PSYCHOLOGICAL EXPLANATION: TACIT THEORY OR SIMULATION?

by

ANGELA J. ARKWAY

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1995

UMI Number: 9605569

UMI

300 North Zeeb Road
Ann Arbor, MI 48103

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

6/1/95
Date

Stephen Schiffer
Chair of Examining Committee

6/2/95
Date

Executive Officer

Hartry Field

Stephen Stich

William Earle

John Greenwood

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Acknowledgments

I would like to thank Professor Stephen Schiffer for his insights, guidance and encouragement throughout this project. It is through trying to meet his high standards that I made the progress that enabled me to write this dissertation.

I thank warmly also Professor Hartry Field for his participation and constructive criticism. Thanks are due to Professor Stephen Stich whose advice on the simulation theory came at just the right moment. Professor William Earle and Professor John Greenwood were always encouraging and for this too I am grateful.

Finally, I thank my husband, Richard, for his patience and supportiveness. Had he not been generously endowed with these qualities, this dissertation would never have been written.

Table of Contents

Chapter One

The Present State of Affairs

After the thoroughgoing rejection of the conceptual scheme and vocabulary of folk psychology by radical experimental psychologists and the re-definition of its vocabulary and its re-interpretation by invoking behavior and behavioral dispositions by less radical philosophers in the 1950s, folk psychology made a comeback on the psycho-philosophical scene in the 1960s. Not only was the error of such behavioristic ways recognized because of their inadequacy to describe complex, intelligent, innovative human behavior but folk psychological explanations of behavior in terms of beliefs, desires, knowledge or plans seemed to provide a more promising starting point for scientific theories aimed at explaining human behavior. During the 1970s the use of commonsense mentalistic terms proliferated in psychological theorizing and it is still the idiom of cognitive science today.

While some philosophers welcomed this re-valorization of folk psychology in the halls of academia and the attempt to construct psychological theories using the conceptual scheme and the vocabulary of folk psychology, others were skeptical that theories using such concepts could play any role in a science aimed at explaining human behavior. The purpose of this study is to explore the nature of commonsense psychological explanation of behavior as it is construed by theorists in the first category. The way we shall proceed is to focus our attention on two philosophical theories. The first is the theory theory, the theory of the nature of folk psychology and of psychological explanation the most widely

accepted today. The second is the simulation theory, a theory that has been offered recently as a contender to the theory theory.

Before we begin our discussion of these theories however, it will be useful to state clearly the philosophical question that we consider it incumbent upon any account that purports to describe the nature of psychological explanation to answer. It can be put succinctly: what is it that constitutes a correct explanation of the behavior of a person in terms of his beliefs and desires? Although we consider that an inquiry into the nature of commonsense explanation should concentrate first and foremost on the discussion of what would constitute a correct explanation of that type, the questions that arise when this topic is examined in the philosophical literature are several. It will be worth our while to mention two of these other questions here in order to clearly identify and differentiate our own focal point.

An issue that is sometimes confused with the question of what it is that constitutes a correct explanation is the epistemological question of what it is to know that our explanations are correct. These are different questions however. Indeed theorists who agree about the answer to the first could have quite different opinions about the answer to the epistemological question. In fact one could have very definite views about the conditions that a correct explanation meets and hold that it is impossible to know when those conditions are met, that there is no determinate answer to the epistemological question.

Perhaps the question that is the most often confused with what we take to be the main *philosophical* question about psychological explanation is that of the process that takes place within the organism in the course of the production of explanations of behavior in terms of beliefs and desires.

In this case the chief concern is to describe the cognitive organization, the belief-desire mechanism, that is instantiated by any organism that has the concepts of belief and desire and explains behavior by ascribing those intentional states. The problem presented by discovering the way this mechanism works, the "black box" problem, falls largely within the domain of investigation of cognitive scientists. In spite of the fact that the question of how the organism works and the question of what constitutes a correct explanation that issues from that process clearly seem to be different questions we shall see in the course of our discussion that these two issues are sometimes conflated, and sometimes even appear to be confused.

Although we maintain that it is a separate issue, we do not deny that the question of the cognitive organisation underlying explanation is important to philosophers looking for an account of the nature of correct explanation. Indeed such an account will not even qualify for serious consideration as a viable account unless it is compatible with some plausible model of the way the belief-desire mechanism works. The philosophers discussed below lean heavily on data obtained from psychological experiments to support their views. If it were to turn out that the view of the nature of explanation which they consider to be the most plausible on conceptual grounds points to the organism's having a certain belief-desire mechanism and if empirical evidence seems to confirm that this is the mechanism that human beings instantiate, then this finding will indeed add strength to their over-all argument in favor of that view of explanation.

The theories under consideration here are examined with regard to the plausibility of their answer, or with regard to their potential to answer, what we have described as the principal question about commonsense

psychological explanation: what it is that constitutes a correct one. In this chapter, sketches of the theory theory and of the simulation theory are given and the current debate between the proponents of these views is laid out. In Chapter Two we take a closer look at the theory theory. In that chapter we examine the motivation behind the theory theory of psychological explanation and consider several possible versions of it. Chapter Three consists of a close examination of the simulation theory. Our aim in the first part of this discussion is to describe what it is that the simulation theory purports to account for, and to make clear how it proposes to do so. We then examine the potential of the simulation theory to provide a replacement for the theory theory view of what it is that constitutes a correct commonsense explanation of behavior. In Chapter Four, after bringing to the fore the questions that our examination of commonsense psychological explanation of behavior has elicited, we suggest and discuss different answers to those questions.

1:1 The Theory Theory: A General View

Theory theorists hold that folk psychology is a theory comprised of a vast network of folk psychological causal laws that link mental states to behavior, to other mental states, and to sensory input. This collection of laws is internally represented and tacitly known by all who have the concepts of belief, desire, and of all the other intentional states and of the qualitative mental states also.[1]

---

[1]Some theorists, in particular Stephen Stich and Hartry Field, would argue that the view of the theory theory of folk psychology laid out here is a narrow interpretation of that view. Whereas for Hartry Field a broader interpretation would exclude the assumption that the laws are tacitly known in some sense by the folk, for Stephen Stich a broader interpretation would be one that does not assume that the knowledge structure which composes the tacitly known theory include lawlike generalizations. Our discussion of the theory theory centers on the interpretation that posits tacit knowledge of internally-represented laws because on

According to the theory theory, a commonsense explanation of behavior that is correct is so because the event being explained is subsumed by one of these internally-represented tacitly-known psychological laws, and because the explainer alludes to this law in giving the explanation. The law states which mental events are causally relevant, typically, or in cases where all else is equal, or in certain idealized conditions, to the event to be explained. It is for this reason that by alluding to the law an explainer is able to say correctly which mental states are explanatorily relevant to the occurrence of a particular piece of behavior.

The existence of such laws seems assured by the fact of our impressive success at prediction. Theory theorists hold that the reason that we are so often correct in predicting what people will do when we ascribe certain beliefs and desires to them indicates that there are reliable but contingent generalizations about human behavior and the mental states that typically underlie it. In fact, they say, these generalizations are so reliable that they warrant the status of laws.

Theory theorists further assume that if indeed reference to such laws can account for our evident success at prediction then it can also account for our other folk psychological practices, practices where our success is less evident: for example, our explanations, interpretations, and descriptions of the behavior of ourselves and others in intentional terms. That some of our explanations, interpretations and descriptions are indeed

---

our understanding the assumption that what underpins our explanations are *folk laws* is an integral part of the theory theory. In the first section of Chapter Two we say what we think is the underlying motivation of theory theorists of folk psychology. If this hypothesis is correct it will follow that the knowledge structure underlying our explanations includes some kind of *law*. Also, it seems to us that if, as the theory theorists represented here argue, these internally-represented laws are *folk* laws, then, by definition they will be known in some sense by the folk.The simulationists' arguments against the theory theory laid out in the next section are levelled at the interpretation presented here.

correct seems to be confirmed by the fact of our success in our interaction with the world, our attaining most of the time the everyday goals we set ourselves. Theory theorists maintain that the existence of, and our reference to, psychological causal laws would account for this success as it would for the confidence with which we use subjunctive and counterfactual conditionals when describing hypothetical situations in common discourse. This is how the assumption of the existence of psychological laws is justified for the theory theorist.

It is generally agreed among these theorists that the laws that constitute this theory cannot be strict laws of universal form, laws that state that whenever you have a G event you have an F event, where the G event is a mental event and the F event is a behavioral event, as it is simply not the case that mental events and behavioral events are linked in that way. A suggestion that is popular is that these laws include clauses that state that they hold in cases where all else is equal. In this view commonsense psychological laws are considered to be unstrict laws, laws that admit of exceptions. Such "ceteris paribus" or "hedged" laws are already used in special sciences other than psychology, say these theorists. They are usually expressed as universally quantified conditional statements. These conditionals have in their antecedent a conjunction of the relevant explanatory factors and the consequent is the behavioral event to be explained. Were a tacitly-known psychological ceteris paribus law to be made explicit, it might take the following form: If x wants that p, and x believes that not-p unless q, and x believes that x can bring it about that q, then, ceteris paribus, x tries to bring it about that q. What is expressed by such a law is that if an organism has the beliefs and desires subsumed by the law and if all else is equal or if certain ideal conditions obtain, then the

behavior connected to those states by that law will occur. In cases where the explanation in which this law is alluded to turns out to be wrong it is because the ideal conditions implicitly alluded to in the law did not obtain on that occasion. This does not detract from the lawlike status of these conditionals nor from the plausibility of the collection of them qualifying as a theory, according to the theory theorist.

It is one of the main advantages of the theory theory of folk psychology that this view of psychological explanation fits the deductive-nomological model of the covering law model of explanation. Indeed, according to theory theorists, commonsense psychological explanations of behavior are deductive arguments in which the covering law, usually an unstrict cp law of one sort or another, is implicitly deployed. We shall have more to say about this in Chapter Two.

Another important advantage of the theory theory of folk psychology is that its view of the mechanism that underlies our capacity to give commonsense explanations and predictions of behavior fits a well-established model in cognitive science of the cognitive organisation that underlies our various capacities. These cognitive science theories say that certain of our abilities can be accounted for by the fact that we have at our command, and tacitly deploy, a theory of how things work in certain domains. Thus tacit knowledge of various theories accounts for our ability to produce and comprehend natural language, to solve mathematical problems, to recognize objects visually, and to manipulate and predict the behavior of middle-sized physical objects. Theory theorists of psychological explanation hold that the cognitive organization responsible for our ability to give commonsense psychological explanations is just another instantiation of this model.

1:2 The Debate

The current debate between simulationists and theory theorists arose when simulationists, in presenting their account, indicated that they were offering it as a replacement for, what they considered to be, the flawed theory theory of folk psychology.[2] Let us look at their objections to the theory theory and at the theory theorists' rebuttals.

The theorists who introduced, or re-introduced as we shall see in Chapter Three, the simulation theory level their objections to the theory theory of folk psychology at the suggestion that we have tacit knowledge of complex folk psychological laws that link mental states to input, to each other, and to behavior. The issues that worry them in particular are the anomalous precocity of young children that such knowledge presupposes, the method of acquisition by human beings of knowledge of the postulated commonsense psychological laws, and lastly, the fact that no-one has been able to formulate satisfactorily a single law of this type. Let us take the problem posed by the assumed necessity of anomalous precocity first.

The empirical data tells us roughly the ages at which children show signs of having certain abilities. Developmental psychologists say that at around the age of three years children can anticipate that others might differ from themselves with respect to actual or potential targets that they see, want, like, expect, know, and that at around the age of four years "Children acknowledge that people can diverge not just with respect to the targets of their mental stance, (but that) they can also construe the very same target differently. They acknowledge that the beliefs, perceptual

---

[2] Robert Gordon, "Folk Psychology as Simulation," Mind and Language 1 (Summer 1986), 158-71; Alvin Goldman, "Interpretation Psychologized," Mind and Language 4 (Autumn 1989), 162-85.

experiences and emotions of two people may conflict with respect to the same situation; in addition, they acknowledge that the construal, and consequent attitude of the same person, with respect to a given situation may conflict over time (Flavell, 1988; Perner, 1991). For example, they understand that whereas they see or believe that X is the case with respect to a given situation, another person--or they themselves at some earlier point in time--might see or believe that not X is the case with respect to that situation."[3] Thus children of the age of four years show signs of having the capacity to attribute to others beliefs that differ from their own in a given situation and they understand that beliefs, their own and others', about a particular situation can change over time, capacities which experiments indicate they do not have earlier.

Theory theorists account for this accrual of capacities by saying that the folk theory is acquired by increments. This is why, to extrapolate from their view, whereas at the age of three children have acquired in some manner or other tacit knowledge of the personal maxims that connect the beliefs and desires that they themselves experience with respect to different targets, it is not until around the age of four that they acquire knowledge of the general laws that connect target-directed beliefs and desires to their real causes, these being the behaver's actual beliefs about the way the world is. Up to that point children assume that the causes of other peoples' beliefs

---

[3] Paul Harris, "From Simulation to Folk Psychology: The Case for Development," Mind and Language 7 (Spring/Summer 1992): 126. In this paper Harris refers to J. H. Flavell, "The Development of Children's Knowledge about the Mind: From Cognitive Connections to Mental Representation," in Developing Theories of Mind, ed. J. W. Astington, P. L. Harris and D. R. Olson (New York: Cambridge University Press, 1988), and also to J. Perner, Understanding the Representational Mind (Cambridge, MA.: MIT Press, 1991).

and desires are the same as what causes their own, that is that they are caused by the way the world is as they themselves see it. [4]

Simulation theorists point out that this account of what underlies a child's ability demonstrates amply the anomalous precocity that theory theorists assume that children have at the age of four as it takes for granted that they have at their command and frequently deploy complex laws that link overt behavior not only to mental states similar to the ones they have experienced in similar situations, but behavior to a whole range of other possible mental states which they themselves have not experienced, mental states that represent the world as in fact it is not, according to their own experience, but as it might be from another's point of view.

In an effort to overcome the incredulity evoked by the suggestion that young children have knowledge of such complex laws, theory theorists point out that such anomalous precocity is already demonstrated in a child's capacity to understand natural language. In this case it is widely accepted that children know and use a complex theory of grammar when they understand and use language.

One simulation theorist, Robert Gordon, balks at the comparison to grammar to counter the anomalous precocity argument. The difference, he says, lies in the fact that generative grammarians, in order to explain commonsense intuitions of grammaticality, have no compunction about introducing technical vocabulary. They do not claim that what explains our grammatical competence is knowledge of *folk* grammar. Thus there is no requirement that the theory of grammar be framed in terms of the child's own concepts. The question of anomalous precocity thus does not arise in

---

[4] Stephen Stich and Shaun Nichols, "Folk Psychology: Simulation or Tacit Theory," Mind and Language 7 (Spring/Summer 1992): 60.

this context. Theory theorists, however, says Gordon, "*do* generally make the corresponding claim: that what explains our competence in predicting behavior is *folk psychology* -the principles that underlie our mental attributions and explanations of behavior."[5]

The problem of the requisite anomalous precocity, however. is not the only problem that the theory theorist must solve in relation to the developmental data, according to simulationists. For even if one grants for a moment that the data can be accounted for by knowledge of such laws and that young children are anomalously precocious in just this way, the question then arises of how they get to know complex laws that link mental states to each other, to input and to behavior. There are two popular hypotheses about this and both have met with strong objections. One suggestion is that the apparatus for intentional explanation is part of our innate endowment. Another is that knowledge of these laws is culturally produced and transmitted.

The latter suggestion represents the standard view about the acquisition of knowledge of the laws of a theory of folk psychology. This position is summarized by Paul Churchland: "All of us learn that framework (at mother's knee, as we learn our language), and in doing so we acquire the common-sense conception of what conscious intelligence *is*. We may call that theoretical framework 'folk psychology'. It embodies the accumulated wisdom of thousands of generations' attempts to understand how we humans work."[6]

---

[5] Robert Gordon, "Reply to Stich and Nichols," Mind and Language 7 (Spring/Summer 1992): 91

[6] Paul Churchland, Matter and Consciousness (Cambridge: MIT Press, 1988), 59

The view that the laws of the folk theory of psychology are culturally produced and culturally transmitted is found to be implausible by at least one simulation theorist, Alvin Goldman. He points out that the cultural transmission model is a non-starter if it implies that the platitudes themselves are learned from mothers' formulations since it is only philosophers who attempt to articulate the laws and few children have mothers who are also philosophers.[7] In answer to the suggestion that children generate these laws for themselves after picking up clues from the speech of their elders, Goldman queries whether this would lead to the generation of the same laws. He reminds us that findings about the standard methods of theory-generation and acceptance indicate that usually scientists, when exposed to the same data and when working independently, come up with different theories. Since theory theorists generally hold that children do come up with the same theory their view seems to contradict these findings.

The innateness hypothesis represents a shift from the standard picture of folk psychology being constituted by the incremental accumulation of wisdom. One nativist theory theorist is Jerry Fodor. He argues that besides its being conceptually the most plausible suggestion there is impressive empirical evidence that knowledge of commonsense Homo Sapiens psychology is innate.[8]

The superior conceptual plausibility of this hypothesis, according to him, is constituted by the fact that this is likely to be a better explanation

---

[7] Alvin Goldman, "Interpretation Psychologized," Mind and Language 4 (Autumn 1989): 168; "In Defense of the Simulation Theory," Mind and Language 7 (Spring/Summer 1992): 106-7.

[8] Jerry Fodor, "Creation Myth," in Psychosemantics (Cambridge: MIT Press, 1987), 132.

for the intricate, varied nature of human behavior and its adaptability to rapidly changing environments than the hypothesis that each organism has to spend an important part of its life learning enough about the psychology of its conspecifics to allow it to integrate its own behavior with theirs. Insofar as empirical evidence is concerned, one point that this theorist makes is that it appears to be the case that acceptance of some form of intentional explanation is a cultural universal. He also interprets the fact that developmental psychologists find that the apparatus of mentalistic explanation is operative early as confirming evidence for the innateness hypothesis. He adds that it is significant that the results of the empirical tests that confirm early acquisition of the framework of folk psychology are matched by results gathered by psycholinguists that indicate the early presence of knowledge of the universal rules of grammar. And lastly Fodor holds that the innateness hypothesis is supported by the fact that there is, as yet, no suggestion as to how a child might acquire the apparatus of intentional explanation from experience. In answer to the suggestion that children learn mental notions at the same time as they learn to use mental terms, this theorist asks how a child who has no previous idea of what remembering is goes about learning that 'remember' means remember.

