indicator in the rat, and an oak tree indicator in the child" (6). Since he also requires that the

indicators indicate infallibly ("if and only if"), it is hard to make a distinction on Dretske's account between "discriminating meanings" and "discriminating things". Thus Dretske doesn't have to make the concession Dennett does, that an organism only seems to discriminate meanings, but the result is that Dretske's story doesn't seem to be about beliefs in general so much as it seems to be about reacting to things in one's immediate environment. Or, rather, about navigating in one's environment. Dretske appeals to Ramsey's characterization of beliefs as "maps by means of which we steer." I am in agreement with an approach that pays attention to how beliefs are employed, but, taken literally, this metaphorical characterizaion is far too narrow a notion of belief.

Dretske pleads for some indulgence here, saying that his account should be seen as suggesting what the basic building blocks of a complex interactive structure could be. So, we can speculate on how the placebo effect could be accounted for on Dretske's view of beliefs. Since beliefs are "indicators", he might be able to say that the belief that one has taken a stimulant could sometimes lead one to exhibit the symptoms of having taken a stimulant, if the belief is regarded as an "indicator" of something for which it would be beneficial to exhibit the same symptoms as a stimulant does. For, since, on his account, beliefs are acquired by conditioning --- i.e., response A is rewarded when it occurs in conditions C --- his account does allow for the phenomenon that an organism can be conditioned to respond physiologically so that, for instance, someone who has been given a pink pill that is a stimulant in conditions when it has been beneficial to exhibit the responses of increased heart rate, may still respond with increased heart rate when a pink pill not containing the stimulant is taken. But even if Dretske would want to give this story (and I'm not sure he would) it still does not provide a satisfactory account, for it requires us to hold that all cases of physiological responses to beliefs -- such as blushing from embarassment or becoming weak-kneed from fear --- result, at bottom, from such conditioning. Recall that Dennett's appeal to natural selection faltered here, too, and for analogous considerations: we cannot explain the physiological effects of belief by appeals to either natural selection or conditioning by rewards, for the physiological effects we seek to explain are not always beneficial. And, anyway, my

speculations about what Dennett and Dretske would say really go beyond their accounts, for each is concerned to explain the role of belief in causing the organism's intentional actions. Neither means to count mere physiological responses as intentional actions.

In spite of this evaluation, I have mentioned Dretske's view because of two differences from the picture Dennett and Davidson provide: (i) Dretske attempts what Dennett and Davidson deem unattainable even in principle: an account on which the causal relations in reason-giving explanations obtain in virtue of the meaning of the reason , i.e., in virtue of the characterizations of cause and effect given in psychological explanations, and (ii) Dretske identifies learning, in addition to natural selection, as a method by which (what I've judged to be) ersatz rationality may be achieved.

What is common to Dennett's and Dretske's pictures is the idea of a thinker as an infomation processing system. In both their accounts, internal states of an organism mediate between the thinker's environment and the actions he takes. The organism is thought to react to its own internal states, and is configured in such a way that its own internal states reflect how things are in its environment. It is conceived as producing actions accordingly, i.e., in response to its internal states. Both Dennett and Dretske take it to be part of their picture that they must account for how some information about the environment is "converted" into an internal state or a feature of an internal state. Dretske talks of indicators, Dennett of abilities to "seem to discriminate meanings by discriminating things".

Viewing a physical organism as a control system of the sort Dennett and Dretske do --- one that contains transducers that produce signals indicative of features in the organism's environment, can produce motions or changes that affect the environment outside the system, and whose behavior exhibits some sort of rule-governed relationship between the transducer signals and the motions regarded as the system's output ---- appeals because of the ability to characterize a

control system functionally as well as describe one in physical terms that can be shown to meet those functional requirements. Hence it allows one to characterize beliefs functionally (i.e., in terms of their role in producing a system's behavior), and yet to identify beliefs with physical states. We have seen that such accounts can yield only ersatz rationality. Further, even if we set the issue of ersatz versus genuine rationality aside, there is another problem: these accounts do not seem able to give satisfying explanations of the efficacy of belief such that the same belief could effect both a placebo-style physiological change and an intentional act.