Simulationists, however, are no more convinced by the argument that knowledge of these complex laws is innate than they are by the argument that it is culturally acquired and transmitted. Alvin Goldman points out that the results of developmental psychologists attest only to the early awareness of mentalistic phenomena and not to the awareness of psychological laws.[9] He adds that with regard to the early use of the word 'remember', what this may suggest is merely that children have an innate

[9] Goldman, "Interpretation Psychologized," 168.

propensity to identify mental categories within themselves just as they have one to identify and individuate material bodies. He finds that the use of such a term by young children is no indication of the innateness of the mental notion of remembering.

A third batch of problems that simulationists challenge the theory theorist to solve is that surrounding the difficulty in the formulation of the laws of folk psychology. In Goldman's words:

> [A]ttempts by philosophers to articulate the putative laws or 'platitudes' that comprise our folk theory have been notably weak. Actual illustrations of such laws are sparse in number; and when examples are adduced, they commonly suffer from one of two defects: vagueness and inaccuracy ... . But why, one wonders, should it be so difficult to articulate laws if we appeal to them all the time in our interpretive practice?[10]

The theory theorist's answer is merely to remind us that the problems of articulating the laws of folk psychology are no different from the problems of articulating the knowledge structures underlying all sorts of cognitive capacities, for example "the principles underlying a speaker's capacity to judge the grammaticality of sentences in his language," "the principles underlying people's everyday judgements about the behavior of middle sized physical objects," or about "their ability to solve mathematical problems, or .... to play chess, etc. "But," they add, "of course, in all of these domains, the theory-theory really is the only game in town."[11]

---

[10] Goldman, "Interpretation Psychologized," 167.

[11] Stich and Nichols, "Folk Psychology: Simulation or Tacit Theory," 49.

1:4 The Simulation Theory: A General View

The simulationists' position is that the arguments based on anomalous precocity, and on the difficulty of the acquisition and the formulation of appropriate laws show that insofar as accounting for our commonsense practices of predicting, explaining, interpreting and describing behavior by ascribing beliefs and desires, the theory theory of folk psychology just does not do the job. A simulation-based theory of what underlies these practices is more plausible than one that posits tacit knowledge of the laws of a folk psychological theory, they say.

The basic idea of the simulation theory is that human beings, when explaining behavior, use the component of their cognitive mechanism devoted to practical reasoning rather than the one devoted to nomological reasoning, the component which entails the deployment of such difficult-to-account-for psychological laws. They present their case by showing how plausible it is as an account of the practice of prediction.

In making a prediction of our own behavior, or of the behavior of others, the practical reasoning component of our cognitive mechanism, the decision-making mechanism which governs the daily interaction of our beliefs, desires, and other intentional and qualitative states, is simply taken 'off-line', some simulationists say. That is, the belief-desire mechanism is disengaged from its actual inputs, - external stimuli and the predictor's own salient beliefs and desires, and is fed pretend input, those beliefs and desires the predictor attributes to his future self or to the person whose behavior he is about to predict. The practical reasoning mechanism then processes these inputs 'off-line', and a pretend decision to behave is generated. This pretend decision is what constitutes the prediction of behavior. In this way, via the process of performing a mental simulation of another, one is able to

reliably predict his behavior. Simulationists hold that the conceptual plausibility of this account is backed by its intuitive appeal; prima facie, it appears to be a better description of what we actually do when we predict behavior than the view that we deploy tacitly-known laws.

Simulationists go on to make the reasonable claim that the mechanism that serves predictive practice also serves explanatory practice. A first step in explaining a behavioral episode is to try to find some beliefs and desires, which, when fed into our practical reasoning mechanism operating off-line will produce a decision to perform the behavior we want to explain. The way we do this is by projection, either total or partial, they say. Total projection is the default mode of simulation and is our first stab at understanding the behavior of those around us, at explaining it to ourselves and others. The way that total projection works in explanation is that after projecting one's entire knowledge base onto the other we compare the decision-to-behave-in-such-and-such-a-manner generated by the internal processing of its components with the behavior to be explained. If they do not match then it is a case where partial or "patched" projection is in order. In this event we make adjustments in the salient facts, norms and values, that is, in some of the data on which our practical reasoning mechanism performs its operations when it is on line, and/or examine the environment from our subject's point of view for different input until the generated decision-to-behave and the actual behavior of the other do match. Typically there will be many sets of adjustments that will render the desired outcome. The set we will choose is the one that requires the least amount of pretending, the shifts in input that are the least different from

those upon which we usually operate.[12] The simulationist view is that we are constantly and automatically projecting our own beliefs onto others in the daily understanding and predictin of their behavior.

Simulationists insist that this operation is not theory-driven but process-driven. Although there are mental simulations that are likely to be theory-based, for example those in which a person seeks to simulate the weather or the economy in the sense of mentally constructing an actual or genuinely feasible sequence of its states, not all mental simulations need to be theory-driven in order to succeed. A simulation of some target systems might be accurate if the agent lacks such a theory. This is likely to be the case if;

> (1) the *process* that drives the simulation is the same as (or relevantly similar to) the process that drives the system, and (2) the initial states of the simulating agent are the same as, or relevantly similar to, those of the target system. Thus, if one person simulates a sequence of mental states of another, they will wind up in the same (or isomorphic) final states as long as (A) they began in the same (or isomorphic) initial states, and (B) both sequences were driven by the same cognitive process or routine.[13]

The general theme of simulationists then is that human beings are able to predict and explain behavior by using the resources of our own minds, in particular their capacity for projection and their decision-making mechanism, to simulate the psychological precursors of that behavior. Insofar as our ability to give commonsense explanations and predictions of behavior is concerned no use is made of the mechanism that serves our capacity for nomological reasoning.

---

[12] Robert Gordon, "The Simulation Theory: Objections and Misconceptions," Mind and Language 7 (Spring/Summer 1992): 18.

[13] Goldman. "Interpretation Psychologized," 173.

Although the tenor and emphasis of the simulationists' account seems to indicate that their view is that the theory theory of folk psychology is a misguided theory tout court, both the thrust of their arguments against the theory theory and the type of theory they are offering as a replacement seem to indicate that what they are objecting to is in fact the strand of the theory theory that deals with the *mechanism* underlying our explanation-giving. What simulationists appear to be arguing is that it cannot be the case that the way our belief-desire mechanism works in the giving of commonsense psychological explanations is that it deploys a tacitly-known theory because positing tacit knowledge of the laws of such a theory raises the problems of anomalous precocity, acquisition and formulation described above. Their objections then are usually understood as being aimed at the thesis that our ability to give commonsense psychological explanations and predictions of behavior can be satisfactorily accounted for by the cognitive science model that accounts for some of our other abilities, the model that holds that such abilities are constituted by tacit knowledge of a theory of the domain in which we have expertise.[14]

That this thesis is the object of their attack on the theory theory seems to be confirmed by the fact that there is no mention in the simulationists' objections to the theory theory to the latter's thesis that what

---

[14] Articles have been published recently with regard to whether or not simulation can account for some of our other cognitive capacities, and thus pose a serious threat to the reigning cognitive science model. Since the purpose of this study is not to examine the value of the simulation theory of our ability to predict and explain behavior as a basis on which to establish a new cognitive science model of our cognitive abilities, we shall leave this issue to one side. We shall assume that the simulationist view about the mechanism responsible for our folk psychological practices is correct. For a critical survey of suggestions for a simulation account of other capacities, see S. Nichols, S. Stich, A. Leslie, and D. Klein, "Varieties of Off-Line Simulation," to appear in Theories of Theories of Mind, ed. P.Carruthers and P. Smith (Cambridge: Cambridge University Press).

constitutes a correct explanation of behavior is the deployment of the appropriate psychological law. It is clear that an objection to the theory theory thesis about the underlying mechanism, does not entail an objection to its thesis about what constitutes a correct explanation. For it could be the case both that the theory theory has got it wrong about the workings of the belief-desire mechanism, that law-deployment plays no role in our giving of commonsense psychological explanations of behavior, and that such an explanation, when it is correct, it is so because there is some kind of underlying psychological law that subsumes the event being explained and which is "internally-represented" and "tacitly-known" in some sense of those terms.

We shall see in Chapter Three that the simulation theory has received serious consideration in the philosophical literature as a real contender to the theory theory of psychological explanation. Our work in that chapter then will be to attempt to extend the simulation theory in order to see if it has the potential to substantiate a claim to being a replacement for the theory theory of folk psychology with regard to its thesis about the nature of psychological explanation. For this purpose we shall assume, (although not without misgivings), that the simulationists' arguments against the theory theory's thesis about the belief-desire mechanism underlying our ability to give commonsense psychological explanations of behavior are well-founded.[15]

---

[15] For a discussion of the difficulties inherent in the simulationists' thesis about the underlying cognitive mechanism see S. Stich and S. Nichols in "Folk Psychology: Simulation or Tacit Theory?" and "Second Thoughts on Simulation," both in Mental Simulation: Philosophical and Psychological Essays, ed. Martin Davies and Tony Stone. (Oxford: Blackwell, 1994). The former paper also appears in Mind and Language 7. (Spring/Summer 1992). The refs. to this paper throughout this work are to Mind and Language.

Before considering the simulation theory, however, we shall examine the theory theory of commonsense psychological explanation in order to see in detail how this account proposes to answer the question which is of interest to us here, what it is that constitutes a correct commonsense psychological explanation of behavior, and to see whether or not there are any well-founded objections to *that* thesis.

Chapter Two

The Theory Theory

The theory that our everyday commonsense explanations of the behavior of ourselves and our fellows are correct because they are partly constituted by the implicit deployment of tacitly-known complex psychological laws can hardly be said to be motivated by intuition. Before we examine different interpretations of this idea, therefore, we shall look at a view which appears to be what motivated this theory and to have earned it the kind of consideration it has received, and continues to receive, in the philosophical literature.

2:1 The Covering Law Model

We mentioned earlier that theory theorists see commonsense psychological explanations as fitting the covering law model of explanation (CLM). Wesley Salmon reminded us recently of the long-lasting impact of the CLM, of its exhaustiveness with regard to the categories of scientific explanation, and of the legitimacy it confers on any explanation it subsumes.[1]

Introduced in the climate of logical empiricism in 1948 by Hempel and Oppenheim in its original form of a deductive-nomological explanation of particular occurrences (D-N explanations), it was eventually extended by Hempel in 1962 and 1965 to include three other categories of scientific explanation: deductive-nomological explanations of general regularities, inductive-statistical explanations of particular facts (I-S explanations) and

---

[1] Wesley Salmon, Four Decades of Scientific Explanation (Minneapolis: University of Minnesota Press, 1989).

deductive-statistical explanations of general regularities.[2] Salmon points out that "According to the received view (i.e. accounts similar to that given by Hempel in his 1965 *Aspects*), ... , every legitimate scientific explanation belongs to one of [these] four sectors."[3] Although not much was written about Hempel and Oppenheim's 1948 paper in the ten years following its publication, by the late 1950s it had some vocal critics, namely Michael Scriven and Norwood Hanson.[4] This did not prevent deductive-nomological explanation from becoming the view about the explanation of particular events the most widely accepted by scientifically-minded philosophers by the 1960s. In spite of several decades of discussion of the fundamental philosophical issues to which it gives rise, the CLM of explanation is still endorsed by many theorists today. It is not unreasonable to think that this long-lasting impact is due to the legitimacy it appears to confer automatically on any explanation that meets the conditions it specifies for correct explanation. Let us look at these conditions in order to see the source of this authenticity.

As our concern here is with the explanation of particular behavioral events, we shall restrict our discussion to the two categories of the CLM that subsume such explanations. These categories are the pertinent D-N

---

[2] Carl Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," Philosophy of Science 15 (1948): 135-75. Carl Hempel, "Deductive-Nomological vs. Statistical Explanation," in Minnesota Studies in Philosophy of Science III, ed. Herbert Feigl and Grover Maxwell (Minneapolis: University of Minnesota Press, 1962), 98-169. Carl Hempel, "Aspects of Scientific Explanation," in Aspects of Scientific Explanation and Other Essays in the Philosophy of Science (New York: The Free Press, 1965), 333-496.

[3] Salmon, Four Decades of Scientific Explanation, 10.

[4] Michael Scriven, "Definitions, Explanations and Theories," in Minnesota Studies in Philosophy of Science II, ed. Herbert Feigl, Michael Scriven and Grover Maxwell (Minneapolis: University of Minnesota Press, 1958), 99-195. Norwood Russell Hanson, "On the Symmetry between Explanation and Prediction," Philosophical Review 68 (1959): 349-58.

model and the I-S model. A D-N explanation of a particular event is a valid deductive argument whose conclusion, the explanandum statement, states that the event to be explained did occur and whose premises, known collectively as the explanans, must include a statement of at least one general law that is essential to the validity of the argument. The premises must also include statements which make assertions about particular facts. Other general conditions of adequacy specified by Hempel and Oppenheim are that the explanans must have empirical content and that the sentences constituting the explanans must be true.[5] The laws required for deductive-nomological explanations are statements of universal form, statements that assert a uniform connection between different empirical phenomena or between different aspects of an empirical phenomenon.[6] This type of law is a statement to the effect that whenever and wherever conditions of a specified kind F occur, then so will, always and without exception, certain conditions of another kind, G.

Although Hempel and Oppenheim recognized in their 1948 paper that not all scientific explanations are based on laws of strictly universal form, it was not until 1962 that Hempel presented a fully worked-out version of another alternative, inductive-statistical explanations.[7] He pointed out that sometimes the connection between the empirical phenomena subsumed by a law will obtain only in a high percentage of cases and not always and without exception. What is appealed to in this case are laws of probabilistic form, laws that state that the probability of

---

[5] Salmon, Four Decades of Scientific Explanation, 12.

[6] Carl Hempel, Philosophy of Natural Science (Englewood Cliffs, N.J. : Prentice Hall, 1966), 54.

[7] Hempel, "Deductive-Nomological vs. Statistical Explanation," 98-169.

conditions of kind F occurring, given that conditions of kind G obtain, is high. Explanations of this type are constituted by inductive arguments whose explanans consist of probabilistic laws and statements of particular fact and which imply the explanandum statement not with deductive certainty, but with a high degree of probability. The laws invoked in scientific explanation, both those of probabilistic form and those of universal form, are called covering laws for the explanandum phenomenon.[8]

What emerges from this overview of these two forms of the CLM is that whatever authority such explanations have appears to be derived from the authority attached to the notion of law and, secondarily, from that attached to the rules of deductive and inductive inference. The notion of law involves notorious problems and is one of the issues that has plagued proponents of the CLM since its inception. Notwithstanding these problems, the fact is that according to the CLM, if an explanation in any domain is to take its place in the ranks among genuine scientific explanations then that domain must be one in which the individuals or events are subsumed by laws.

It seems clear then that for a theorist who holds that the CLM of explanation is the correct model for scientific explanations, if commonsense psychological explanations of behavior are bona fide explanations, they will fit that model. But in order for psychological explanations to be covering law explanations there must be psychological laws which subsume the events to be explained. What seems likely is that the need for psychological laws in order to account for the nature of psychological explanation is motivated by the need for them in the covering

---

[8] Hempel, Philosophy of Natural Science, 51.

law model.[9] It also seems likely that another theory, in fact the cognitive science theory mentioned earlier about what underlies our cognitive abilities, played an important role the development of this idea.

2:2 A Cognitive Science Model

It is often pointed out that the dominant explanatory strategy in cognitive science of positing tacit knowledge of a theory was first used in linguistics by Chomsky in 1965 to account for the adult's ability to produce and comprehend natural language sentences and for the child's ability to acquire a natural language[10] and that a little later Jerry Fodor used this same strategy in giving a philosophical account of the mental competences responsible for behavior.[11] In the following paragraphs we take a glimpse at a paper written by Fodor in 1968, a paper in which he gives an account of our ability to perform specific behavioral tasks. The model of the cognitive organization underlying the ability to perform these tasks which is described in this paper is often cited in relation to the model which accounts for our cognitive ability to give commonsense psychological explanations of behavior, the one that underpins the theory theory of the nature of commonsense psychological explanation, the object of our discussion. Although the relation between these models is usually assumed to be that of identity, there seems to be a significant difference between them. This difference lies in the fact that whereas one posits tacit

---

[9] Stephen Schiffer, "Ceteris Paribus Laws," Mind 100 (January 1991): 11.

[10] Noam Chomsky, Aspects of the Theory of Syntax (Cambridge: MIT Press, 1965).

[11] Jerry Fodor, "The Appeal to Tacit Knowledge in Psychological Explanation," Journal of Philosophy 65 (1968): 627-40 and Psychological Explanation: An Introduction to the Philosophy of Psychology (New York: Random House, 1968).

knowledge of rules or instructions, the other refers to tacit knowledge of the laws of a theory.

In his 1968 paper Fodor undertakes to defend an intellectualist account of our mental competences, that is of the cognitive capacities that allow us to perform activities such as tying one's shoe laces, playing chess, or typing "Afghanistan." What he proposes is that there are mechanisms in the organism which is performing a specific behavior that mediate instructions that describe the psychological functions sufficient for the performance of that behavior. Although instructions for, say, tying ones shoes will designate other behaviors, (e.g. take the left shoe lace in the left hand), which are themselves composed of complex patterns of behavior, each of these complex patterns will be broken down into their component parts which again will be broken down, and this process will be repeated again and again until only elementary functions are specified in the instructions. In Fodor's view, if this is indeed the case, then, "A completed psychological theory must provide systems of instructions to account for the forms of behavior available to an organism, and it must do so in a way that makes reference to no unanalyzed psychological processes."[12] In an effort to clarify this requirement he adds: "Assume that there exists a class of elementary instructions which the nervous system is specifically wired to execute. Each elementary instruction specifies an elementary operation, and an elementary operation is one which the normal nervous system can perform but of which it cannot perform a proper part." These elementary operations, he goes on to say, have no theoretically relevant internal

[12] Fodor, "The Appeal to Tacit Knowledge in Psychological Explanation," 629.

structure, they are operations which the nervous system simply performs and about which it makes no sense to ask the question "how?"

Fodor's suggestion then is that a paradigmatic psychological theory is a list of elementary instructions for producing behavior, and that the vocabulary in which the instructions are formulated should be constrained so that its terms designate only operations that are elementary for the organism in question.

What seems to follow from this view of Fodor's is that if the instructions that compose a *completed* psychological theory will be in terms of the elementary operations necessary for the performance of different behaviors, then the higher level instructions, those specifying such complex functions as taking the left shoe lace in the left hand, compose an *incomplete* psychological theory. We may conclude that this is the type of theory to which scientific psychologists subscribe at the present time and that they can look forward to these functions being broken down and eventually described in terms of basic elementary functions when cognitive science achieves its goal.

But Fodor's view is not only that we perform the behavioral tasks we do because our nervous systems are wired to execute certain instructions. He also holds that we have tacit knowledge of these instructions. With regard to what it is that we tacitly know, Fodor says: "if X is something an organism knows how to do but is unable to explain how to do, and if S is some sequence of operations, the specification of which would constitute an answer to the question 'How do you X?,' and if an optimal simulation of the behavior of the organism Xs by running through the sequence of operations specified by S, then the organism tacitly knows the answer to the question 'How do you X?,' and S is a formulation of the organism's tacit

knowledge." Fodor justifies his inference from facts about the computational operations of an optimal simulation of an organism to facts about the mental life of an organism by appealing to an inductive principle that, he says, when suitably hedged, justifies one to infer like causes from like effects.

The fact that there is no introspective evidence to support this suggestion about our having tacit knowledge of the operations specified by S is of no consequence, says Fodor, as much of our behavior is responsive to internal and external stimuli we are unable to report.[13] He gives an example to demonstrate how it is the case that although some of our internal processes are caused by external stimuli, we are aware of neither the stimuli nor the processes. He points out that if it is true that, say, texture gradients contribute to the causal determination of our visual estimates of depth, and if the central nervous system is indeed the kind of organ most sensible people now think it is,

> then some of the things the central nervous system does, some of the physical transactions that take place in the central nervous system when we make estimates of depth, must satisfy such descriptions as 'monitoring texture gradients', 'processing information about texture gradients', 'computing derivatives of texture gradients', etc. For, on one hand, varying texture gradients does, in fact, cause concomitant variation in depth estimates, and on the other, the central nervous system is the organ whose functioning mediates this causal relation. And it is a point of logic that the central nervous system *can* be the organ that mediates the causal relation between texture gradients and depth estimates only if its operations satisfy the sort of descriptions mentioned above, since whatever physical system mediates our perception of depth must, *ipso facto*, perform whatever operations are nomologically necessary and sufficient for depth perception. If, then, our perceptions of depth are mediated by the central nervous system and if depth estimates vary as a function of the first derivative of texture gradients, then among the operations of the central nervous system

[13] Ibid., 631.

must be some which satisfy such descriptions as 'is computing the first derivative of texture gradient *t*'[14]

Fodor foresees objections to his referring to the specification of internal operations as tantamount to explaining how one performs a certain piece of behavior and to his talking about such operations in terms of our having tacit knowledge of instructions. What makes it reasonable to talk this way, he says, and what constitutes the kind of evidence relevant to assessing claims that some organism tacitly knows some proposition are the following facts: since there are behaviors that an organism knows how to perform and about which it cannot answer the question "How does one perform behaviors of that kind?," and since it cannot know how to perform that behavior unless there *is* an answer to that question, the organism must tacitly know that answer, i.e. the instructions or rules which it executes in performing that behavior.[15] It follows that we must be able to "acknowledge" the principles or rules which control our behavior "in whatever sense of acknowledging is relevant to the operation of that principle in controlling behavior."[16]

Fodor draws the conclusion from these considerations that one kind of requirement that it would be rational to place on a psychological theory would be: "for every behavior an organism knows how to perform, a psychological theory of that organism must supply an answer to the question 'How does one produce behaviors of that kind?' "

---

[14] Ibid., 632.

[15] Ibid., 637-8.

[16] Ibid., 634.

Fodor states clearly at this point that this stipulation entails that those putative psychological theories that are formulated as a set of laws that relate behavior to inputs, to neurological states or to both would *not* qualify as psychological theories. Although such theories might "correctly predict each occurrence of a certain type of behavior, thus explaining the occurrences by subsuming them under general laws, ...this kind of theory would not, in any obvious sense, provide answers to the question 'How does one produce behaviors of that type?,' and hence would fail to satisfy the condition we have stipulated."[17] He points out that the theories proposed by psychological behaviorists who see psychological theories formalized according to the D-N model of the CLM would not qualify as psychological theories according to this stipulation. Psychological theories built on the D-N model are not designed to answer "how"-questions, he says.

So Fodor, in his 1968 paper, draws a clear distinction between theories that account for *how* something occurs and those that predict and explain particular occurrences of behavioral events. It is clear that this does not entail that this view cannot be used to account for *how* we explain and predict behavior. In fact it is widely held that this account of the mental competences responsible for the performance of certain behavioral tasks is identical not only to the present day cognitive science view that cognitive abilities, such as the ability to estimate depth visually, and our mathematical and linguistic abilities, are accounted for by the fact that we refer to a tacitly-known theory, the axioms of which we are hard wired to execute, but that it is also identical to the view that our ability to give commonsense explanations and predictions of the behavior of middle-sized physical objects, and of the behavior of ourselves and our fellows, can be accounted

---

[17] Ibid., 637.

for in terms of our reference to the *laws* of a *commonsense* theory which is represented in our heads and which we tacitly know.

The question that arises is why, in order to account for the *mechanism* responsible for our commonsense explanatory ability, it is deemed necessary to posit tacit knowledge of commonsense laws instead of relying on hard-wired instructions, which are "tacitly known" in some sense, and which are assumed to be sufficient to account for our other abilities. We shall see in the following sections that positing psychological laws incurs all sorts of difficulties, difficulties that are not incurred by saying that the cognitive mechanism responsible for explanation-giving works according to certain rules or instructions. It seems likely then that the transition from rules to laws has a motivation that is independent of the account of the cognitive mechanism responsible for our explanation-giving. In fact it seems likely that the transition has something to do with the covering law model of explanation.

Theory theorists hold that commonsense psychological explanations are covering law explanations. Such explanations are correct, they say, because a psychological law of a tacitly-known folk theory plays the role of covering law premise in the deductive arguments of which they are composed. We have also seen that cognitive scientists hold that many of our abilities can be accounted for by the fact that the component of our cognitive mechanism which is responsible for that ability works according to tacitly-known instructions, the collection of which constitutes a theory. It would be reasonable for theorists giving an account of the nature of commonsense explanation to assume that our ability to give such explanations can be accounted for in the same manner. But if they do make that assumption they are left with the choice of positing a set of tacitly-

known instructions according to which the component of the cognitive mechanism responsible for our explanation-giving works *and* a set of tacitly-known folk laws which play the role of covering law in our explanations *or* of positing one set that will fill both functions. Considerations of simplicity and economy would dictate the latter option and the requirements of the covering law model would make that set laws, not rules or instructions. If this hypothesis is true then it seems likely that the widely-accepted view that folk psychology is a theory composed of tacitly-known internally-represented folk psychological laws was motivated by the fact that it allows commonsense psychological explanations to fit the covering law model of explanation.

This hypothesis about the origin of commonsense psychological laws is supported by the fact that the type of law that is posited by the version of the theory theory that is the most widely accepted today is one that seems designed to incorporate these instructions into our commonsense folk psychological laws.

2:3 Ceteris Paribus Laws with Lower-Level Completions

Now that we have some understanding of the motivation lying behind the theory theory and a suggestion for the origin of its laws, we shall look at different suggestions for the laws themselves. We shall start by examining the popular suggestion just mentioned, that the folk psychological laws that play the role of covering law in our commonsense explanations of behavior are ceteris paribus laws. Since Jerry Fodor is the chief proponent of this view let us look at his theory.

Fodor holds that the tacitly-known psychological laws that we deploy in giving our commonsense explanations of behavior are unstrict ceteris paribus (cp) laws, laws such as: .

If x wants that p, and x believes that not-p unless q, and x believes that x can bring it about that q, and the cetera are paria, then x tries to bring it about that q.

He says that the domain of folk psychology is in the same position with regard to the existence of causal laws that subsume nonphysical properties as any of the special sciences since the laws of these sciences subsume properties such as being a mountain, being an airfoil, and being a lever. He points out that if we accept that there are causal laws in the special sciences involving these nonphysical causally responsible properties, then we have no grounds for rejecting the hypothesis that there are causal laws in intentional psychology.

Fodor explains that the fact that psychological causal laws are hedged does not prevent the psychological properties cited in them from being nomologically sufficient for behavior. Although he admits that one cannot have it "... *both* that special laws only necessitate their consequents ceteris paribus *and* that we must get Bs *whenever* we get Ms?"[18], he holds that what one can have still allows a substantive claim to be made, one that allows serious scientific business to be done. He says that if it is a law that Ms cause Bs ceteris paribus, then it follows that you get Bs whenever you get Ms and the ceteris paribus conditions are satisfied. He maintains that

---

[18] Jerry Fodor, "Making Mind Matter More," in A Theory of Content and Other Essays (Cambridge: MIT Press, 1990), 152.

there is a difference between this claim and the vacuous claim that Ms cause Bs except when they do not. The supposed difference, we take it, lies in the assumption that since ceteris paribus conditions are determinate, although they may be undefined and undefinable, it is possible to recognize situations in which they are satisfied, while in the vacuous claim, the fact that no determinate conditions are referred to at all prevents one from recognizing situations in which the law obtains. He concludes that even if it is true that causes need to be covered by laws that necessitate their consequents, it does not follow that they need to be covered by strict laws. Hedged laws necessitate their consequents in worlds where their ceteris paribus conditions are satisfied, he says.

Fodor sees no problem in requiring that the cp conditions be satisfied in order for a law to hold in a given case even though these conditions cannot be enumerated. Whether they can be enumerated or not does not prevent their being discharged, he says. "What determines whether a given law can cover a given event is whether the law is determinately satisfied by the event." It is not necessary that the conditions in which it is satisfied are well defined. What is important is that we can recognize situations in which these conditions are satisfied, says Fodor.

Having argued for the existence of causal laws that subsume nonphysical intentional properties and for the lawlikeness of psychological ceteris paribus sentences, Fodor says: "I digress to remark that hedged laws can play the same role as strict ones in covering law explanations, so long as it's part of the explanation that the ceteris paribus conditions are satisfied."[19] He remarks that if the explanation "you got a B because you had an M and it's a law that you get a B whenever you get an M" is

[19] Fodor, "Making Mind Matter More," 154.

satisfying then so too ought the explanation "you got a B in world w because you had an M in world w and its a law that ceteris paribus you get a B whenever you get an M, and the ceteris paribus conditions were satisfied in world w." It is a mistake to think that covering law explanations depend on causes being covered by strict laws. "Strict laws and hedged laws with satisfied ceteris paribus conditions operate alike in respect of their roles in covering law relations and in respect of their roles in covering law explanations. Surely this is just as it should be: strict laws are just the special case of hedged laws where the ceteris paribus clauses are discharged *vacuously*; they're the hedged laws for which 'all else' is *always* equal."

Fodor attempts to clarify the nature of the propositions that these commonsense laws express, the terms in which these laws would be formulated if they could be made explicit. As we saw above, Fodor's suggestion for the type of proposition expressed by our commonsense generalizations linking behavior to beliefs and desires is one that has the following form: If x wants that p, and x believes that not-p unless q, and x believes that x can bring it about that q, and the cetera are paria, then x tries to bring it about that q. It is because the term "ceteris paribus" is an idiom that it needs to be explicated in order for the full implications of his suggestion to come to light,

Fodor's general thesis is that sentences containing cp clauses express complete propositions although the part of the proposition expressed by the cp clause cannot be cashed out in the same vocabulary as that used to express the rest of the proposition. However there is no reason to doubt, he says, that the cp clauses can be discharged in the vocabulary of some lower level science. Ceteris paribus laws are non-basic and any non-basic law, any law that is not a law of fundamental physics, is to be understood as

requiring mediation by a physical mechanism. They are macro-level laws requiring micro-level mediation. He suggests that the cp clauses of such laws can have the effect of existentially quantifying over these mechanisms.[20] In this case "As cause Bs ceteris paribus" can mean something like "There exists an intervening mechanism such that when it's intact, As cause Bs." He remarks that a standard way to account for the failure of a ceteris paribus condition is to point to the breakdown of the intervening mechanism.

It follows from this general thesis about the cp clauses of nonbasic laws that psychological cp laws, although expressed in purely intentional terms, have implementing physical mechanisms that underpin them and which are referred to idiomatically by the cp clause. "If it's nonbasically lawful that Ms cause Bs, there is always a story to tell about how (typically, by what transformations of microstructures) instantiating M brings about the instantiation of B." Thus there is a physical mechanism in the organism, describable in the vocabulary of, say, neurology, bio-chemistry or physics, that makes it the case that the satisfaction of the conditions described in the antecedent of the intentional law is non-superfluously nomologically causally sufficient for the occurrence of the event described in the consequent. If that mechanism fails, those intentionally-described conditions will not be satisfied. He says, "Nonbasic laws rely on mediating mechanisms which they do not, however, articulate (sometimes because the mechanisms aren't known; sometimes because As can cause Bs in many different ways, so that the same law has a variety of implementations)."[21] We take this to indicate that the complete propositions that psychological cp

---

[20] Ibid., 155.

[21] Ibid., 155.

laws express idiomatically, quantify over the beliefs, the desires *and* the mediating mechanism that are together non-superfluously nomologically sufficient for the occurrence of the behavioral event. The question of whether or not such propositions do indeed determine anything worth calling psychological laws has been brought under close scrutiny by Stephen Schiffer.

Schiffer suggests that one plausible way to interpret the thesis about the completion of psychological cp laws in the vocabulary of a lower level science is to see the antecedent of the conditional of which such laws are purportedly composed as constituted by a conjunctive condition, one conjunct of which describes (say) a neurological condition and the other a psychological condition.[22] In this case the satisfaction of both conjuncts will be causally sufficient for the satisfaction of the consequent although the satisfaction of neither conjunct alone will be. He discusses whether or not there are true propositions that such conditionals might express and if there are, whether or not such propositions are, or determine, laws that can be used as the covering law premise in psychological explanations.

Schiffer points out that the interpretation of this suggestion in terms of a conjunctive condition is coherent only on the understanding that, for example, in the case of the human organism, it is a neurological realization of the psychological condition that conjoins with the implementing neurological condition to make a causally sufficient conjunctive condition for the behavioral event subsumed by the law. He argues, however, that in this case the proposition expressed by the conditional cp sentence would not be true. Since the only requirement on being a realization of a psychological state is that it have a certain functional role and that it stand

---

[22] Schiffer, "Ceteris Paribus Laws," 5.

in a certain relation to distal objects and properties, it is easy to imagine realizations that can cohere with the basic neurological condition and not form with it a causally sufficient condition for the occurrence of the behavior. For a proposition which specifies this kind of a conjunctive sufficient condition for behavior to be true, what would need to be shown is that every physically possible realization of the psychological conjunct of the condition that could cohere with the unspecifiable, but nevertheless specific, neurological condition would form with it a sufficient condition for the behavioral event to occur.

Schiffer suggests that in modifying Fodor's view in order to take into account the multiple realizability of psychological states, one should take into account the fact that even if there is no *lower* level completing condition with which every nomologically possible realization of the psychological condition can cohere in order to make a causally sufficient condition for the behavior, there might well be some *same* level condition that can conjoin with each realization to form with it a nomologically sufficient condition for the occurrence of the behavioral event.[23] He proposes that Fodor's suggestion that the proposition expressed by "Ms cause Bs cp" is

> (EC)(C is a condition specifiable in the language of some more basic science (e.g. neurophysiology) & it's a law that Ms cause Bs when C is satisfied)[24]

be replaced by

---

[23] Ibid., 6.

[24] Ibid., 5.

> *Ms cause Bs cp iff for each realization D of M there is a same level condition C such that D-&-C is nonsuperfluously causally sufficient for a B event.[25]

Schiffer points out that a minor problem with this suggestion is that since there may well be realizations of M for which there is no C, and since some realizations of M might be among the defeating conditions referred to in the cp clause, the proposition as it stands might still not be true. He suggests that this can be remedied by replacing "each realization D of M" with "each of sufficiently many realizations D of M". It is not necessary for the truth of the proposition "Ms cause Bs cp" that every realization of M have a completing condition C, but just that most of them do.

However in clarifying the suggestion for the proposition expressed by a psychological cp sentence and in adapting it so that it will accommodate the difficulty presented by the multiple realizability of psychological states a major problem with the initial suggestion has been revealed, says Schiffer. The above proposal as to what is meant by a psychological cp sentence does not warrant talk of a psychological covering law, he says. If the right hand side of * is what is meant by "Ms cause Bs cp," then a true substitution instance of the right hand side would be the putative psychological law. But it is evident, says Schiffer, that a substitution instance of "For each (or each of sufficiently many) realization(s) D of M there is a same level condition C such that D-&-C is nonsuperfluously causally sufficient for a B event" will be no kind of law and consequently not a law that could figure as a covering law in CLM explanations. Traditionally covering laws must be either laws of strict

---

[25] Ibid., 6.

universal form or laws of probabilistic form. It is clear that a substitution instance of the above would be neither.

Schiffer next considers the possibility that this proposition, although not qualifying as a law itself, might determine a law, and that that law might be of the kind needed for commonsense psychological explanations of the covering law model.[26] Cp laws in general, and psychological cp laws in particular, can play the role of covering law in covering law explanations just as strict laws do, we were told by Fodor. In this case commonsense psychological explanations would be true substitution instances of the following:

Ms cause Bs, cp
Mx
Cetera are paria

Bx

But if we accept that the complete proposition expressed by "Ms cause Bs cp" is the right hand side of *, then since the right hand side of * is not a law itself, if we are dyed-in-the-wool covering law model theorists, we might want to say that the right hand side of * will determine a law. Adopting the position of such a theorist for a moment, Schiffer explores the full story underlying an explanation using "Ms cause Bs, cp" as the covering law premise when the right hand side of * is understood as being

[26] The following was taken from notes on a lecture on this topic by Stephen Schiffer in October, 1991.

the proposition that "Ms cause Bs cp" expresses. He suggests that such an explanation would consist of the following argument:

(a) Mx.

(b) Since Mx, there is a realization D of M such that Dx.

(c) *For each (or each of sufficiently many) realization(s) D of M there is a same level condition C such that D-&-C is nonsuperfluously causally sufficient for a B event.

(d) Since Bx, it's reasonable to suppose that Cx also where C is a relevant same-level condition.

(e) (ED)(EC)(Dx & Cx & Dx-&-Cx) is causally sufficient for Bx

---

Bx

Although Schiffer concedes that since M does occur in the premises of this argument one might want to give M something of an explanatory role, he denies that any explanatory role that M might play here is one that comports with the covering law model of explanation. According to that model, to have an explanatory role is to appear in a law that can be used to explain. And it is clear that M does not appear in any law in the above explanation. The only law here, Schiffer points out, is the neurophysiological law alluded to at (e) and M does not appear in it. So if M plays an explanatory role at all it is not the kind of explanatory role that the covering law model requires since it does not appear in a behavior-explaining law. We should not be surprised at this, adds Schiffer, since not only does it not seem plausible that our commonsense explanations entail anything like what is expressed in (a)-(e), but neither does it seem plausible

that such a story is entailed by the accepted fact (stated at (e)) of the multiple realizability of psychological states and their realization by more basic properties.