Thus considering Dennett's and Dretske's accounts of how beliefs are causal has only led us to a functionalist version of the challenge. I do not mean to be arguing for the impossibility in principle of an account of the placebo effect: of course there is no reason why an entity functionally characterized in one way cannot meet the conditions for an entity characterized in another way as well; meeting various functional requirements is what the discipline of design is all about. So the complaint is not that it is impossible that a single belief could satisfy two different functional characterizations; but that in response to being asked why a belief happens to satisfy the two different characterizations that it does satisfy, we are given an explanation that is in essence the statement that the entity that answers the call for one characterization will just happen to satisfy the other; or at least that this has happened, lots of times, in lots of cases, for a lot of different people. The placebo effect is a special kind of self-fulfilling prophecy: the kind of placebo effect that calls out for explanation is the effect whereby the belief that X causes Y is causally efficacious in bringing about Y as well as in getting someone to do X. Not all self-fulfilling prophecies pose this kind of explanatory problem; in contrast, we can explain some cases of self-fulfilling prophecies -- inflation, for instance --- via the operation of genuine rationality.

4. Beliefs and Ontology

In this appendix, I have only been concerned to explain why, given my interest in explaining how beliefs are causally efficacious, I did not think the approach of functionalizing beliefs and employing the "double-nexus" background described above would be very fruitful. I make no pretense of having surveyed all the relevant philosophical views. I would like to comment on a remark in a major book that appeared after I had developed my view, though; Jaegwon Kim's Mind in a Physical World.[62] Kim is concerned about the problems that arise with interlevel explanations and offers his own view, viewing philosophy of mind as one would chemistry and biology: chemical and biological properties can be functionalized so as to be identifiable with what he calls micro-level properties. However, even Kim, who feels his proposal does solve the problem of mental causation for mental properties that are functionalizable, points out that, on his view "the causal powers of mental properties turn out to be just those of their physical realizers, and there are no new causal powers brought into the world by mental properties" (118). He also thinks that there are mental properties that are not functionalizable, and so, on his view, there may be mental properties whose causal efficacy cannot be accounted for, even on his view, "within a physicalist scheme" (119). Kim explicitly grants that one could justifiably see his view as another form of "the irreality of the mental" (119).

Where Kim ends, then: "...physicalism, as an overarching metaphysical doctrine about all of reality, exacts a steep price" (Kim 120) is just about where Mach begins. But whereas Mach was inveighing against the dogma of physicalism, Kim is unapologetic about his commitment to physicalism. He tells us why, too: he sees physicalism as the only alternative to dualism, and "For most of us," he says, "dualism is an uncharted territory, and we have little knowledge of what possiblities and dangers lurk in this dark cavern." (120) It is not clear to me why Kim thinks

---

[62] Thanks to Richard Gale for suggesting I read Kim's book, and to Clark Glymour for discussing it with me.

physicalism is the only alternative to dualism: given his remarks allowing the causal efficacy of some mental properties he thinks cannot be accounted for "within a physicalist scheme", the alternative of expanding the "scheme" seems an obvious one to consider. On the other hand, Mach --- whom Kim does not mention, but who spent a great deal of time in empirical psychophysical research and wrote a book about the foundations of the subject -- did look into an alternative to physicalism other than dualism. I am unapologetic about having done so as well.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, John R. Learning and Memory: An Integrated Approach. New York: John Wiley and Sons, 1995.

Anderson, John R., ed. Cognitive Skills and Their Acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1981.

Arrow, Kenneth J. , Enrico Colombatto, Mark Perlman, and Christian Schmidt, eds. The Rational Foundations of Economic Behavior: Proceedings of the IEA Conference held in Turin, Italy. New York: St. Martin's Press, 1996.

Arthur, W. Brian, Steven N. Durlauf, and David A. Lane, eds. The Economy as an Evolving Complex System II. Reading, MA: Addison-Wesley, 1997.

Atkeson, Andrew and Robert E. Lucas, Jr. "On Efficient Distribution with Private Information." The Review of Economic Studies 59 (July 1992): 427-453.

Axelrod, Robert. The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration. Princeton: Princeton University Press, 1997.

---. The Evolution of Cooperation. New York: Basic Books, 1984.

Bannock, Graham, R. E. Baxter and Evan Davis. The Penguin Dictionary of Economics, 5th ed. London: Penguin Books, 1992.

Barro, Robert J. Getting It Right: Markets and Choices in a Free Society. Cambridge, MA: The MIT Press, 1996.

Bartlett, F. C. Remembering: A Study in Experimental and Social Psychology. London: Cambridge University Press, 1932.

Bechtel, William. "Studies of Categorization: a review essay of Neisser's 'Concepts and Conceptual Development' and Harnad's 'Categorical Perception'". Philosophical Psychology 1 (1988): 381 - 389.

Beinsen, Lutz and Ulrike Leopold-Wildburger. "Towards Bounded Rationality Within Rational Expectations -- Some Comments from and Economic Point of View", in W. Leinfellner and E. Kohler (eds.), Game Theory, Experience, Rationality. 141-152.

Bicchieri, Cristina. Rationality and Coordination. New York: Cambridge University Press, 1993.

Binkley, Robert, Richard Bronaugh, Ausonio Marras, eds. Agent, Action, and Reason. Toronto: University of Toronto Press, 1971.

Boden, Margaret A., ed. The Philosophy of Artificial Intelligence. New York: Oxford University Press, 1990.

Bodenhausen, Galen V., and Wyer, Robert S., Jr. "Social Cognition and Social Reality: Information Acquisition and Use in the Laboratory and the Real World." Social Information Processing and Survey Methodology. Eds. Hans-J. Hippler, Norbert Schwartz, and Seymour Sudman. New York: Springer-Verlag, 1987. 6-41.

Bogdan, Radu J., Ed. Belief: Form, Content and Function. Oxford: Clarendon Press, 1988.

Bolles, Edmund Blair. A Second Way of Knowing: The Riddle of Human Perception. New York: Prentice Hall, 1991.

---. Remembering and Forgetting: Inquiries into the Nature of Memory. New York: Walker and Company, 1988.

Bomfim, Antulio, Robert Tetlow, Peter von zur Muehlen and John Williams. "Expectations, Learning and the Costs of Disinflation: Experiments Using the FRB/US Model." Board of Governors of the Federal Reserve System, August, 1997.

Brandenburger, Adam. "Knowledge and Equilibrium in Games." Journal of Economic Perspectives 6 (1992): 83-101.

Brandom, Robert. Making It Explicit. Cambridge, MA: Harvard University Press, 1994.

Bratman, Michael. Intentions, Plans, and Practical Reason. Cambridge, MA: Harvard University Press: 1987.

---. "Cognitivism About Practical Reason." Ethics 102 (1991): 117-128.

Brody, Howard. Placebos and the Philosophy of Medicine. Chicago: University of Chicago Press, 1977.

Crocker, Jennifer, Brenda Major, and Claude Steele. "Social Stigma." The Handbook of Social Psychology. Eds. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. New York: Oxford University Press, 1998. 504-553.

Cudd, Ann Elizabeth. Common Knowledge and the Theory of Interaction. (Ph.D. dissertation), University of Pittsburgh, 1988.

---. "Game Theory and the History of Ideas About Rationality: An Introductory Survey." Economics and Philosophy. 9 (1993): 101-133.

Davidson, Donald. "Actions, Reasons, and Causes." Journal of Philosophy 60 (1963): 685-700. Rpt. in Essays on Actions and Events. Oxford: Clarendon Press, 1985. 3-20.

---. Essays on Actions and Events. Oxford: Clarendon Press, 1985.

Davidson, Paul. "Rational Expectations: A Fallacious Foundation for Studying Crucial Decision-Making Processes." Journal of Post-Keynesian Economics V (Winter 1982-1983): 182 - 198.

Dennett, Daniel C. Brainchildren: Essays on Designing Minds. Cambridge, MA: MIT Press, 1998.

---. "Three Kinds of Intentional Psychology." The Intentional Stance. Cambridge, MA: MIT Press, 1987. 43-81.

---. "True Believers." The Intentional Stance. Cambridge, MA: MIT Press, 1987. 13-42.

---. The Intentional Stance. Cambridge, MA: MIT Press, 1987.

Dewey , John. Human Nature and Conduct. John Dewey: The Middle Works, 1899-1924, Volume 14. Ed. by Jo Ann Boydston. Carbondale: Southern Illlinois University Press, 1983.

Dretske, Fred. Explaining Behavior: Reasons in a World of Causes. Cambridge, MA: MIT Press, 1991.

----. "Reasons and Causes." Philosophical Perspectives 3: Philosophy of Mind and Action Theory. Atascadero, CA: Ridgeview, 1989.

Dupont, Dominique. "Trading Volume and Information Distribution in a Market-Clearing Framework." Board of Governors of the Federal Reserve System, August 1997.