Fodor understands the implications of Schiffer's argument against the existence of psychological cp covering laws to be the rejection of nonbasic explanations on the grounds that they would be overdetermined by basic explanations.[27] Schiffer assures us however that Fodor assumes wrongly in this case and that his argument does not imply anything about overdetermination or the rejection of nonbasic explanation. His point, he repeats, is that Fodor's suggestion for what constitutes nonbasic explanations is not plausible. Such explanations cannot consist of deductive arguments whose covering laws are psychological cp laws, since there are no laws that could play the required role in commonsense psychological explanations.

The cp-law view of the theory theory just described, when considered in relation to a version that preceded it, can be construed as a "weak" version of the theory theory of psychological explanation. A much stronger version of the theory theory was presented by David Lewis in 1972. In his view the psychological laws of our folk theory, the covering laws implicitly referred to in our commonsense explanations of behavior, are our commonsense platitudes. Since our ordinary mental terms are the theoretical terms of the folk theory, that is they derive their meaning from the causal roles which are described by the folk theory, and our mental states are defined in terms of what occupies those causal roles, the laws of

---

[27] Jerry Fodor, "Hedged Laws and Psychological Explanation," Mind 100 (January 1991): 19-34.

the theory include only those platitudes that are common knowledge.[28] Although there are many well-known objections to this thesis, we have not rehearsed them here since this strong version of the theory theory is defeated on independent grounds. Indeed the refutation of the strong thesis that our commonsense platitudes play the role of covering law in our commonsense explanations of behavior is entailed by the refutation of the much broader or weaker thesis which allots that role to psychological cp laws

The situation so far, as we see it, is that the most popular candidate for the psychological laws proposed by theory theorists, psychological cp laws with lower-level completions, does not appear to qualify since it does not meet a condition which they deem essential for such laws: they are not of a kind that comports with the covering law model of explanation. There remains another likely candidate, or category of candidates, however, candidates which seem to represent our pretheoretical notion of what a psychological law would look like.

2:4 Probabilistic Laws with Same-Level Completions

In this section we shall investigate the question of whether or not there are probabilistic psychological laws of a type that would accommodate theory theorists' needs with regard to playing the role of covering law premise in covering law explanations. We shall examine the hypothesis that there are psychological cp laws that have same-level probabilistic completions such

---

[28] David Lewis, "Psychophysical and Theoretical Identifications," in Readings in the Philosophy of Psychology 1 ed. Ned Block (Cambridge: Harvard University Press, 1980), 207-15.

that they qualify to play the required role in deductive-nomological explanations of the covering law model.[29]

In examining a proposition in view of its qualification for the role of probabilistic covering law four steps seem to be indicated. Let us suppose that a probabilistic psychological law will be a substitution instance of the following formula:

There is such-and-such a probability that an M event will cause a B event.

It seems that in order to consider whether or not a sentence that is a substitution instance of this formula would indeed express a law, we must first discuss whether or not the proposition expressed by the sentence playing that role would be true. Since the answer to this question will depend on the nature of the notion of probability expressed by the probabilistic sentence, in order to consider different propositions for their qualification as probabilistic laws we must first examine the notion of probability each candidate proposition is assumed to project. Only then will it be appropriate to examine the proposition with regard to its truth, and, if it qualifies on this account, with regard to its potential for being a psychological law. Finally, if the proposition in question has passed all these tests, we must examine it to see whether or not it can play the requisite role in CLM explanations.

Before beginning this step-by-step examination however, a certain clarification needs to be made. One might be inclined to think that the

---

[29] This possibility was suggested by Schiffer in his paper "Ceteris Paribus Laws" and developed more fully in his seminar in the Fall of 1991. Much of what follows in this section is influenced by Schiffer's comments in those lectures.

complete proposition expressed by a substitution instance of the above formula, a substitution instance such as

There is such and such a probability that if a person really wants something then he will take steps to get it

would be

It is likely/reasonable for someone in the same epistemic position as oneself to have some non-negligible degree of expectation that a person who really wants something will take steps to get it."[30]

However although such a proposition might be true, it does not appear to be a candidate for lawhood. It seems clear that the intended probability in this case is epistemic or subjective. And if this is so, then it does not seem that it could be the proposition expressed by a law. For how could it be the case that a proposition about what it is reasonable for one person to believe be used in an explanation of what another person does, or that a proposition about what it is reasonable for a person of a certain kind to believe explain why I did something? On the face of it, at least, such a suggestion appears implausible. It would seem that if our behavior is to be explained by reference to a probabilistic law then some notion of objective probability is required. The notions of probability that we examine in this chapter are of this kind.

---

[30] Schiffer, "Ceteris Paribus Laws", 8.

2:4:1 Laws of Single-Case Objective Chance

Let us first look at the hypothesis that there are platitudinous, cp psychological laws of objective chance. In this case "Ms cause Bs, cp" will be read as "Every M event has such-and-such an objective chance of causing a B event". An example of the kind of proposition that might be expressed by a psychological law of objective chance is

If a person really wants something then there is such-and-such an objective chance that he will take steps to get it.

Before examining such a proposition for its truth and for its qualification for lawhood, let us first look at what is involved in the notion of objective chance.

David Lewis explains that we do have such a concept in "A Subjectivist's Guide to Objective Chance": "We think that a coin about to be tossed has a certain chance of falling heads, or that a radioactive atom has a certain chance of decaying within the year, quite regardless of what anyone may believe about it and quite regardless of whether there are any other similar coins or atoms." [31] So we have a concept of probability such that the objective chance that an event has of occurring is something that it has absolutely. It is not relative to a particular reference class or to a knowledge situation, but to everything, past, present and future. The objective chance that a particular event has of occurring supervenes not just on "the actual arrangement of qualities, but on that plus all the chances

---

[31] David Lewis, "A Subjectivist's Guide to Objective Chance," in Philosophical Papers 2 (Oxford: Oxford University Press, 1986), 90.

there are, at various times, of that arrangement continuing one way or another."[32]

Lewis explains that he thinks of objective chance as attaching in the first instance to propositions and he makes a wary claim about the type of proposition involved: "the chance of an event, an outcome, etc., is the chance of the truth of the proposition that holds at just those worlds where that event, outcome, or whatnot occurs... I have foremost in mind the chances of truth of propositions about localized matters of particular fact - a certain toss of a coin, the fate of a certain tritium atom on a certain day- but I do not say that those are the only propositions to which chance applies."[33] This is the type of context then in which he considers that the notion of probability projected is the most likely to be the notion of objective chance.

Lewis holds that objective chance, although not the same thing as credence or frequency, is time dependent and world dependent. He sees it as a function of three arguments. "To a proposition, a time, and a world it assigns a real number." The objective chance of that proposition being true at time *t* in world *w* is some number *r*. It is a function that maps the types of particular states of affairs described in the proposition onto a number. Lewis does have misgivings about the general notion of objective chance. He remains hopeful however that they can be dispelled.

Now that we have clarified, to some extent at least, the notion of objective chance, let us return to our query about the truth of the proposition expressed by a psychological probabilistic sentence when the

---

[32] Lewis, Philosophical Papers 2, xiv.

[33] Lewis, "A Subjectivist's Guide to Objective Chance," 90-91.

probability involved is understood as objective chance, the truth of the proposition expressed by the following sentence:

If someone really wants a million dollars, then there is such-and-such an objective chance that he will take steps to get it.

At first glance, however, if objective chance is the notion of probability projected here it appears doubtful that such a proposition can be true. It seems unlikely that there can be a number that represents the objective chance of this proposition being true, since it seems that there will be no single objective chance that every token of a mental event that has the property of being the desire for a million dollars, could have of causing behavior that could be described as "taking steps to get it." Consider a particular event token that has the property of being that desire. This same event token could have another property, and it could be the case that having this latter property precludes it from causing that behavior. In this case this particular token of the mental event with the property of being the desire for a million dollars has no objective chance of causing the behavior. However if we take a different token of that mental event, a token that does cause the behavioral event, the taking of steps, then the objective chance that that token has of causing that behavioral event will be high. So it seems that there can be no single objective chance that a type of mental event can have of causing a type of behavioral event. And in this case a proposition that says that a specific type of mental event has such and such an objective

chance of causing a type of behavioral event must be false. And if it is false it cannot be a candidate for lawhood.[34]

2:4:2 Laws of Relative Frequency

In the category of objective notions of probability, besides single case objective chance, there are also relational notions. A relational notion of objective probability is a probabilistic relation between properties. The idea is that the probability of one property, say G, occurring, given that the property F occurs, is some number r. This kind of probability can be represented as a function that maps pairs of properties onto numbers:

$$P\ (G/F) = r$$

The most common way to interpret the kind of probability involved here is in terms of relative frequency. So if the notion of probability projected by a probabilistically completable psychological cp law is relative frequency, a likely reading of "Ms cause Bs, cp" is "The relative frequency with which an M event causes a B event is r".

Let us start our examination of this suggestion by looking at the notion of probability involved. One way of understanding the notion of relative frequency is in terms of the *actual* relative frequency with which two properties co-occur. An example of a sentence expressing a proposition that incorporates this kind of probability is:

---

[34] Hartry Field pointed out that this problem might be a problem for all laws of objective chance.

the actual relative frequency of cancer-getters among cigarette-smokers in some finite sample is found to be some number r.

Let us consider the proposition that the above sentence about smokers expresses. It seems, prima facie at least, that although such a proposition might be a true proposition about the proportion of cancer-getters among smokers in some sample, there is reason to doubt whether or not it is, or determines, a law. For if it is a law that the actual fraction of cancer-getters among smokers is such-and-such then it seems that that proportion, being a fact subsumed by a law, should not depend on accidental contingencies. On the other hand it seems clear that the proportion of cancer-getters among smokers is contingent on facts about the age at which people start smoking, whether or not they exercise, how much nicotine is in the brand of cigarettes they smoke, etc. In other words, if a mark of a law is that it support counterfactuals, then propositions that state the actual relative frequency with which two events co-occur do not appear to qualify on that account.

There is, however, another way to understand the relational notion of probability in terms of relative frequency, one that says that the relative frequency of Gs among Fs *in the limit* is r. In this case the proposition will specify the frequency with which one will find property G instantiated, given that property F is instantiated, if one were to consider sufficiently many cases for one to reach the limit at which the proportion remains the same. In this case the frequency with which one will find cancer-getters among cigarette-smokers when one has examined enough cigarette-smokers to reach the limit will be r.

However, whereas it seemed likely that propositions about actual relative frequency were true, although not lawlike or law-determining, it is not clear that propositions about frequency in the limit are even true. In order for a proposition about the proportion of cancer-getters among smokers in the limit to be true it would have to be determined what is to be understood by "in the limit." If we suppose that what is meant is "if the number of smokers were to increase without end", then a suggestion for the true proposition expressed by a statement of relative frequency in the limit would be "If the number of smokers were to increase without end, then in the limit such and such proportion of them would develop cancer." But it seems that there would be no single number that would truly represent that proportion since once again the number will depend on contingent facts about who smokes and who does not, etc. It seems that there is no determinate answer to the question of what percentage of smokers would get cancer were the number of smokers to increase without end.

These observations about the ill-suitedness of the notion of relative frequency to be one projected by a law are prima facie. They have been used in this section to demonstrate why we think it is not likely that probabilistically completable psychological cp propositions will express or determine laws that are qualified to enter into D-N explanations, when the notion of probability involved is the notion of relative frequency.

It is tempting at this point to argue that although probabilistic generalizations of relative frequency might not qualify as laws of the kind that can play the role of covering law in D-N explanations, that these remarks about the notion of relative frequency being unsuited to be that projected by laws might be justified when it is a matter of the posited laws

playing *that* role, the remarks are not justified when it is a question of their being qualified to play a different role, the role of covering law in inductive-statistical covering law explanations. This of course was the role that was originally allotted to them by Hempel.

In fact, one might support this thesis by adopting a position that Nicholas Rescher puts forward in another context, that is by holding that laws are made, not found.[35] The suggestion is that whether or not a proposition is lawful is a matter of decision. Rescher points out that since empirical evidence can support only a general claim about *actual* cases, any assertion of lawfulness regarding a generalization goes beyond the empirical evidence. He concludes that lawfulness is not something we discover about generalizations but something we impute to them: "It reflects the use to which we want to put a generalization in organizing and systematizing our knowledge of the world." If we adopt this position, we can see Hempel's use of statements of relative frequency as probabilistic laws as reflecting his decision to speak of them as laws in order to accommodate the I-S model of the CLM.

2:5 Inductive-Statistical Explanation

Although our examination of what seemed to be the most likely candidates for laws that could play the role of covering law in commonsense psychological explanations of the deductive-nomological kind has shown that none appear to qualify, if one accepts Rescher's view of laws, there seems to be a possibility that has remained unexplored by covering law theorists, viz., that such explanations are inductive-statistical explanations using probabilistic commonsense psychological laws of relative frequency

[35] Nicholas Rescher, Scientific Explanation (New York: Free Press, 1970), 97-121.

as covering laws. The problem with this suggestion however lies in the nature of inductive statistical explanation itself.

One difficulty with I-S explanations is their intrinsic fallibility. The conditions which this form of the CLM specifies do not appear to be sufficient for correct explanation. We recall that explanations of this type are constituted by inductive arguments whose explanans consist of probabilistic laws and statements of particular fact. The laws state that the probability of conditions of kind F occurring, given that conditions of kind G obtain, is high. The premises imply the explanandum statement not with deductive certainty, but with a high degree of probability. So it is consistent with the satisfaction of the conditions specified by the I-S model that the conclusion of the inductive argument of which such an explanation is constituted be false, that in spite of the high probability of the explanandum event occurring in certain circumstances and in spite of the fact of those circumstances obtaining, the event referred to in the explanandum does not in fact occur. Although this problem can be avoided by adding the stipulation that the conclusions of such arguments be true, there is another, more serious, problem incurred by their intrinsic fallibility.

It can also be the case that *both* the premises *and* the conclusion of the inductive argument that constitutes the explanation are true and that the explanation is still false. This type of situation occurs when the facts cited in the explanans are irrelevant to the occurrence of the explanandum event: when, for example, although it is true that almost all colds clear up within a week after the administration of vitamin C, and it is true that Jones, who had a cold, took vitamin C and that his cold cleared up within a week, it is not true that his cold cleared up within a week *because* he took vitamin C.

Since the probability of colds clearing up within a week whether or not the patient takes vitamin C is high, the high probability of colds clearing up after taking vitamin C is not relevant to the probability of colds clearing up after a week and invoking such a fact does not constitute an explanation or part of an explanation of the fact that Jones' cold cleared up after a week.[36] The point is that because of the fallibility of I-S explanations, their dependability as bona fide explanations is questionable in a way that that of D-N explanations is not.

Besides these problems of fallibility, there are also difficulties surrounding the objectivity of I-S explanations. As a solution to a problem that arose with the I-S model of 1962, that of its explanatory ambiguity, Hempel introduced in his 1965 paper the requirement of maximal specificity. Hempel was concerned that the I-S model allows it to be the case that an inductive argument that strongly supports its conclusion may be transformed by the addition of a new premise consistent with the original premises into an argument that strongly undermines that conclusion.[37] As he believed that the reason for this is that all of the relevant evidence is not included in the statements in the explanans, Hempel suggested the following remedy: if we think of K as the knowledge situation represented by the set K of all statements accepted in that situation, then "In formulating or appraising an I-S explanation, we should take into account all that information provided by K which is of potential *explanatory* relevance to the explanandum event; i.e., all pertinent statistical laws, and such particular facts as might be connected, by the

---

[36] Salmon, Four Decades of Scientific Explanation, 58-9.

[37] Hempel, "Aspects of Scientifc Explanation," 394-5.

statistical laws, with the explanandum event."[38] Hempel goes on to declare that "The preceding considerations show that the concept of statistical explanation for particular events is essentially relative to a given knowledge situation as represented by a class K of accepted statements. Indeed, the requirement of maximal specificity makes explicit and unavoidable reference to such a class, and it thus serves to characterize the concept of 'I-S explanation relative to the knowledge situation represented by K.'"[39] Thus the requirement of maximal specificity resulted in making the I-S form of the CLM essentially epistemically relevant.

That correct inductive explanations are relative to a knowledge situation in the particular way that Hempel suggests in this paper is much disputed in the philosophical literature. Pretheoretically it does seem that there can be correct explanations of events quite independently of the fact of whether anyone knows them or not. Nor is it generally accepted that making explanations relative to a knowledge situation solves the problem of explanatory ambiguity. It was even argued by Coffa in 1974 that, "to accept Hempel's thesis of epistemic relativity amounts to accepting the claim that there are no inductive explanations."[40]

Since the objections to the I-S model laid out here are generally considered to be well-founded it is not surprising that theory theorists offering an account of commonsense psychological explanation ignore the inductive-statistical model of the CLM and attempt to explain why these explanations, when they are correct, are of the deductive-nomological type.

---

[38] Ibid., 400.

[39] Ibid., 402.

[40] Salmon, Four Decades of Scientific Explanation, 72. Salmon refers to Albert Coffa's paper "Hempel's Ambiguity," Synthese 28 (1974): 141-63.

Let us review the situation so far. Our discussion in this chapter has revealed that the most likely candidates for the laws needed by theory theorists to play the role of covering law in deductive nomological explanations, do not in fact qualify. The candidates discussed were suggestions for various types of cp law, those that could be completed in the vocabulary of some lower-level science and those that could be completed probabilistically at the level of psychology.[41] If the arguments that we have used against the existence of psychological cp laws apply, as is likely, to cp laws across the board, then besides undermining the theory theory account of the nature of commonsense psychological explanation, we will have raised questions about the nature of other types of explanation which are thought to use this type of law and to be covering law explanations. The explanations affected will be other types of commonsense explanation, explanations that account for the behavior of middle-sized objects for example, and special science explanations. Denying cp sentences their lawhood, therefore has dire consequences for all but the explanations of basic physics.

The disappearance of *commonsense* cp laws will also have some affect on the cognitive science model of our abilities mentioned in section 2:2. Whereas the version of that model that posits a rule-following mechanism and purports to account for our linguistic and mathematical abilities and our ability to estimate depth visually will remain unaffected by

[41] A suggestion for cp laws which has not been discussed here is Horgan's and Tienson's suggestion that they are "soft" laws. See Terence Horgan and John Tienson, "Soft Laws," in Midwest Studies in Philosophy XV (1990): 256-79. Since Horgan and Tienson make no reference to the set of soft psychological laws constituting a tacitly known folk theory, we do not consider it to be a version of the theory theory of psychological explanation. This type of law is discussed in a different context in Chapter Four however.

the fact that no commonsense cp laws exist, what appears to be a later development of that model, the one that posits tacitly-known laws of a commonsense theory and purports to account for the abilities to give commonsense psychological explanations of behavior and to give commonsense explanations of the behavior of middle-sized objects, will be affected. Not only will this later development no longer have a raison d'etre if the theory theory of the nature of commonsense psychological explanation is wrong, but if there are no commonsense cp laws, then law-deployment cannot be the mechanism that underlies those abilities. A requirement on a theory of explanation that purports to replace the theory theory of commonsense psychological explanation then is that besides accounting for the nature of commonsense explanations of the behavior of people and of the behavior of middle-sized objects and for the nature of special science explanations, it must also either provide, or be compatible with, some cognitive science account of our explanatory abilities.

Chapter Three

The Simulation Theory

In this chapter we describe the theory which Robert Gordon and Alvin Goldman offer as a replacement for the theory theory of folk psychology, the simulation theory. This theory was mounted as a challenge to the theory theory by Gordon in 1986 and by Goldman in 1989. While Gordon presents it as an hypothesis about the nature of folk psychology, Goldman discusses it in the context of an examination of three different theories about the procedure that guides our interpretation of others in terms of the mental states we judge responsible for their behavior, a discussion in which he gives pride of place to the simulation theory. A theme common to their papers is that human beings are able to predict and interpret behavior by using the resources of their own minds to simulate the psychological precursors of that behavior. There is no reference by simulators to a theory in the performance of these simulations, they say. While Gordon holds that we simulate by identification in the imagination with the person whose behavior we are predicting or interpreting, Goldman describes the simulations as being "process-driven" and not "theory-driven." Pretend inputs are fed into the simulator's decision-making mechanism whose processes then run "off-line" so that no action is actually produced but a decision to perform some action is the end of the "run," the output of the processing. When we predict, interpret and explain behavior we are simulators, not theorizers, in their view.

Gordon's and Goldman's discussions about mental simulation revolve around the issue of the method we use, or which is implemented by our

decision-making mechanism, when we attribute mental states in the performance of our folk psychological practices of interpretation, prediction and explanation. Their view is that the mechanism described by the simulation theory is a more accurate description of what underlies our actual practice than that described by the theory theory. While Gordon is confident that he can extract philosophically fundamental accounts of mentality from the simulation theory,[1] Goldman is pessimistic with regard to the possibility of simulation's giving constitutive accounts of anything. His need of the simulation procedure stems from the necessity to provide his introspectionist model of the mind with an account of third person mental ascriptions.[2]

These different views of the power of the simulation theory represent a strong and a weak replacement thesis with regard to the theory theory. When the simulation theory is thought of as a worthy replacement for the theory theory across the board, the implication is that it has the resources to replace not only its thesis about the mechanism responsible for our folk psychological practices, but also the more philosophical strands of the theory theory, including those concerned with the nature of mental states, the nature of our mastery of mental concepts, the meaning of mental terms and, what is of particular interest to us, the nature of psychological explanation. On the other hand, when simulation is offered just as a replacement for that strand of the theory theory to do with the mechanism, the assumption is either that there is some version of the theory theory with whose different strands the simulation procedure is compatible, or that

---

[1] In particular, Gordon says that he hopes to address the issue of the nature of psychological explanation in a forthcoming paper. In correspondance, June 1994.

[2] Alvin Goldman, "The Psychology of Folk Psychology," Behavioral and Brain Sciences 16 (1993): 15-28.

there is some other non-simulation-based hypothesis that is compatible with the simulation account of our practices and that will replace some or all of the philosophical strands of the theory theory. While Goldman appears to favor this last position, Gordon is convinced that the simulation theory can sustain a strong replacement claim.

It is important to mention here the position of these simulationists with regard to the mechanism underlying our other cognitive abilities. Gordon and Goldman appear to agree with the dominant explanatory strategy of cognitive science with regard to the mechanism underlying many of our abilities, viz., that it consists of the deployment of a tacitly known theory. In particular they appear to agree that our ability to give commonsense explanations of the behavior of middle-sized objects, of folk physics, follows this model.[3] We saw in Chapter One that their objections to the theory theory of commonsense psychological explanation seem to indicate that they hold that our ability to give commonsense psychological explanations of behavior cannot depend on a law-deploying mechanism because of the problems incurred by the anomalous precocity that young children would need in order to deploy such complex laws, the difficulty of accounting for their acquisition and the apparent impossibility of their formulation. We have two comments to make about their position with regard to the supposed difference of the mechanism underlying our ability to give commonsense psychological explanations.

First, there is some confusion as to whether or not simulationists really are talking about the *mechanism* that underlies our ability to give commonsense psychological explanations in their discusssions of the theory

---

[3] Robert Gordon, "Reply to Stich and Nichols," Mind and Language 7 (Spring/Summer 1992): 90. Alvin Goldman, "In Defense of the Simulation Theory," Mind and Language 7 (Spring/Summer 1992): 108.

theory. The objections they raise against the theory theory seem to indicate that their concern might be epistemological rather than to do with the cognitive organisation responsible for our explanation-giving. But whether or not young children are capable of knowing such complex laws, and how we acquire knowledge of them seem to be different problems to the problem of whether or not the belief-desire mechanism works according to laws.

Secondly, their acceptance of the theory theory of commonsense explanations of folk physics seems to imply that the problems of acquisition etc., do not arise for commonsense laws concerning the behavior of middle-sized objects. But again the fact, if it is a fact, that we can account for the acquisition of the laws of folk physics, and that their deployment occurs only when a certain late stage in development is achieved, and the "fact" that the mechanism underlying the ability to give that kind of commonsense explanation is indeed law-deployment are different. What we need to know in order to assess the merit of the simulation theory as a theory about the mechanism that underlies our commonsense psychological explanations of behavior is what it is that makes it likely that the mechanism responsible for our ability to give this kind of explanation is *different* from the mechanism that underlies our ability to give explanations of folk physics such that the latter is law-deploying and the former is not.

Notwithstanding the seeming confusion of issues, simulationists see commonsense psychological explanations of human behavior as representing a model of explanation that is different from the model that accounts for many other types of explanation; some aspects of commonsense psychological explanation cannot be accounted for

satisfactorily by the theory theory, they say. It seems clear however that neither the rejection of the theory theory view that we have knowledge of such complex laws, nor the rejection of the thesis that the mechanism responsible for our giving commonsense psychological explanations of behavior is law deployment entail the view that commonsense psychological explanations are not covering law explanations. The view that we cannot have knowledge of complex commonsense psychological laws and the view that the belief-desire mechanism does not involve the deployment of laws are both compatible with the view that commonsense psychological explanations are covering law explanations, that they are correct when the events being explained are subsumed by some type of psychological law. The tenor of the discussion surrounding the simulation theory, however, seems to indicate that it is not generally taken merely as a better suggestion for how we get to know which beliefs and desires to ascribe or for the cognitive mechanism underlying the practice of giving commonsense psychological explanations, but that it is thought of as potential replacement for the theory theory of psychological explanation across the board.

In this chapter we shall set out the simulation theory as it seems to have been understood in the literature, that is, as a theory about the cognitive mechanism which underlies the production of our commonsense predictions, explanations and interpretations of behavior. This will allow us to examine the theory with regard to the resources it might have to substantiate a replacement claim which is strong enough to offer a replacement for the theory theory's account of the nature of psychological explanation.

The relation between the simulation theory and a weaker replacement claim will not concern us here. For in the event that it were

indeed intended to be merely a theory about how we get to know which beliefs and desires to attribute in explanation or a theory about the mechanism underlying our explanation-giving, it would either invoke some version of the theory theory to account for what it is that constitutes a correct explanation, (and we have already argued that it is unlikely that any laws exist that could play the required role), or else it would invoke a view of the nature of psychological explanation that does not involve simulation, and thus be beyond the scope of this chapter which is to be devoted to discussing the potential of the simulation theory to challenge the theory theory's account of psychological explanation.

Before we turn to laying out the simulation theory of the cognitive mechanism responsible for the practice of giving commonsense psychological explanations of behavior, it is worth our while to look briefly at the historical context in which it is proposed. Gordon and Goldman are usually discussed in the literature as presenting more or less the same point of view about simulation. On our understanding, however, their views differ substantially. We suggest that a possible reason for the two views having been conflated and treated as one by their critics is that their proponents mistakenly claim a common heritage. Both Gordon and Goldman, when introducing their view of simulation, refer to its relation to a notion used by theorists in the hermeneutic tradition as a method of inquiry into the moral or social sciences, the notion of *verstehen*. They also both refer to Collingwood's view of history as the re-enactment of past experiences as the direct precursor of their own view. However it appears to be the case that while Collingwood holds one view on this topic, the *verstehen* philosophers have at least two different views, only one of which, the one we shall refer to as their earlier view, bears a resemblance

to Collingwood's suggestion. And it seems that while Gordon might be justified in claiming descent from the early Collingwoodian version of *verstehen*, it seems that Goldman is justified in claiming it only from the other, later development of the *verstehen* view, a development which Collingwood explicitly rejects. This difference in views enters into our discussion of the potential of the simulation theory to provide a theory of the nature of psychological explanation that makes no reference to the covering laws of a theory.

3:1 The Early *Verstehen* View

The characteristic of the hermeneutic tradition of thought which is pertinent to the present discussion is its reactionary stance to positivist philosophy of science at the end of the nineteenth century.[4] The philosophical descendants of Comte and Mill held that the same scientific method should be used in the scientific investigation of all fields and that the ideal to be used throughout was that used in the exact natural sciences, in particular mathematical physics. They also held that scientific explanation consists in the subsumption of individual cases under hypothetically assumed general laws of nature, including human nature. Many German philosophers, historians and social scientists, however, among whom are notably Wilhelm Dilthey in his early work[5], Karl Jaspers and Max Weber, rejected both the methodological monism of positivism and its view of explanation.

---

[4] Georg Henrick von Wright, Explanation and Understanding (Ithaca, N.Y.: Cornell University Press, 1971), 4.

[5] Wilhelm Dilthey, Einleitung in die Geisteswissenschaften. (Introduction to the Sciences of Mind). 1883.

A distinction was made by these non-positivist theorists between the method to be used in exploring the exact sciences, and that to be used in investigating the human sciences, or the *Geisteswissenchaften*, as they came to be called after this term was used in a translation of the title of one of Mill's works. They held that a different approach is justified since the aims of these investigations are different. While the scientists working in natural science are aiming at giving an explanation of the phenomena in their domain by observing their relation to each other and bringing them under a general law of nature, the aim of those involved in the second type of investigation is to "understand" the phenomena in their field by discovering the thought expressed by the event. In the latter case no further inquiry into causes of such events is necessary since "cause" refers here not to the kinds of occasions when such and such happens, but to the particular thought that caused the particular event. The hermeneutic philosophers readily admitted that since every explanation can be said to further understanding, "understanding" when used in this sense is the goal of all sciences. The type of understanding to which they are referring, and which they call "*verstehen*," is not the *goal* of an investigation into a particular field but a *method* of investigating a field. When the term is used in this manner it has a psychological connotation that it does not have when used in relation to scientific explanation. These anti-positivist philosophers thought that understanding, when considered as a method for investigating the moral sciences, was a form of empathy or "re-creation in the mind of the scholar of the mental atmosphere, the thoughts and feelings and motivations, of the objects of his study."[6]

---

[6] Von Wright, Explanation and Understanding, 6.

Let us look briefly at one interpretation of what was involved in the general notion of *verstehen*. In a paper published in the American Journal of Sociology in 1948, Theodore Abel, a positivist philosopher, undertakes to describe and delimit the operation of *verstehen*.[7] He uses examples to show that we frequently link two disparate observed events by introducing unobserved intervening events. The way we do this is by performing the operation of *verstehen* whose characteristic feature is "the postulation of an intervening process 'located' inside the human organism, by means of which we recognize an observed - or assumed - connection as relevant or 'meaningful.' *Verstehen*, ..., consists of the act of bringing to the foreground the inner-organic sequence intervening between a stimulus and a response."[8]

Abel tells us that this operation involves three steps: internalizing the stimulus, internalizing the response and applying behavior maxims. The arbitrary procedure involved in internalizing the stimulus consists in imagining the emotions that may have been evoked by the impact of a given situation. Gestures or exclamations during the impact may provide clues to which emotions are aroused but whether they do or not, we proceed by imagining how we would have been affected by such an impact. Abel says that our ability to do this imagining depends on our ability to describe a situation by categorizing it and our ability to evoke a personal experience which fits into the category. When it comes to internalizing the behavioral response we also use the imagination. Although no specific techniques are known which permit a definite association between a certain behavior and a

---

[7] Theodore Abel, "The Operation called *Verstehen*," American Journal of Sociology 54 (1948). Reprinted in Readings in the Philosophy of Science, ed. Herbert Feigl and May Brodbeck (New York: Appleton-Century-Crofts, Inc. 1953), 677-687.

[8] Abel, "The Operation called *Verstehen*," 682.

certain internal state, a motive or an emotion, we use our personal experience to infer the motive of an act from the known or observed modification in behavior that it produces. Among the possible inferences, we select those that can be seen as the solution to the problem created by the impact of the stimulus event.

Once we have internalized the stimulus and the response we can apply a behavior maxim which links the two feeling states in a uniform sequence and implies a functional dependence between them. It is the application of the behavior maxim which links the two internal states, which makes the connection between the two observed factors relevant. Thus we 'understand' a given human action if we can apply to it a generalization based upon personal experience. Behavior maxims are generalizations of direct personal experience derived from introspection and self-observation. They are not the same thing as universal psychological generalizations, maintains Abel.[9]

3:2 Collingwood, Dilthey, and Historical Re-Enactment

Whereas the German *verstehen* philosophers' concern was to account for the assumed connection between observed outer events by using one's imagination and own experience to postulate an intervening inner process, Collingwood's main concern was to understand historical events by discovering the thoughts that lay behind them. In 1946 Collingwood wrote that historical knowledge is "knowledge of what mind has done in the past, and at the same time it is the redoing of this, the perpetuation of past acts in

---

[9] Abel, "The Operation called *Verstehen*," 683.

the present."[10] Collingwood holds that the method used in gaining knowledge of minds past is also that used in gaining knowledge of minds present.

If it is by historical thinking that we re-think and so rediscover the thought of Hammurabai or Solon, it is in the same way that we discover the thought of a friend who writes us a letter, or a stranger who crosses the street. Nor is it necessary that the historian should be one person and the subject of his inquiry another. It is only by historical thinking that I can discover what I thought ten years ago, by reading what I then wrote, or what I thought five minutes ago, by reflecting on an action that I then did, which surprised me when I realized what I had done. In this sense, all knowledge of mind is historical. The only way in which I can know my own mind is by performing some mental act or other and then considering what the act is that I have performed. If I want to know what I think about a certain subject, I try to put my ideas about it in order, on paper or otherwise; and then having thus arranged and formulated them, I can study the result as an historical document and see what my ideas were when I did that piece of thinking: if I am dissatisfied with them I can do it over again.[11]

Our knowledge of human psychology then, our own and others', is gained, according to Collingwood, by using the powers of our own minds to re-think past thoughts. With regard to awareness of one's actual mental states, he says, "On what I have only begun and am still doing, no judgment can as yet be passed."[12] He makes no mention of prediction.

---

[10] R.G. Collingwood, The Idea of History (London: Oxford University Press, 1946), 218.

[11] Collingwood, The Idea of History , 219.

[12] Ibid.

Although Collingwood's view seems similar to the *verstehen* view as portrayed by Abel, and espoused by the early Dilthey[13], Collingwood explicitly rejects a method of historical inquiry developed by Dilthey in his later works.[14] He agrees with Dilthey that historical data offer the historian the occasion "for reliving in his own mind the spiritual activity which originally produced them," and thus that "It is in virtue of his own spiritual life, and in proportion to the intrinsic richness of that life, that he can ... infuse life into the dead materials with which he finds himself confronted," that, in fact, historical knowledge is inner experience.[15] He also agrees that since to be Caesar does not constitute knowledge of him, a further step is needed. However Collingwood points out that in acknowledging this latter fact Dilthey had come up against the question of how there could be "a knowledge, as distinct from an immediate experience, of the individual," a question that the early *verstehen* philosophers had not recognized.

Collingwood does not agree with the step that Dilthey takes to solve this problem. Dilthey reasons that since it is only by psychological analysis that I come to know the structure of my own personality, so an historian, if he is to understand the past which he is reliving, must understand, by means of psychology, the new experiences which are incorporated into that structure. Collingwood argues, however, that psychology is not history but science, a science constructed on naturalistic principles, whose proper domain is that of the non-cognitive mental states.[16] To say that historical

---

[13] Abel cites an early work of Dilthey's as one of his sources: W.Dilthey, Ideen ueber eine beschreibende und zergliedernde Psychologie (Leipzig: Teubner, 1884).

[14] Wilhelm Dilthey, Gesammelte Schriften VII (B.G. Teubner: Leipzig, 1914-27).

[15] Collingwood, The Idea of History, 172.

[16] Ibid., 173.

knowledge, now thought of as knowledge of the contents of one's own mind, is only intelligible when conceived of in terms of the laws of psychology is to fall back on the positivist view that the only kind of knowledge is scientific knowledge.

Collingwood's view is that understanding oneself does not involve the use of psychological generalizations. He says that when he recognizes in himself an immediate feeling of discomfort and asks himself why he has this feeling he might answer that question by reflecting that that morning he received a letter criticizing his conduct in a valid manner. What he is doing in this case, he says, is "recognizing in its detail a certain individual event, or series of events, which are already present to my consciousness as a feeling of discomfort or dissatisfaction with myself."[17] He adds, "To understand that feeling is to recognize it as the outcome of a certain historical process. Here the self-understanding of my mind is nothing else than historical knowledge." In the same way, he says, when I relive in my own mind a certain experience of Caesar I am not being Caesar but am myself and know that I am myself. I incorporate his personality into my own not by confusing myself with him but by distinguishing myself from him and at the same time making his experience my own. Since the past is not a dead past but lives on in the present, the historian's knowledge is knowledge of the past in the present, "the self-knowledge of the historian's own mind as the present revival and reliving of past experiences."[18]

This glimpse at the recent forerunners of the simulation theory shows that the early *verstehen* theorists and Collingwood, just as we shall see that their successors do, deny the use of psychological generalizations in

---

[17] Ibid., 174.

[18] Ibid., 175.

explaining human behavior. Although the concern of the *verstehen* theorists and Collingwood was to discover and describe the correct method to be used by theorists investigating certain fields, we shall see that the simulation theorists's concern is to describe the method we, the folk, actually use in giving our everyday commonsense accounts of why people perform some action or other. Our review has also shown a difference between the early *verstehen* view, which closely resembled Collingwood's, and the later one developed by Dilthey. We shall find this difference echoed in the present-day views discussed below

3:3 Simulation and our Folk Psychological Practices

In his 1986 paper about simulation Gordon claims to offer an account of the nature of folk psychology. His thesis is that folk psychology is not a theory, but that it might be "the capacity for practical reasoning, supplemented by a special use of a childish and primitive capacity for pretend play."[19] He illustrates his point by showing how we exercise these capacities when we give predictions of human behavior. What follows in the first part of his paper is a detailed description of the process that takes place within us, of the method that we unwittingly use, in making these predictions. Let us look at his argument for simulation being the best account of our predictive success.