Dutta, Jayasri, and Stephen Morris. "The Revelation of Information and Self-Fulfilling Beliefs," Journal of Economic Theory 73 (1997): 231-244.

Eatwell, John, Murray Milgate, and Peter Newman, eds. The New Palgrave: The World of Economics. New York: W. W. Norton, 1991.

Einstein, Albert. "Autobiographical Notes." Albert Einstein, philosopher-scientist. Library of Living Philosophers, Vol. 7. Eds. Paul Arthur Schilpp. Evanston, Ill.: Open Court Press, 1949.

Ericsson, K. Anders, and Herbert A. Simon. Protocol Analysis: Verbal Reports As Data. Cambridge, MA: MIT Press, 1984.

Farmer, Roger E. A. "Sticky Prices." The Economic Journal, Vol. 101 (November 1991): 1369-1379.

---. "The Lucas Critique, Policy Invariance and Multiple Equilibria." The Review of Economic Studies. 58 (April1991): 321-332.

Fechner, Gustav Adolphus. Elements of Psychophysics. Trans. Helmut E. Adler. New York: Holt, Rinehart and Winston, 1966.

Fiske, Susan T. "Stereotyping, Prejudice, and Discrimination." The Handbook of Social Psychology. Eds. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. New York: Oxford University Press, 1998. 357-411.

---. "What Does the Schema Concept Buy Us?" Personality and Social Psychology Bulletin. 6 (1980): 543-557.

Forges, Franciose. "Self-Fulfilling Mechanisms and Rational Expectations." Journal of Economic Theory 75(1997): 388-406.

Frydman, Roman. "Towards an Understanding of Market Processes: Individual Expectations, Learning, and Convergence to Rational Expectations Equilibrium." The American Economic Review 72 (1982): 652-668.

Frydman, Roman, and Edmund S. Phelps. Individual Forecasting and Aggregate Outcomes: "Rational Expectations" Examined. New York: Cambridge University Press, 1983.

Gardner, Howard. The Mind's New Science: A History of the Cognitive Revolution. New York: Basic Books, 1985.

Garfield, Jay L. Belief in Psychology:A Study in the Ontology of Mind. Cambridge, MA: MIT Press, 1988.

Gatch, Loren. "To Redeem Metal with Paper: David Hume's Philosophy of Money." Hume Studies. 22 (1996): 169-191.

Gauthier, David. Morals By Agreement. Oxford: Oxford University Press, 1986.

---. Moral Dealing: Contract, Ethics, and Reason. Ithaca: Cornell University Press, 1990.

Gibson, James J. The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin Co., 1966.

Gilbert, Margaret. On Social Facts. Princeton: Princeton University Press, 1992.

---. "Rationality, Coordination, and Convention", Synthese 84 (1990): 1-21.

Gregory, Richard L. Eye and Brain: The Psychology of Seeing. 4th ed. Princeton: Princeton University Press, 1990.

Grossman, Sanford J. "An Introduction ot the Theory of Rational Expectations Under Asymmetric Information." The Review of Economic Studies 48 (Oct., 1981): 541-559.

Grunbaum, Adolf. The Foundations of Psychoanalysis. Berkeley: University of California Press, 1984.

---. Validation in the Clinical Theory of Psychoanalysis. Madison, CT: International Universities Press, 1993.

Grunberg, Emile, and Franco Modigliani. "The Predictability of Social Events." The Journal of Political Economy 42 (December 1954): 465-478.

Grush, Rick. "Yet Another Design for a Brain? Review of Robert Port and Timothy van Gelder's Mind as Motion: Explorations in the Dynamics of Cognition." Philosophical Psychology 10 (1997): 5-23.

Hamlyn, D. W. Understanding Perception: The Concept and Its Conditions. Brookfield, Vermont: Ashgate Publishing Company, 1996.

Hardin, C. L. Color for Philosophers. Indianapolis: Hackett Publishing Company, 1988.

Hargreaves-Heap, Shaun P. The New Keynesian Macroeconomics: Time, Belief, and Social Interdependence. Aldershot: Edward Elgar, 1992.

Harre, Rom. The Anticipation of Nature. London: Hutchinson & Co., 1965.

---. "Wittgenstein and Artificial Intelligence". Philosophical Psychology 1 (1988): 105-115.

Harvey, John T. "The Nature of Expectations in the Foreign Exchange Market: A Test of Competing Theories." Journal of Post-Keynesian Economics 21 (Winter 1998-1999): 181-200.

Hausman, Daniel M. and Michael S. McPherson. Economic Analysis and Moral Philosophy. Cambridge, UK: Cambridge University Press, 1996.

---. "Taking Ethics Seriously: Economics and Contemporary Moral Philosophy." Journal of Economic Literature. 31: 671-731.

Hausman, Daniel M. The Inexact and Separate Science of Economics. Cambridge, UK: Cambridge University Press, 1992.

Hausman, Daniel M., ed. The Philosophy of Economics: An Anthology. New York: Cambridge University Press, 1984.

Hayek, F. A. Individualism and Economic Order. Chicago: University of Chicago Press, 1948.

---. "Economics and Knowledge." Individualism and Economic Order. Chicago: University of Chicago Press, 1948. 33-56 .

---. "The Use of Knowledge in Society." The American Economic Review 65(1945): 519-530. Rpt. in Individualism and Economic Order, Chicago: University of Chicago Press, 1948. 77-91.

Heidelberger, Michael. "Fechner's Impact for Measurement Theory" Behavioral and Brain Sciences 16 (1993): 146-148.

---. "The Unity of Nature and Mind: Gustav Theodor Fechner's Non-Reductive Materialism", in Poggi, Stefano and Maurizio Bossi, Eds., Romanticism in Science: Science in Europe, 1790-1840 (Boston Studies in the Philosophy of Science 152). Dordrecht: Kluwer 1994. 215-236.

Heil, John. Perception and Cognition. Berkeley and Los Angeles: University of California Press, 1983.

---. Philosophy of Mind: A Contemporary Introduction. London and New York: Routledge, 1998.

Heil, John and Alfred Mele. Mental Causation. Oxford: Clarendon Press, 1993.

Hippler, Hans-J., Norbert Schwartz, and Seymour Sudman, eds. Social Information Processing and Survey Methodology. New York: Springer-Verlag, 1987.

Hogarth, Robin M., and Melvin W. Reder, eds. Rational Choice: The Contrast between Economics and Psychology. Chicago: The University of Chicago Press, 1987.

Holland, John H. Hidden Order: How Adaptation Builds Complexity. Reading, MA: Addison-Wesley, 1995.

Hollis, Martin. Reason in Action: Essays in the philosophy of social science. Cambridge: Cambridge University Press, 1996.

---. The Cunning of Reason. Cambridge: Cambridge University Press, 1987.

Hollis, Martin and Edward Nell. Rational Economic Man: A Philosophical Critique of New-Classical Economics. New York: Cambridge University Press, 1975.

Hoover, Kevin D., ed. Macroeconometrics: Developments, Tensions and Prospects. Boston: Kluwer Academic Publishers, 1995.

---. The New Classical Macroeconomics. Oxford: Basil Blackwell, 1988.

Hume, David. Essays: Moral, Political, and Literary, Revised Edition. Ed. Eugene F. Miller. Indianapolis: Liberty Fund, 1985.

---. "Of Money." Essays: Moral, Political, and Literary, Revised Edition. Ed. Eugene F. Miller. Indianapolis: Liberty Fund, 1985. 281 - 294.

---. "Of Interest." Essays: Moral, Political, and Literary, Revised Edition. Ed. Eugene F. Miller. Indianapolis: Liberty Fund, 1985. 295 - 307.

---. A Treatise of Human Nature, Second Edition. Ed. L. A. Selby-Bigge. Oxford: Oxford University Press, 1978.

---. An Enquiry Concerning the Principles of Morals. Ed. J. B. Schneewind. Indianapolis: Hackett, 1983.

Hylleberg, Svend, and Martin Paldam, eds. New Approaches to Empirical Macroeconomics. Oxford, England and Cambridge, MA: Blackwell Publishers, 1991.

James, William. A Pluralistic Universe. Lincoln, NE: University of Nebraska Press, 1996.

----. The Principles of Psychology. Cambridge, MA: Harvard University Press, 1983.

---. "The Will to Believe." The Writings of William James. Ed. John J. McDermott. Chicago: University of Chicago Press, 1977. 717-734.

Kelley, David. The Evidence of the Senses: A Realist Theory of Perception. Baton Rouge and London: Lousiana State University Press, 1986.

Khalfa, Jean, ed. What is Intelligence? Cambridge: Cambridge University Press, 1994.

Kim, Jaegwon. Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, MA: MIT Press, 1998.