He builds his case for prediction on the reliability of our own declarations of immediate intention, the reliability of utterances such as "I shall now pour some coffee." It is more likely, Gordon says, that our declarations of immediate intention are the result of practical reasoning about what we should do in a particular case, rather than the result of

---

[19] Robert Gordon, "Folk Psychology as Simulation," 171.

nomological reasoning about what people usually do in cases of that type. For one thing our declarations of immediate intention are more reliable than any inference from a law of typical causation and an inventory of our relevant beliefs and desires could warrant. This is shown in part by our capacity to predict our behavior even when the action we choose is atypical. Another factor that indicates that nomological reasoning is not the kind of reasoning used in the first person case is that we are incapable of making an inventory, let alone a reliable inventory, of the beliefs and desires that would be subsumed by a law since we do not keep tabs on all of the relevant propositional attitudes. Gordon uses these observations as prima facie support for the view that our declarations of immediate intention are the result of practical reasoning based on salient facts, salient norms and values and a background of other facts, norms, and values that we are unable to list exhaustively.[20]

Gordon extrapolates from the hypothesis that practical reasoning is used to predict our own immediate behavior to propose that it is used in the prediction of our behavior in hypothetical situations also. On these occasions simulation of the self in counterfactual circumstances occurs, he says. While our knowledge base of salient facts, norms, and values and, of course, unlistable background facts remains the same, a pretend "fact," an imagined modification of the actual world, is fed into our practical reasoning system. This pretend fact is then processed by the system and the self-simulator merely awaits the outcome. This processing is what Gordon means by "practical simulation." It is a simulated "deciding-what-to-do," a process that is fully automatic and usually not introspectible by the

---

[20] Ibid., 160.

organism in which it is taking place. The outcome will be a simulated decision about what to do, a decision upon which the self-predictor will not act immediately and perhaps not act at all. In other words, this simulated decision will be a hypothetical prediction, "a prediction of what I would do in the specified hypothetical circumstances, other things being as they are."[21]

From here it is but a short step for Gordon to predicting the behavior of others by simulation. Gordon emphasizes that simulation in this case is not merely a matter of deciding what oneself would do in a situation in which another finds himself. That would be just another kind of self prediction and irrelevant to the prediction of the behavior of others. He does not deny that this kind of reasoning by analogy sometimes takes place but that is not the kind of simulation that underlies our interpretive, explanatory or predictive ability in his view. Simulation when used for these latter purposes requires either total or partial projection of the self onto another.

Gordon first describes the case of total projection. Members of a close-knit community who are in spatiotemporal proximity can often predict each other's behavior by projecting themselves in imagination onto the other without making any adjustments at all in their knowledge base, not even for a slight difference in spatiotemporal location. In this case what is projected onto the other is the totality of the simulator's beliefs, the salient facts, norms and values and the unlistable background facts that make up his knowledge base. The predictor/simulator however is not able to discriminate the individual components of what is projected. The projection occurs automatically whenever triggered by a situation calling

---

[21] Ibid., 161.

for the anticipation of another's behavior. For the duration of the simulation the simulator becomes the other such that "I" and "now" would not have their ordinary references were they to be uttered by him. However these and other indexicals do retain certain ties, says Gordon, those they have because one's " perceptions, memories, actions and emotions are keyed to [one's] egocentric map."[22] What do shift, he says, are "the locations and vectors of environmental features on [one's] egocentric map--that is, the mental map in which things and events are represented in relation to [oneself], here and now, ..."

The next step in this process of predicting the behavior of a very similar other is for the simulator to search the environment for features that would engage his own emotional and practical responses. Gordon points out that it is features that are emotionally and practically relevant to him that set his practical reasoning mechanism in motion. This is a more plausible suggestion, he insists, than that it is the nomological reasoning mechanism that is triggered by those features of the environment that are relevant to the simulator because they could be the referents of terms used in the antecedent of some tacitly-known law.[23] In the case of the simulation of others, just as in the case of simulation of oneself in hypothetical conditions, the simulated decision-what-to-do need not be translated into action. However it is the simulated process of deciding-what-to-do which terminates in a decision to do so-and-so which enables one to predict another's behavior.

---

[22] Gordon, "The Simulation Theory: Objections and Misconceptions," Mind and Language 7 (Spring/Summer 1992): 30-1.

[23] Ibid., 15.

Gordon admits that the above strategy will give successful predictions of behavior only when the other is relevantly similar to oneself. Often however when one projects oneself in imagination onto the other, one has to make adjustments to one's baggage of salient facts, norms and values in order for one's predictions to have the best chance of coming out right. This type of situation calls for patched projection. One of the first adjustments that is often made is to project oneself into the other person's shoes, into his spatiotemporal location. We then search the environment from this vantage point for a stimulus. If one's emotional and practical responses are still not engaged, one attempts to project oneself into his mind, to make adjustments in one's own attitudes, one's skills or lack of them, one's idiosyncrasies etc. in an effort to match his. Then it is into the practical reasoning system with this adjusted set of salient data, beliefs and desires, that we feed facts about the environment, facts that we imagine we would apprehend were we in the shoes of the person whose behavior we are trying to predict.

Goldman's emphasis is on the interpretation and intentional description of others - how we come to describe them as having some mental concepts as opposed to others. He works from the assumption that we have mental concepts and declares that his inquiry is into a subset of them, beliefs about beliefs. For him simulation is the procedure we use as a means of getting to know which beliefs to attribute to others in order to interpret their behavior and in order to give a description of them in intentional terms. Goldman states clearly at the beginning of his discussion that his question is "how does the (naive) interpreter arrive at his/her judgments about the mental attitudes of others?"[24]

---

[24] Goldman, "Interpretation Psychologized," 162.

Like Gordon, Goldman builds his case for simulation being the mechanism underlying in our third person ascriptions on the plausibility of its being what guides our predictive practice. However, in contrast to Gordon, in Goldman's view the simulator does not project himself in terms of his undiscriminated mind-set onto the other, nor does he become the other in his imagination such that the references of "I" and of other indexicals change.

When the simulator attempts to interpret the behavior of the other in a particular situation, he identifies with *himself* in that situation. Thus the indexical "I" is a rigid designator whenever it is uttered by him. There is no projection. He remains *himself* with his own baggage of salient and non-salient beliefs and desires, and in his imagination he immediately makes adjustments among them for the different situation in which *he* now finds himself. This is the first step. The assumption is made at the start that the other has the same basic likings and needs as oneself and that the other has the perceptual beliefs that oneself would have in that situation.[25] The next step for the simulator is to imagine what *he* himself would do in this situation, that is with this adjusted set of beliefs, desires, hopes, fears etc. This is a simulated *self*-prediction. The third step is for him to infer from *himself*, the model, to the other. In other words he uses the knowledge of his own adjusted set of mental states, those that serve as input to the system, and his knowledge of the outcome of their interaction, the simulated decision-what-to-do, as a model from which to infer the probable behavior of the other. Thus it appears to be the case that, for Goldman, a hypothetical self-prediction and a prediction of the behavior of another amount to one and the same thing.

---

[25] Ibid., 170.

Both Gordon and Goldman assume that if simulation of another's intentional states, or of one's own in counterfactual circumstances, is a correct description of the mechanism that underlies our predictions of behavior, then that will also be the mechanism that underlies the practices of explaining or interpreting behavior.

Although Gordon speaks mostly of the practice of prediction in his 1986 paper and of how it is underpinned by simulation, he does speak briefly here of the interaction of explanation and prediction in the process of hypothetico-practical reasoning which, he says, underlies them both. However, it is in his paper of 1992 on this topic that he describes the use of simulation in our explanatory assignment of mental states.

Gordon says that we are constantly automatically projecting our own beliefs about the environment onto others in our daily understanding of their ordinary behavior. Total projection, the default mode of simulation, is our first stab at understanding their behavior, at explaining it to ourselves or to others. The way that total projection works in explanation is that after projecting one's entire knowledge base onto the other we compare the decision-to-behave-in-such-and-such-a-manner generated by the internal processing of its components with the behavior to be explained. If they do not match then a change has to be made in one's knowledge base and it is a case where "patched" projection is in order. In this case we make adjustments in the salient facts, norms and values, some of the data on which our practical reasoning mechanism usually performs its operations, and/or examine the environment from our subject's point of view for different input until the generated decision-to-behave and the actual behavior of the other do match. Typically there will be many sets of adjustments that will render the desired outcome. The set we will choose is

the one that requires the least amount of pretending, the shifts in input that are the least different from those upon which we usually operate. As Gordon puts it, we will follow the principle of least pretending.[26]

Goldman merely mentions explanation in passing after saying how attractive simulation is as a view of our predictive practice: "The simulation procedure can also be used for explanatory, or 'retrodictive', assignment of mental states," he says.[27] If I wanted to explain a surprising move of yours in chess, he continues, I would infer from your move a new strategy that you might be using. In order to check whether my choice of strategy was likely to be correct I would see if this inferred strategy is one that I might arrive at from your presumed antecedent states and one that would lead me to make the move that you have just made. Adjustments would have to be made of course for individual differences. And, he says, "there is no assumption here that people are always successful or optimal simulators." His conjecture is that "simulation -- whether explicit or implicit-- is the fundamental method used for arriving at mental ascriptions to others."

3:4 Theoretical Commitments of the Simulation Theory

Gordon states repeatedly that his view of the simulation procedure does not involve a model. We do not introspect knowledge of own mental states and cognitive organization and use the latter as a model from which to infer the organization and behavior of another, according to Gordon. How we would behave in another's situation is irrelevant to how the other would behave in

---

[26] Gordon, "Folk Psychology as Simulation," 164 ; "The Simulation Theory: Objections and Misconceptions," 22.

[27] Goldman, "Interpretation Psychologized," 169.

that situation. When we project onto the other we are not mentally *transferring* our own states to him. What we are doing is identifying with him in the imagination, thereby *transforming* ourselves into him.[28] A decision to behave made in this context is a decision about how *he* would behave, maintains Gordon. So simulation is not simulation of oneself under counterfactual circumstances followed by inference to the other.

Gordon holds that the capacity for practical simulation, both in the case of total and partial projection, operates at a sub-verbal level most of the time. In these cases we are unaware of the practical reasoning process, of the pretend inputs and of course of their subdoxastic counterparts, and of the process of projection. This explains why we can anticipate the actions of others even though we often cannot say what we anticipate or why.[29]

Gordon insists that none of these tacit goings-on either presupposes, or is constituted by, general knowledge of the laws of a theory of human psychology. He recognizes that we do use laws or generalizations about human behavior in common discourse. These are not universally known and mechanically applied however but are rules of thumb, or heuristics, learned, applied and interpreted within the context of simulation, he says.[30] It is only after one has constructed certain patterns of behavior from many previous simulations that one may develop inductively-formed generalizations which can be used for prediction and explanation without

---

[28] Robert Gordon, "Simulation without Introspection or Inference from Me to You," in Mental Simulation: Philosophical and Psychological Essays, ed. Martin Davies and Tony Stone. (Page refs. are to a manuscript), 5.

[29] Gordon, "Folk Psychology as Simulation," 170.

[30] Gordon, "Folk Psychology as Simulation," 166; "The Simulation Theory: Objections and Misconceptions," 16.

recourse to simulation. But, he says, the process of simulation will still be used in the application of these generalizations; it is simulation that enables us to see whether or not it is likely that the conditions referred to in the antecedent of any behavioral generalization or law are met in the case at hand. And this includes those conditions referred to in the ceteris paribus clause. By identifying with another in imagination one is in the best position to judge whether or not the other will have the beliefs and desires mentioned in the generalization and whether or not all else is equal in this case, or whether some intervening factors will prevent the event referred to in the consequent from occurring. Thus it is simulation that enables us to know if the law is applicable in a particular case. Gordon's conclusion is that although generalizations may be called upon for the purposes of prediction and explanation, it is our capacity to simulate others which is fundamental to our predictive and explanatory ability, both because it is the basis upon which those generalizations are formed and because our skill at formulating and using these laws depends on our capacity to simulate.

Besides simulation obviating the need for knowledge of laws for prediction and explanation, simulation also obviates the need for our having a prior understanding of the concepts of belief and desire in order to predict and explain the behavior of others, says Gordon. We can simulate and understand others long before are capable of deploying complex laws and attributing beliefs and desires. What we project onto the other in the case of total projection is our own undiscriminated mindset and we can achieve this projection without being aware of either the individual components of what is being projected or of the mechanism which operates on those components. In response to the argument that any adjustments made to our mindset when total projection fails are in the form of

consciously added individual beliefs etc., and that for this intentional concepts are needed, Gordon points out that there are mechanisms for simulation in place at an early age which work without concepts and that this indicates that we have the capacity to make such shifts in imagination without having the concepts of individual mental states. The acquisition of intentional concepts is the result of many previous simulations, says Gordon, just as the acquisition of the knowledge of behavioral generalizations is.[31]

The similarity between this description of the method we supposedly use in our practices and the prescription of the early *verstehen* theorists that the method of empathic understanding be used in investigating fields of inquiry to do with human behavior is now clear. Both views eschew the appeal to generalizations which link types of external stimuli or mental state to a typical behavioral output in favor of the largely tacit reference to, or the factoring in of, particular facts or states of affairs concerning the behaver, a process carried out by simulation.[32]

Both Gordon and Goldman argue at length that the belief-desire mechanism is isomorphic in the human species. Goldman says that simulation cannot account fully for our success in predicting human behavior given the complexity of the human organism unless one makes the additional assumption "that other people whose behavior we predict are psychologically very similar to ourselves."[33] We can successfully simulate

---

[31] Robert M. Gordon, "Reply to Perner and Howes," Mind and Language 7 (Spring/Summer 1992): 98.

[32] Another point on which Gordon's view resembles the earlier one is in his reference to the use of personal behavior maxims. See Gordon, "The Simulation Theory: Objections and Misconceptions," 30.

[33] Goldman, "Interpretation Psychologized," 177.

the internal operations of others because they are largely homologous to our own. Goldman suggests that "people's predictions of other people's behavior, based heavily on attributions of content, are so successful because people operate with the same set of fundamental cognitive constraints." The psychological constraints that we have in common bear on the psychological operations that generate propensities to group together certain sets of elements in the visual and temporal domains, he says.

The aspect of Goldman's view of simulation that differentiates it clearly from Gordon's and that lays it open to the charge of theory-dependence is the assumption that we take our own decision-making mechanism "off-line" when we predict, explain or interpret the behavior of others. His view is interpreted as saying that we use our own system as a model from which we infer what is going on in the cognitive system of another person in certain circumstances.

Critics of the simulation theory point out that the assumption that we are able to predict and explain the behavior of others because of the homologousness of our belief-desire mechanisms, and because of our capacity to take our own system off-line, and then infer from the goings-on in it to the goings-on in another, presupposes extensive self-understanding and that this is just as problematic as other-understanding. In fact, as Paul Churchland puts it, "If one is to be able to apprehend even the first-person intricacies at issue, then one must possess a conceptual framework that draws all of the necessary distinctions, a framework that organizes the relevant categories into the appropriate structure, a framework whose taxonomy reflects at least the more obvious of the rough nomic regularities

holding across its elements, even in the first person case. Such a framework is already a theory."[34]

However Goldman holds that knowledge of a theory is not necessary for simulation. We can simulate the target system by letting our own system run off-line without knowing how either system works because the simulation is process-driven and not theory-driven, he says. Although he acknowledges that some mental simulations do rely on the simulator having knowledge of a theory, for example those that mimic the weather and those that mimic the economy, he insists that the kind of mental simulation that he and Gordon are talking about does not. All that is needed is that the process that drives the simulation in the simulator be relevantly similar to the process that drives the system to be simulated and that the initial states of the simulator be relevantly similar to those of the target system.[35] When this is the case the reasoning operations in the simulator will generate similar successive states to those that the target's mechanism will generate.

Churchland's objection to Goldman's view reminds us of Collingwood's objection to Dilthey. Indeed Dilthey was criticized by Collingwood for assuming that in order to understand past events by re-thinking the thoughts that lay behind them, the historian had to have recourse to the principles of psychology to understand what was going on in his own mind. Although there is no mention by Dilthey of a subsequent inference to the other, as the aim is to understand the thoughts of those who performed past actions, it seems clear that the inference was intended. And, just as Collingwood considered this move equivalent to falling back on the

---

[34] Paul Churchland, A Neurocomputational Perspective: The Nature of Mind and the Structure of Science (Cambridge: Cambridge University Press, 1989), 120.

[35] Goldman, "Interpretation Psychologized," 173.

positivist practice of using the same method of inquiry for history as for the exact sciences, viz., the application of laws that subsumed the events to be explained, so the theory-theorist critics of the simulation heuristic insist that if one has a model then in order to understand the model itself and to justify the inference from the model to the other, one must have knowledge of the principles of a theory of folk psychology.

3:5 Developmental Evidence

To support their view, Gordon and Goldman both turn to data from developmental psychology, and in particular to a set of experiments conducted by the psychologists Wimmer and Perner in 1983 and suggested as an experimental paradigm by the philosophers J. Bennett, D. Dennett, and G. Harman. Gordon appeals to this empirical data in order to substantiate his thesis about belief attribution; he says, "...to attribute a belief to another person is to make an assertion, to state something as a fact, *within the context of practical simulation.* Acquisition of the capacity to attribute beliefs is acquisition of the capacity to make assertions in such a context."[36] There are two sets of experiments that support his conclusion about belief attribution, says Gordon.

In the Wimmer-Perner experiments of 1983 children of different ages observed the puppet-child Maxi put a chocolate in a box and go out to play.[37] While he is out his mother transfers the chocolate from the box to a cupboard. When the five-year-olds were asked where Maxi will look for the chocolate when he returns, they answered correctly that he will look in

---

[36] Gordon, "Folk Psychology as Simulation," 168.

[37] H. Wimmer and J. Perner, "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception," Cognition 13 (1983): 103-28.

the box. However when the three to four-year-olds were asked the same question, they replied that Maxi would look in the cupboard. Both age groups were then asked where Maxi would tell his brother the chocolate was if he wanted to mislead him. While the five-year-olds said that he would tell him it was in the cupboard, the four-year-olds indicated the box.

In a related experiment in 1985, Baron-Cohen, Leslie and Frith studied the ability of autistic children and children with Down's syndrome to perform this sort of task.[38] Whereas only 20% of the autistic children gave the correct answers, 86% of the Down's syndrome children did. Since both types of impairment involve mental retardation, and on average autistic children seem to suffer retardation to a lesser degree than Down's syndrome children, the disparity was attributed to a cognitive deficit specific to autistic children; the lack of the capacity for pretend play.