Kosslyn, Stephen Michael. Ghosts in the Mind's Machine: Creating and using images in the brain. New York: W. W. Norton, 1983.

Krugman, Paul. "History Versus Expectations." The Quarterly Journal of Economics 106 (May, 1991): 651-667.

Laffont, Jean-Jacques. The Economics of Uncertainty and Information. Trans. John P. Bonin and Helene Bonin. Cambridge, MA and London, England: The MIT Press, 1989.

Lando, Henrik, and Michael Teit Nielsen. "Flexibility and Uncertainty in the Housing Market," International Review of Law and Economics 18 (1998): 419-431.

Lee, Bong-Soo. "On the Rationality of Forecasts." The Review of Economics and Statistics 73 (May, 1991): 365-370.

Leinfellner, Werner and Eckehart Kohler. Game Theory, Experience, Rationality: Foundations of Social Sciences, Economics and Ethics. In Honor of John Harsanyi. Vienna Circle Institute Yearbook 5 (1997). Dordrecht, Netherlands: Kluwer Academic Publishers,1998.

Lengwiler, Yvan. "Certainty Equivalence and the Non-Vertical Long Run Phillips-Curve." Board of Governors of the Federal Reserve System. June 1998.

Lewis, David. Convention: A Philosophical Study. Cambridge, MA: Harvard University Press, 1969.

Lucas, Robert E., Jr. "Asset Prices in an Exchange Economy." Econometrica 46 (1978): 1429-1445.

---. "Econometric Policy Evaluation: A Critique." Studies in Business Cycle Theory. Cambridge, MA: MIT Press, 1981. 104-130.

---. "Expectations and the Neutrality of Money." Journal of Economic Theory 4 (1972): 103-124. Rpt. in Lucas, Robert E., Jr. Studies in Business Cycle Theory. Cambridge, MA: MIT Press, 1981. 67-89.

---. Models of Business Cycles. Oxford: Basil Blackwell, 1987.

---. "Monetary Neutrality." Nobel Lectures in Economic Sciences 1991-1995. Ed. Torsten Persson. Singapore: World Scientific Publishing, 1997.

---. Studies in Business Cycle Theory. Cambridge, MA: MIT Press, 1981.

Lucas, Robert E., Jr., and Thomas J. Sargent, eds. Rational Expectations and Econometric Practice. Vol. 1. Minneapolis: University of Minnesota Press, 1981.

Luce, R. Duncan, and Howard Raiffa. Games and Decisions. New York: Dover Books, 1989.

Lycan, William G., ed. Mind and Cognition. Oxford: Basil Blackwell, 1990.

Lyon, Robert. "Notes on Hume's Philosophy of Political Economy." Journal of the History of Ideas 31 (1970): 457-461.

Lyon, Gordon. "The Experience of Perceptual Familiarity." Philosophy 71 (1996): 83-100.

Mach, Ernst. The Analysis of Sensations. London: Routledge/Thoemmes Press, 1996.

Maki, Uskali, Bo Gustafsson, and Christian Knudsen. Rationality, Institutions and Economic Methodology. New York: Routledge, 1993.

Marr, David. Vision. San Fransisco: W. H. Freeman and Sons, 1982.

McDowell, John. Mind and World. Cambridge, MA: Harvard University Press, 1994.

---. "The Content of Perceptual Experience." Philosophical Quarterly 44(1994): 190-205.

Michon, John A. and Janet L. Jackson, eds. Time, Mind, and Behavior. Berlin and Heidelberg: Springer Verlag, 1985.

Mills, Terence C. "Signal Extraction and Two Illustrations of the Quantity Theory." The American Economic Review 72 (December 1982): 1162-1168.

Mishkin, Frederic S. "The Rational Expectations Revolution: A Review Article of: Preston J. Miller, Ed.: The Rational Expectations Revolution, Readings From the Front Line." National Bureau of Economic Research Working Paper Series, No. 5043. Cambridge, MA: National Bureau of Economic Research, February 1995.

Morton, Adam. "Folk Psychology is Not A Predictive Device". Mind 105 (1996): 119-137.

Moser, Paul K. Rationality in Action: Contemporary Approaches. New York: Cambridge University Press, 1990.

Moyers, Bill. Healing and the Mind. New York: Doubleday, 1993.

Muth, John. "Rational Expectations and the Theory of Price Movements." Econometrica 29 (1961):315-335. Rpt. in Lucas, Robert E., Jr. and Sargent, Thomas J., Eds. Rational Expectations and Econometric Practice. Vol. 1. Minneapolis: University of Minnesota Press, 1981. 3-22.

Nash, John F. "The Bargaining Problem." Econometrica 18 (1950):155-162.

Nash, John. "Two-Person Cooperative Games." Econometrica 21 (1953):128-140.

Neilson, Lars Tyge, Adam Brandenburger, John Geankoplos, Richard McKelvey, and Talbot Page. "Common Knowledge of an Aggregate of Expectations." Econometrica 58 (1990): 1235-1239.

Neisser, Ulric. Cognition and Reality. New York: W. H. Freeman, 1976.

---. "On 'Social Knowing'." Personality and Social Psychology Bulletin 6 (1980): 601-605.

---. "Perceiving, Anticipating, and Imagining". Perception and Cognition. (Minnesota Studies in the Philosophy of Science IX). Ed. C. Wade Savage. Minneapolis: University of Minnesota Press, 1978.

---, ed. Memory Observed: Remembering in Natural Contexts. New York: W. H. Freeman, 1982.

Neisser, Ulric and Eugene Winograd, Eds. Remembering Reconsidered. New York: Cambridge University Press, 1988.

Neisser, Ulric and Robyn Fivush. The Remembering Self. New York: Cambridge University Press, 1994.

Oakes, Penelope J., Alexander Haslam, and John C. Turner, eds. Stereotyping and Social Reality. Oxford: Blackwell Publishers, 1994.

Oakeshott, Michael. Rationalism in Politics and other essays. London: Methuen & Co., 1967.

Osborne, Martin J., and Ariel Rubinstein. A Course in Game Theory. Cambridge, MA: MIT Press, 1994.

Persson, Torsten, Ed. Nobel Lectures in Economic Sciences 1991-1995. Singapore: World Scientific Publishing, 1997.

Phelps, Edmund S. "Equilibrium: an Expectational Concept." The New Palgrave: The World of Economics. Eds. John Eatwell, Murray Milgate, and Peter Newman. New York: W. W. Norton, 1991. 224-227.

---. "Recent Studies of Speculative Markets in the Controversy Over Rational Expectations." EUI Working Paper No. 87/267. San Domenico, Italy: European University Institute, 1987.

Port, Robert F. and Timothy van Gelder, eds. Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA: MIT Press, 1995.

Price-Williams, D. R., ed. Cross-cultural studies. Middlesex, England; Penguin Books, 1969.

Rapoport, Anatol. N-Person Game Theory: Concepts and Applications. Ann Arbor: University of Michigan Press, 1970.

Rashid, Salim. "David Hume and Eighteenth Century Monetary Thought: A Critical Comment on Recent Views." Hume Studies 10 (1984):156-164.

Reed, Edward and Rebecca Jones, eds. Reasons for Realism: Selected Essays of James J. Gibson. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

Resnick, Michael D. Choices: An Introduction to Decision Theory. Minneapolis: University of Minnesota Press, 1987.

Rotemberg, Julio J., and Michael Woodford. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: Expanded Version." National Bureau of Economic Research Technical Working Paper Series. No. 233. Cambridge, MA: National Bureau of Economic Research, May 1998.

Russell, Thomas. "Macroeconomics and Behavioral Finance: A Tale of Two Disciplines." Game Theory, Experience, Rationality. Eds. W. Leinfellner and E. Kohler. Dordrecht, Netherlands: Kluwer Academic Publishers,1998. 153-159.

Ryan, Alan, ed. The Philosophy of Social Explanation. Oxford: Oxford University Press, 1973.

Ryle, Gilbert. The Concept of Mind. Hammondsworth: Penguin, 1963.

---. On Thinking. Oxford: Basil Blackwell, 1979.

Sacks, Oliver. Seeing Voices. Berkeley and Los Angeles: University of California, 1989.

Salyer, Kevin D., and Steven M. Sheffrin. "Spotting Sunspots: Some Evidence in Support of Models With Self-fulfilling Prophecies." Journal of Monetary Economics 42 (1998): 511 - 523.

Sargent, Thomas J., ed. Energy, Foresight, and Strategy. Washington, D. C.: Resources for the Future, 1985.

Sargent, Thomas J. Rational Expectations and Inflation, Second Edition. New York: Harper Collins, 1993.

---. "Interpreting Economic Time Series." The Journal of Political Economy 89 (April 1981): 213-248.