Gordon's interpretation of the Maxi results is that since the younger children frame their predictions in terms of what they themselves know, they either lack the concept of belief altogether or the capacity to attribute beliefs that differ from their own. These, and other experiments, show that at the age of about four years children develop the ability to make allowances for what another is not in a position to know. In gaining this ability they gain the concept of belief, says Gordon, and at the same time the capacity to attribute beliefs that differ from their own. Gordon suggests that instead of accounting for this increase in ability by assuming that it is not until the age of four that the child has fully internalized a system of laws and generalizations concerning belief, this could be accounted for if it were the case that "the child of four develops the ability to make assertions,

---

[38] S. Baron-Cohen, A.M. Leslie, and U. Frith, "Does the Autistic Child have a 'Theory of Mind'?" Cognition 21 (1985): 37-46.

to state something as fact, *within the context of practical simulation.* That would give her the capacity to overcome an initial egocentric limitation to the actual facts (i.e. as she sees them)."[39] This hypothesis is supported by the results obtained from the experiments with the autistic children, says Gordon, since they strongly indicated that only those children who can engage in pretend play can master the concept of belief.

Goldman appeals to the very same evidence to support his thesis about what is responsible for a child's interpretational ability. The fact that children who gain the ability "to clearly distinguish between someone else's belief state and reality" in the normal course of development can predict correctly, and the fact that so many autistic children, who are known to lack that capacity, cannot, might indicate that "pretend play is a preliminary exercise of a mechanism the primary function of which is the simulation of real people's mental states."[40] Notwithstanding his recognition of the fact that the psychologists who did this experiment suggest that their findings might be due to the fact "that autistic children as a group fail to acquire a 'theory of mind'," he concludes that it is likely to be due to the fact that the mechanism for simulation is what underlies interpersonal interpretation, prediction and communication.

The interpretation that Gordon and Goldman give to the results obtained from these experiments provoked a huge response in the literature where different interpretations were offered and the conflicting results of other experiments cited. Gordon and Goldman each responded to these objections, either discrediting the soundness of the methodology used in the

---

[39] Gordon, "Folk Psychology as Simulation," 169.

[40] Goldman, "Interpretation Psychologized," 175.

experiments or presenting still other results in support of their view. We continue below by describing a small part of the follow-up debate, in particular, some of that surrounding the interpretation of the Maxi results.

Paul Harris, a psychologist and proponent of the simulation theory, points out the interesting fact that the false belief tasks described above were originally designed to rule out the possibility of simulation's being the procedure used in making successful predictions.[41] Since it is impossible to tell whether a subject is using a tacit theory of mind or the simulation procedure when making predictions that could be based on what he himself knows, thinks or feels, it was thought that by creating a situation in which he and the other have access to different facts, that a successful prediction would have to result from knowledge of a theory since projection of oneself onto the other would result in a false prediction of what the other would do. Harris points out the ironical fact that contrary to the designers' intentions, the results of this experiment can be easily be accounted for by a more sophisticated type of simulation, Gordon-style patched projection, for example.

Harris looks at a possible simulation interpretation of the results of two sets of experiments reported by Wimmer, Hogrefe and Sodian in 1988 concerned with children's knowledge assessment.[42] One set had to do with the causal link between seeing and knowing, and the other with the link between inference and knowing. With regard to the first set, the results indicated that it is between the ages of three and five years that children

---

[41] Paul Harris, "From Simulation to Folk Psychology," Mind and Language 7 (Spring/Summer 1992): 123.

[42] H. Wimmer, J. Hogrefe, and B. Sodian, "A Second Stage in Children's Conception of Mental Life: Understanding Informational Accesses as Origins of Knowledge and Belief" in Developing Theories of Mind, ed. J. W. Astington, P.L. Harris, and D. R. Olson (New York: Cambridge University Press, 1988).

come to understand that another child's seeing something or being told something leads to his knowing about that thing. Wimmer et al. account for the lack of this understanding in younger children by suggesting that when the younger children assess whether they themselves know something -what is in a closed box for example - they carry out an answer-check procedure to see if they can come up with the answer. In doing so they do not consider whether they have looked in the box or been told what is in it. Wimmer et al. suggest that the younger children attempt to perform the same procedure when asked about another's knowledge. There too when checking if the other has the answer they ignore whether or not he has seen or been told what is in the box. What they do, suggest Wimmer and his colleagues, is to check whether or not the other child has uttered anything that will indicate that he has the answer. This accounts for why, if the other child has not actually stated the answer, the three-year-olds either make errors or deny that the other knows. Five-year-olds however carry out an *access*-check procedure. They evaluate the other's access to information about what is in the box. Wimmer et al. go on to suggest that whereas children of the age of three have acquired the fragment of the theory of mind that says that people who say that p typically know or believe that p, children of the age of five have acquired an additional fragment that says that people who are told that p or see that p typically come to know that p.

Harris objects that these conclusions of Wimmer et al. are not borne out by the results of the earlier 1983 Maxi experiment. If the three-year-old's assessment of another's knowledge does proceed by the answer-check method then the children in the previous experiment, after attempting and failing to perform the answer check procedure with regard to Maxi, since they are fully aware that Maxi has not declared where he thinks the

chocolate is, should predict that Maxi will go to the box as that is where Maxi put the chocolate before he went out to play. If these children *do* use the answer-check procedure what they should *not* do is to credit Maxi with knowledge he shows no sign of having, that is, knowledge that the chocolate has been moved to the cupboard. In fact, says Harris, subsequent research has shown that three-year-olds do quite well when asked which of two people knows what is in a box when one has looked inside and the other merely lifted the box up. This finding is consistent with what one would expect if it were the case that children proceed by *imagining* themselves looking into the box and *simulate* seeing the contents irrespective of whether they themselves have looked into the box.

Harris holds that Wimmer et al. are mistaken also in thinking that children do not credit others with inferential knowledge. Harris describes the smarties box experiment. Children are shown that a well-recognized smarties box contains pencils rather than smarties (M & Ms). The lid is put back on the box and the children are asked to predict what other children will say is in the box before opening it. It turns out that whereas four-year-olds predict smarties, three-year-olds predict pencils. If it were the case that the four-year-olds proceeded by using the access check method, they should say that the others will have no idea what is in the box since they have had no access to its contents. The fact that they make a correct prediction seems to indicate that they credit the others with making the same inference from the familiar outside of the box to what is likely to be inside as they would themselves. Harris suggests that the way this works is that they set aside what they themselves know about the contents of the smarties box and feed into their decision-making mechanisms the fact that the box is a smarties box.

The above is a brief summary of some of the evidence marshalled by the simulation theorists in support of their view that simulation provides a better account of the mechanism that underlies our folk psychological practices than the theory theory does. Both sides of the debate, however, acknowledge the inconclusiveness of the experiments conducted to date.

3:6 The Nature of Psychological Explanation

The interest that the simulation theory has generated in the philosophical world since it was introduced by Gordon and Goldman seems to indicate that this theory is generally considered to be of greater depth and broader scope than the modesty of its initial thesis might lead one to suppose. For not only has it been taken very seriously by philosophically-minded cognitive scientists, notably Stephen Stich, who find the widely accepted theory theory of the mechanism that underlies our different abilities seriously challenged,[43] but the simulation theory, considered as a view about the nature of folk psychology, is also strongly criticized by eliminativists such as Paul Churchland, as we saw above.[44] It has been suggested that the reason for this is that they find the premise that folk psychology is a theory necessary to their argument that it is a false theory.

However the simulation theory is also discussed in a positive way in relation to its potential to solve certain purely philosophical problems. Martin Davies takes up Gordon's suggestion that the simulation theory has

---

[43] Stephen Stich and Shaun Nichols, "Folk Psychology: Simulation or Tacit Theory," Mind and Language 7 (Spring/Summer 1992): 87-97. See also Shaun Nichols, Stephen Stich, Alan Leslie, and David Klein, "Varieties of Off-line Simulation," to appear in Theories of Theories of Mind, ed. P. Carruthers and P. Smith (Cambridge: Cambridge University Press).

[44] Paul Churchland, A Neurocomputational Perspective: The Nature of Mind and the Structure of Science (Cambridge: Cambridge University Press, 1989), 118.

the potential to give "a way of interpreting ordinary discourse about beliefs in terms of pretend play and practical simulation." He understands this as the suggestion that "the simulation view will yield an alternative to the theory theorist's understanding of mental states themselves."[45] Davies also thinks that there might be a way to get a simulation-based philosophically fundamental account of the conditions on the possession of mental concepts. Simon Blackburn is another theorist who discusses the simulation theory with regard to its providing an alternative to the theory theory account of understanding "the sayings and doings of others."[46] These suggestions, and others, for philosophical uses of simulation indicate that Gordon is not alone in thinking that the simulation theory might have the potential to be much more than an empirical hypothesis about what guides us in the carrying-out of our folk psychological practices.

It is in light of this possibility that we are led to consider the resources of the simulation theory to give a philosophically fundamental account of the nature of commonsense psychological explanation. It is clear however that replacing the strand of the theory theory of folk psychology that is to do with the nature of psychological explanation will only be motivated if it can be shown by simulationists that psychological explanation is radically different from the kinds of non-basic explanation that they appear to agree are covering law explanations. If there were nothing special about the nature of folk psychological explanation, there would be no reason to say that it is not subsumed by the covering law

---

[45] Martin Davies, "The Mental Simulation Debate," to appear in Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind, Proceedings of the British Academy 83, ed. C. Peacocke, 7. (Page refs. here and in what follows are to an unpublished manuscript.).

[46] Simon Blackburn, "Theory, Observation and Drama," Mind and Language 7 (Spring/Summer 1992): 187-203.

model as the simulationists under discussion here appear to assume that other non-basic explanations are.

Indeed we have seen that the widely-accepted view of the nature of non-basic explanation is that it consists of a deductive argument, one premise of which is a cp law, or a conjunction of such laws, and the others are particular fact premises. On this model our ordinary commonsense explanations are incomplete since only the particular fact premises are explicit, the covering law premise being tacitly deployed. A feature that makes the theory theory of folk psychology particularly attractive is that it fits this model of non-basic explanation. It is clear that the simulation theory of the mechanism underlying our explanations of behavior is compatible with the theory theory of what constitutes a commonsense psychological explanation, i.e. that it is a covering law model explanation. If one held a hybrid view of folk psychology in which the simulation heuristic accounted for the procedure we follow in order to choose which commonsense psychological explanation is appropriate to the occasion and the theory theory accounted for the nature of that explanation, simulation would just be the heuristic we use to get to know which laws of the tacitly-known theory we should deploy on a particular occasion.

Since simulationists appear to accept the covering law model for other types of non-basic explanation, the suggestion that a simulation-based model of the nature of commonsense psychological explanation replaces the theory theory view will only be motivated if it is the suggestion that *psychological* explanation does not fit the covering law model, that it is a different type of explanation altogether.

What seems to indicate that simulationists *do* think that psychological explanation is different from other kinds of non-basic explanation is the

intimation throughout their account that in the case of psychology a theory is not needed *at all.* Prima facie support for this idea is derived from the intuitively plausible view that since the belief-desire system is part of the very make-up of the human organism and since human beings are cognitively isomorphic, all that is needed when we want to explain why a particular piece of behavior was performed on a certain occasion, or to predict such a performance, is to look and see what our own mechanism would do in similar circumstances.

There are arguments in the literature that support the vague, but plausible, suggestion that we are cognitively similar in this way. Roy Sorenson has an evolutionary approach. He suggests that natural selection has molded us into a population of hyper-similar individuals and this helps support the reliability of our simulations of each other.[47] Goldman suggests that empirical evidence for psychological similarity could be gleaned from seeing whether simulation-based predictions of behavior are accurate. "If it can be settled that many predictions of intentional behavior do in fact use the simulation heuristic, these successful predictions will also have been correct in inferring the agents' (mental) decisions or choices. This already provides support for psychological similarity."[48] Goldman goes on to refer to John Harsanyi's a priori similarity postulate: "This is the assumption that, once proper allowances have been made for the empirically given differences in taste, education, etc., between me and another person, then it is reasonable for me to assume that our basic psychological reactions to any given alternative will be otherwise much the same." Harsanyi insists that

---

[47] Roy Sorenson, "Self-Strengthening Empathy: How Evolution Funnels us into a Solution to the Other Minds Problem," 10. Unpublished manuscript.

[48] Alvin Goldman, "Simulation and Interpersonal Utility," 25. Unpublished manuscript.

such non-empirical postulates are common in science and specifically compares it with a priori non-empirical criteria of theory choice, such as simplicity, parsimony, and preference for the least 'arbitrary' hypothesis.[49] Arguments such as these for the psychological similarity of our intentional-state forming operations have strong intuitive appeal.

The position that our explanations of human behavior are not theory-dependent is also supported by the non-positivist *verstehen* view discussed at the beginning of this chapter: investigation into fields to do with human psychology, and, in particular, the explanation of past human behavior, or of historical events, in terms of the psychological events responsible, is different from other types of investigation and explanation. One does not need to appeal to laws in the case of psychology, the argument went, since in order to find the thoughts that are expressed by the event to be explained, one just needs to use one's own mind to recreate the mental atmosphere of the object of one's study.[50]

But these arguments for the plausibility of commonsense psychological explanation being different are, as we said, merely prima facie. In order to examine whether or not such a view is well-grounded we shall consider a rough, but seemingly plausible, suggestion for a simulation-based account of what it is that constitutes a correct commonsense psychological explanation of behavior. The suggestion to be discussed is the following:[51]

---

[49] Ibid., 16.

[50] Von Wright, Explanation and Understanding, 6.

[51] This formulation comes partly from notes taken during a seminar given by Stephen Schiffer in Fall 1992 at the Graduate Center of the City University of New York and partly from a discussion with Hartry Field in summer 1994.

My statement that Odile went to Lyons because she believed p and desired q is a correct explanation of her going to Lyons iff

(1) she went to Lyons

(2) she believed p and desired q

(3) a decision to go to Lyons is the result of a simulation of Odile run on my decision-making mechanism and that simulation satisfies two conditions:

(a) it essentially involves the belief p and desire q

(b) there are no facts about Odile such that were I to know them a different decision would result from my simulation of her.

This formulation assumes in (3) that in simulating Odile I use a decision-making mechanism that essentially involves my psychological states. It also implies that Odile has a decision-making mechanism that essentially involves her beliefs and desires, a decision-making mechanism that is just like mine and whose processing I am simulating in giving my explanation of her behavior. It is clear from our preceding description of simulation that this assumption and its implication correspond to the simulationist view about the isomorphism between our decision-making mechanisms. However the assumption that Odile has a decision-making mechanism that essentially involves her beliefs and desires seems to present two problems for this formulation.

The first difficulty is that if it is the case that Odile is endowed with such a mechanism, it is not clear what the goings-on in *my* head will have to do with what makes an explanation of *her* behavior correct. If Odile has such a decision-making mechanism then, although it seems true that *part* of what will make my statement that she went to Lyons because she believed p

and desired q correct is that the decision to go to Lyons was the result of a decision-making process in Odile that essentially involved the belief p and the desire q, and the fact that she did go to Lyons, it does not appear likely that whatever *else* is needed in order to make it correct will be anything to do with *my* decision-making mechanism. And if whatever else is needed to make it correct has nothing to do with my decision-making mechanism then the relevance of my simulation of her mindset to the correctness of my explanation of her behavior is not clear. The question that needs to be answered is: What need is there for the simulation-based condition at (3) to be satisfied in order for my explanation to be correct?

Let us imagine that "O did x because she believed that p" is a correct explanation of the fact that O did x. It would seem that I could give this explanation of the fact that O did x, and, at the same time, it could be the case that I know that in O's shoes, with all her relevant beliefs and desires, and with the same decision to do x resulting from the decision-making process taking place in me, I would not do x because I believed that p. Thus it seems that not only does my correct explanation not result from a simulation of O, but that if I had relied on my simulation of O, the correct explanation of the fact that she did x would have eluded me. If we are right about this, and a correct psychological explanation can be given without the simulation-based condition mentioned above being met, then simulation will not be involved in a necessary condition on psychological explanation.

Simulationists might be tempted to argue that in spite of the fact that the simulation condition does not appear to be satisfied by this correct explanation, this explanation does in fact satisfy such a condition. They might suggest that my decision-making system, *malgré moi*, has taken into account all relevant facts about O, and has processed them off-line in a

manner such that *her* motivating beliefs and desires float to the surface of *my* consciousness leaving my own beliefs and desires intact and inert. On this view an incorrect explanation would be the result of an incorrect simulation, one in which my own beliefs and desires have become jumbled up with the pretend O-like ones and have eventually outweighed them.

This hypothetical suggestion of simulationists however goes against the grain of the only non-theory-involving model of simulation discussed here, Gordon's imaginative identification model. In processing the relevant facts about Odile in a manner that will leave my own beliefs intact, the tacit pretend inputs to my system are likely to be hypotheses of the form "I believe that p," and not simply "that p." Martin Davies points out that if one takes the inputs to be of this form, then the processing in the simulator could follow the "contours of the derivational structure of a proof of a conclusion about, say, an intention to act from premises about, say, the agent's beliefs and desires - a proof cast in a psychological theory." Indeed, he adds, some accounts of what it is for a cognitive processing system to embody tacit knowledge of a particular theory is that "... a component processing mechanism embodies tacit knowledge of a particular rule or axiom if it plays a role in mediating causally between representational states that is structurally analogous to the role that the rule or axiom itself plays in mediating derivationally between premises and conclusions."[52] Thus simulationists who take this line of argument are

---

[52] Martin Davies, "The Mental Simulation Debate," 17.

saddled with a model of simulation which appears to rely on the assumption of a tacitly-known theory.

It seems then that satisfaction of the simulation-based condition (3) on the correctness of commonsense psychological explanations in the formulation above is not necessary for correct explanation. It seems likely that many commonsense psychological explanations of behavior that are in fact correct will not satisfy it. .

A closer look will reveal that simulation is unlikely to be involved in a sufficient condition on explanation either. It seems that a simulation of Odile run on my decision-making mechanism could essentially involve the belief p and the desire q and that there could be no facts about O such that were I to know them a different decision to behave would result from my simulation of her and it could still be the case that the beliefs and desires in terms of which I explain her behavior are not those which are in fact explanatorily relevant to it. The simulation-based condition for correct explanation of behavior appears to be too weak for it can be satisfied by an explanation that is false.

We saw above that Gordon attempts to counter this problem by introducing his principle of least pretending, the deployment of which is supposed to select the belief-desire set which is explanatorily relevant to the behavior to be explained. However it seems likely that the deployment of such a principle would not be sufficient to select one set over others, that there would still be several sets that would qualify as being explanatorily relevant to the behavior in question. It would seem then that a simulation-based condition on psychological explanation is likely to qualify as explanations many "explanations" of behavioral facts that are in fact incorrect.