Savage, C. Wade, ed. Perception and Cognition: Issues in the Foundations of Psychology. Minneapolis: University of Minnesota Press, 1978.

Savage, Leonard J. The Foundations of Statistics. New York: Dover Publications, 1972.

Sayre, Kenneth M. and Frederick J. Crosson, eds. The Modeling of Mind. Notre Dame: Notre Dame Press, 1963.

Schelling, Thomas C. "An Essay on Bargaining." The American Economic Review 46 (1956): 281-306.

---. "Experimental Games and Bargaining Theory." World Politics 14 (1961): 47-68.

---. "For the Abandonment of Symmetry in Game Theory." The Review of Economics and Statistics 41 (1959): 213-224.

---. Micromotives and Macrobehavior. New York: W. W. Norton, 1978.

---. The Strategy of Conflict. Cambridge, MA: Harvard University Press, 1960.

Searle, John. The Construction of Social Reality. New York: Free Press, 1995.

Sellars, Wilfrid. Empiricism and the Philosophy of Mind. Cambridge, MA: Harvard, 1997.

Sen, Amartya and Bernard Williams, eds. Utilitarianism and Beyond. New York: Cambridge University Press, 1982.

Sheffrin, Steven M. Rational Expectations, Second Edition. New York: Cambridge University Press, 1996.

Shubik, Martin. Game Theory in the Social Sciences: Concepts and Solutions. Cambridge, MA: MIT Press, 1982.

Simon, Herbert. "Rational Decision Making in Business Organizations." The American Economic Review 69 (1979): 493-513.

---. "Theories of Decision-Making in Economics and Behavioral Science." The American Economic Review 59 (1959): 253-283.

Sklar, Lawrence. Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics. New York: Cambridge University Press, 1995.

Slade, Margaret E. "Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies." Review of Economic Studies 59 (1992): 257-276.

Someran, M. W. van, Y. F. Barnard, and J. A. C. Sandberg. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. San Diego, CA: Academic Press, 1994.

Sorger, Gerhard. Imperfect Foresight and Chaos: An Example of a Self-fulfilling Mistake." Journal of Economic Behavior and Organization 33 (1998): 363-383.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. "Stereotype Threat and Women's Math Performance." Journal of Experimental Social Psychology 35, 1999. pp. 4-28.

Sterrett, Susan G. "Expecting and Predicting." Unpublished manuscript. 1996.

---. "On Social Institutions." Unpublished manuscript. 1991.

---. "Sounds Like Light: Einstein's Special Theory of Relativity and Mach's Work in Acoustics and Aerodynamics." Studies in History and Philosophy of Modern Physics 29B(1998): 1-35.

Stitch, Stephen. From Folk Psychology to Cognitive Science: The Case Against Belief. Cambridge, MA: MIT Press, 1986.

Tanaka, Toshihiro. "Hume to Smith: An Unpublished Letter." Hume Studies 12 (1986): 201 - 209.

Thompson, Evan. Colour Vision: A Study in Cognitive Science and the Philosophy of Perception. New York: Routledge, 1995.

Townsend, Robert M. "Forecasting the Forecasts of Others." Journal of Political Economy 91 (1983): 546-576.

Tuomela, Raimo. "We Will Do It: An Analysis of Group-Intentions." Philosophy and Phenomenological Research 51 (1991): 249-277.

Uleman, James S. and John A. Bargh, eds. Unintended Thought. New York: Guilford Press, 1989.

Velleman, J. David. "Practical Reflection." The Philosophical Review 44 (1985): 33-61.

---. Practical Reflection. Princeton: Princeton University Press, 1989.

Von Neumann, John, and Oskar Morgenstern. Theory of Games and Economic Behavior. Princeton: Princeton University Press, 1944.

Vygotsky, L. S. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press, 1978.

---. Thought and Language. Cambridge, MA: MIT Press, 1986.

Whitley, John D. A Course in Macroeconomic Modelling and Forecasting. Hemel Hempstead: Harvester Wheatsheaf, 1994.

Wilkes, Kathleen V. (1988) Real People: Personal Identity Without Thought Experiments. Oxford: Clarendon Press, 1993.

Wittgenstein, Ludwig. Philosophical Investigations, Third Edition. New York: Macmillan, n.d.

Woodford, Michael. "Learning to Believe in Sunspots." Econometrica 58 (March 1990): 277-307.