One option in this situation is for simulationists to add a further condition to the conditions already stated in the above formulation, one that would rule out the incorrect explanations. The problem here is that since this further condition is likely to be one that is not simulation-based and one whose satisfaction is nevertheless essential to the correctness of the explanation, there would be little or no work left for the simulation-based condition to do in qualifying psychological explanations. The option of settling the problem by appealing to a different, stronger simulation-based condition seems to be ruled out by the fact that such a condition is likely to exclude a greater number of correct explanations than the present condition already does.

So it appears to be the case that satisfaction of the simulation-based condition at (3) above will be neither sufficient nor necessary for my explanation of why Odile went to Lyons to be correct. It seems that if one supposes that a person has a decision-making mechanism that essentially involves her beliefs and desires and if one supposes that this fact is related to her behavior, then what will be relevant to the correctness of a psychological explanation of that person's behavior is likely to be something about her, and not related to another person's simulation of her.

The second difficulty that Odile's having a decision-making mechanism that essentially involves her psychological states poses for a simulationist account of the nature of psychological explanation is to do with the motivation for such an account. We mentioned earlier that in order to motivate a simulation account of the nature of commonsense psychological explanation, simulationists must hold that psychological explanation is a different kind of explanation from the kind that they hold is successfully accounted for by the covering law model. As covering law

explanations are causal explanations, if one sees commonsense psychological explanation as different, one might see it as a species of non-causal explanation. Let us look at this possibility.

It is clear from our discussion of the simulation theory that there are two schools of simulationists. It is also clear that the accounts of both schools rely on an isomorphism between the causal decision-making mechanisms of human beings. While Goldman's off-line simulation view is often criticized on the grounds that positing such a mechanism presupposes tacit knowledge of the laws of an internally represented theory, Gordon's imaginative identification model is not mentioned in this regard. One might be tempted to draw the conclusion that commonsense psychological explanations will be covering law explanations only if the off-line view of the simulation mechanism describes what actually occurs. The reasoning underlying such a conclusion might be that since on the imaginative identification view simulations occur by simple projection of the self, and since no off-line processing takes place, and no inference to the other is made, and no laws are invoked, that the commonsense psychological explanations that result from such simulations are not covering law explanations. A theorist who holds a view that appears to support this conclusion is Martin Davies.

Davies suggests that a close look at the procedure assumed to be taking place by theorists who object to the simulation theory on the grounds that it collapses into the theory theory reveals that what they take to be going on in the first stage of our simulation, albeit tacitly, is that I, the simulator, pretending to be in the shoes of the person I am simulating, entertain in my imagination hypotheses concerning mental states.[53] In the

[53] Martin Davies, "The Mental Simulation Debate," 18.

second stage I proceed to a conclusion about a further mental state, or about an action or an intention to act. Davies suggests that if the first-stage hypotheses are taken to be of the form "I believe that p" and "I desire that q," then, since they exhibit the general form "x believes that p," the states of entertaining them will be appropriate inputs to a mechanism that embodies tacit knowledge of the principles of a psychological theory. However, Davies points out, the simulation procedure can be seen as theory-independent and the simulating system as non-law-embodying if one understands it in terms of Gordon's imaginative identification model as opposed to the off-line model.

Under the imaginative identification construal of simulation the processing mechanisms that mediate transitions among states will no longer appear to be embodiments of tacit knowledge of the principles of a psychological theory. The suggestion is that the simulator, instead of imaginatively entertaining hypotheses about mental states, imaginatively adopts the mental states themselves, that is, he imagines believing that p and desiring that q. Thus the contents of the states processed by the decision-making mechanism while simulating another are simply "that p" or "that q." The assumption is that a decision-making mechanism that mediates transitions between states with such contents will not be one that embodies tacit knowledge of the principles of a psychological theory.

We agree with Davies that the first construal of the simulation procedure presupposes tacit knowledge of a theory. If it were the case that we had such knowledge then it is clear that our psychological explanations of behavior would indeed be covering law explanations and simulation would be the way we pick out the law appropriate to the occasion. Our point, however, concerns the second construal of the simulation procedure.

It seems to us that the decision-making mechanism that is presupposed by the second construal, even though it might not embody tacit knowledge of the laws of an internally-represented theory, is such that an explanation that is generated by it would be causal.

Let us imagine that, just as Davies and Gordon suggest, the part of the cognitive processing mechanism that is responsible for decision-making mediates transitions between states with contents such as "that p" and that this part of the cognitive mechanism does not embody tacit knowledge of the laws of a psychological theory. Then it seems that one might hold that (part of) what makes my explanation of Odile's behavior correct is that the states with the contents "that p" and "that q," are causally responsible by virtue of those contents for the emergence of the decision to go to Lyons from the processing taking place in her. That is, one might hold that this particular decision to behave would not have emerged in her had not states with the content "that p" and "that q" been instantiated by her. It appears then that simulationists who espouse the Gordon-Davies view of the simulation procedure are still likely to maintain that such explanations are *causal* explanations.

But if simulationists do opt for the position that commonsense psychological explanations are causal explanations, then it is incumbent upon them to tell us why these commonsense causal explanations are not subsumed by the covering law model as they appear to maintain that other commonsense causal explanations and other types of non-basic explanations are. If one finds the commonsense explanation "the window broke because it was hit by a rock" an acceptable commonsense causal explanation of the fact that the window broke, and if one holds that this is so because there are (probably unknown) background laws which subsume these events, and if

one accepts that "Odile went to Lyons because she believed p and desired q" is a satisfactory commonsense causal explanation also, then there seems to be no reason to deny that there are underlying laws in this case too. In each case the laws would not have to be internally represented and tacitly known. They could be unknown laws of an unknown type which subsume the events to be explained, laws which are just assumed to exist.

In the psychological case the background laws might be of a type that, were they knowable, could not be expressed in intentional terms. The events described in intentional terms in the particular fact premises of the explanans and in the explanandum of the commonsense explanation might have another description by virtue of which they are subsumed by an unknown law which would play the role of the covering law premise in the explanans of a D-N explanation. On the other hand these background laws might well be laws statable in intentional terms, but not be *commonsense* intentional laws. These laws might be exotic intentional laws which no one has yet discovered and might never discover.[54] In this case the reason our predictions and explanations of our own and others' behavior are so successful is that we are all subsumed by the same unknown exotic intentional laws. We exploit ourselves as functioning exemplars of the laws in order to explain the behavior of others. The correctness of such explanations will depend not on the correctness of our mental simulations but on whether or not we exemplify at the time of the explanation the same law that the person whose behavior we are explaining exemplified at the time of the performance of that behavior.

---

[54] Stephen Stich brought this possibility to my attention. A similar idea is expressed by D.Kahneman and A. Tversky in "The Simulation Heuristic," Judgment Under Uncertainty, ed. D. Kahneman, P. Slovic and A.Tversky (Cambridge University Press, 1982).

It would seem that if we understand the formulation of the simulation-based conditions on correct explanation above as presupposing that the person whose behavior we are explaining has a decision-making mechanism that essentially involves her beliefs and desires, then unless we can account for why it is that commonsense psychological explanations are not satisfactorily accounted for in the above manner, psychological explanations will be subsumed by the covering law model and the motivation for finding an alternative to the theory theory view of commonsense psychological explanation, viz. that it is a radically different kind of explanation from other types of non-basic explanation, will disappear.[55]

The only way out for simulationist seems to lie in the possibility of there being a different understanding of the above formulation. Let us suppose that simulationists are motivated to find an alternative account of psychological explanation because they believe that the ascriptions of beliefs and desires which we make to others in the course of explaining their behavior, unlike those we make to ourselves, are non-factual. In this case they might hold that although there is no determinate fact of the matter about the correctness of my ascriptions to others there is still a sense in which my explanations of their behavior in terms of those ascriptions are correct.

These theorists might hold that although the lack of facticity of our third person ascriptions deprives the beliefs and desires ascribed of their causal relevance, they are still *explanatorily* relevant to the behavior, that

---

[55] Note that although commonsense psychological explanations would be covering law explanations were such laws to exist and play the role described here, this fact would only help theory theorists on the interpretation of the theory theory that does not hold that the laws are tacitly-known in some sense by the folk. See Chapter 1, note 1.

our explanations of other people's behavior in terms of beliefs and desires are bona fide explanations, that many of them are correct in some sense. Exactly what this relevance consists in would need to be spelled out and the sense in which such explanations could be correct accounted for, but, providing this could be done in a satisfactory manner, the explanations of other people's behavior that refer to beliefs and desires would be a species of non-causal explanation. Here, then, is a rationale for saying that commonsense psychological explanations of behavior are radically different from other types of commonsense explanation which are widely acknowledged to be causal covering-law-type explanations.

When we look a little more closely at this suggestion, however, there appear to be consequences that a simulationist might be reluctant to embrace. If the motivation for this realist, though deflationary, non-causal account of explanation is indeed that there is a difference in the facticity of first and third person ascriptions of propositional attitudes, then it seems that some kind of asymmetry is assumed to exist. One possible way to understand this asymmetry is as a "deep" asymmetry, that is, as an asymmetry that applies to the decision-making mechanisms of Odile and myself. In this case the difference in the facticity of my because-statements stems from the fact that although I have a causal decision-making mechanism that essentially involves beliefs and desires, Odile, and the rest of the population, does not. On this interpretation it remains a mystery how it can be the case that I am the proud possessor of such a mechanism when no one else is similarly endowed, as well as what it could mean to simulate someone else in such circumstances.

On the other hand we might take the asymmetry that the difference in the facticity of our ascriptions implies as an asymmetry that is "shallow,"

that is, it is not one that exists between the decision-making mechanisms of Odile and myself but merely between the explanations I utter. In this case my statements about her beliefs and desires and her behavior are non-factual not because she has no beliefs and desires at all, ever, but because "belief" and "desire" used in the context of third-person ascriptions do not refer to the *causes* of behavior. The ascriptions of beliefs and desires to her are made in order to explain behavior, maybe in order to "rationalize" it, and it is assumed that they can do so without referring to causes. These statements will be correct to the degree that they fulfill that purpose. So the asymmetry here would consist in the fact that my explanations of my own behavior in terms of my beliefs and desires and my explanations of another's behavior in terms of her beliefs and desires are different kinds of explanation. On this view the explanations of my own behavior would be causal covering-law-type explanations, and those that I utter concerning the behavior of other people would be non-causal, but correct in whatever sense of "correct" the view mentioned above authorizes.

The advantages of this "shallow" interpretation of the asymmetry is that it agrees both with our intuition that everyone is endowed with a causal decision-making mechanism and with our intuitions that we have privileged access to our own mental states and that the explanations we give of our own behavior are special in that they appear to be "three-dimensional," or more vivid in some way than those we give of the behavior of others. Whether or not the seemingly high number of our successful explanations of other people's behavior can be accounted for if third-person explanations are of the type mentioned here will depend on the as-yet unexplicated notions of "explanation" and "correct" used in relation to third-person explanations.

However it is important to recall that what is under discussion here is the possibility of a *simulation*-based account of psychological explanation. Even though this last view does posit a symmetry or isomorphism between the decision-making mechanisms of myself and the person whose behavior I am explaining, it is clear that that symmetry does not contribute to the *correctness* of my explanations of that other person's behavior. And if the correctness of these explanations does not depend on the symmetry of our decision-making mechanisms, then it is hard to understand what it means to say that we are simulating in these circumstances. In fact one is tempted to draw the conclusion that the most plausible non-causal account we can get of the explanation of the behavior of others seems not to need the simulation-based condition on psychological explanation at all.

In this section we have examined a suggestion for a way in which simulationists could maintain a replacement thesis with regard to the theory theory of folk psychology. This suggestion concerned how they could replace the strand of the theory theory that has to do with the nature of psychological explanation. What was offered appeared to be a plausible formulation of simulation-based conditions on commonsense psychological explanation. It turned out that an assumption underlying this formulation is that we are all endowed with a similar causal decision-making mechanism. Our discussion showed that if it is true that we are so endowed, then although (part of) what will make our explanation of another's behavior correct is that the decision-making mechanism operated in that person on the occasion in question in the way that we say that it did in our explanation and that the event referred to in the explanandum did occur, whatever else is needed does not seem to have anything to do with the decision-making

mechanism of the explainer. This conclusion seems to indicate that simulation is not likely to enter into an account of the nature of explanation. Our discussion also showed that the simulationist's plausible assumption that everyone is endowed with a decision-making mechanism that essentially involves their beliefs and desires seems to imply the view that the explanations we give of other people's behavior in terms of their beliefs and desires are causal explanations. We concluded that if this is indeed the case, such explanations would be candidates for subsumption by the covering law model, and simulation, if it plays any role at all, will merely be the heuristic by which we pick out the appropriate law. When we found a way to interpret the formulation of simulation-based conditions on psychological explanation so that it did not have the unwanted implication that our explanations of others' behavior are causal it turned out that not only did this account of explanation make no appeal to simulation, but that it was hard to see what simulation would amount to on this view.

We conclude from this discussion that the simulation theory looks distinctly unpromising with regard to having the resources to provide an independent, alternative account of psychological explanation to the theory theory's covering-law-model account. Although it remains open to the simulationist to suggest a simulation-based condition that, unlike the one proposed here, is necessary or at least sufficient for explanation, we suggest that the problems outlined above are likely to be inherent to any account of explanation based on the notion of simulation. The main questions about the nature of correct commonsense psychological explanation do not seem to revolve around issues to do with the manner and degree to which our belief-forming mechanisms are similar and our

mental simulations of each other correct. What these questions do concern will be the topic of discussion in our next, and final, chapter.

Chapter Four

A Different Perspective on Psychological Explanation

In this chapter we say what we consider ourselves to have shown so far with regard to the nature of commonsense psychological explanation. In doing so we bring to the fore the constraints that these conclusions impose on a new account of explanation and then look at how some recent suggestions for an account respect those constraints.

We have seen that a great advantage of the theory theory of folk psychology, and perhaps its driving motivation, is that it allows commonsense psychological explanation to fit the covering law model of causal explanation. On the theory theory view psychological explanation is just one more instance of the type of explanation that already accommodates other kinds of non-basic explanations and the scientific explanations whose hidden structure this model was originally assumed to display. When we examined likely candidates to play the role of commonsense psychological covering law in the deductive arguments of which these explanations are thought to be composed, we found that none would do. We concluded from this that it is probable that our ordinary explanations of behavior in terms of the beliefs and desires we judge responsible for it are not ones in which commonsense covering laws are invoked.

The immediate consequence of the conclusion about the unlikelihood of there being commonsense psychological covering laws for our ongoing investigation into the nature of commonsense psychological explanation is that it raises the issue of the kind of explanations commonsense

psychological explanations can be if they are not of the covering law type. A particular problem with regard to this question is whether or not, if there are no commonsense psychological covering laws, it is still possible for propositional attitude 'because'-statements to be literally true, or, in other words, for the intentional properties cited in such statements to play a genuine causal-explanatory role in commonsense explanations. The question arises because of the widespread philosophical assumption, one that dates from Hume, that causes are to be accounted for in terms of regularities or laws. This assumption was recently formulated by Davidson as "events related as cause and effect fall under strict deterministic laws."[1] It would seem to follow that if there are no propositional attitude laws, strict or otherwise, then propositional attitude 'because'-statements cannot be literally true.

The upshot of this is that if one maintains, in spite of the apparent lack of commonsense psychological causal laws, that commonsense psychological explanations are bona fide causal explanations, then one must argue that law subsumption is not a necessary condition for causal-explanatory relevance, that there are other conditions for causal relevance, conditions which are in fact satisfied at least by what correct commonsense psychological explanations cite as causally relevant. Alternatively, one might hold that commonsense psychological explanations are a species of non-causal explanation. We explored the simulation theory for its potential to provide an account of psychological explanation of this latter type.

Our discussion of the simulation theory however showed that there appears to be little hope of getting a non-causal account of psychological

---

[1] Donald Davidson. "Mental Events," in Essays on Actions and Events (Oxford: Oxford University Press, 1980), 208.

explanation from that source. Both Gordon's and Goldman's interpretation of the simulation procedure are such that explanations that devolve from it would be both causal and candidates for the covering law model. When we considered a plausible suggestion for a non-causal deflationary interpretation of Gordon's and Goldman's views, it turned out that such an account presupposes some kind of asymmetry between the explainer and those whose behavior he is explaining and hence it is not clear in what sense, or if, simulation would be involved in such an explanation.

The immediate consequences of our conclusion about the unlikelihood of simulation's being involved in an account of the nature of psychological explanation is that one possibility for psychological explanation being a type of non-causal explanation can be set aside. Our options then are either to look for another way to get a non-causal account or else to look for an account that gives a causal-explanatory role to psychological properties but that does not presuppose the existence of commonsense psychological covering laws.

It is the latter route that we shall pursue here, the philosophical motivation for espousing epiphenomenalism failing to override the strength of our intuition that beliefs and desires actually do cause behavior, that the role played by mental properties in our explanations is a genuine explanatory role. Let us look for a moment at some of the arguments that sway some philosophers in the direction of the counterintuitive conclusion that mental properties are causally inert. The first is usually introduced by way of an example that illustrates the general point that some properties of an event are irrelevant to some of the causal transactions into which that event enters, Dretske's "soprano" example:

Something possessing content, or having meaning, can be a cause without its possessing that content or having that meaning being at all relevant to its causal powers. A soprano's upper register supplications may shatter glass, but their meaning is irrelevant to their having this effect. Their effect on the glass would be the same if they meant nothing at all or something entirely different.[2]

Dretske explains that although the singing consists of singing a phrase with a certain meaning we do not feel compelled to conclude that the meaning contributed causally to the glass's shattering. What distinguishes the mental properties of the events that cause our behavior from the semantic properties of the singing? he asks.

Another motivation for thinking of mental properties as causally inert is sometimes referred to as the problem of explanatory exclusion. In its general form this problem arises when we assume that "(i) every concrete state, event, or process in the world is one that falls within the purview of physics, and (ii) that every event, state, or process within the purview of physics is susceptible, in principle, to a full and complete explanation in terms of the laws and concepts of physics."[3] Such plausible assumptions raise the question of whether or not any other kind of explanation could ever be genuine or legitimate, whether or not the real explanatory work is done only by physics. With regard to psychological explanations the question is whether or not neurobiological explanations, even if they can co-exist with basic physico-chemical explanations, screen off psychological explanations by taking over their explanatory role.

---

[2] Fred Dretske, Explaining Behavior (Cambridge: MIT Press, 1988), 79.

[3] Terence Horgan, "The Explanatory Role of Content," in Dretske and His Critics, ed. Brian McLaughlin (Oxford: Blackwell, 1991), 80.

The tension for many theorists is between conceptual difficulties such as these with the causal-explanatory role of mental properties and a deep-seated reluctance to embrace epiphenomenalism as an intellectually attractive alternative. Tyler Burge suggests that this reluctance is grounded in the fact that the very interest of psychological explanation consists in that it helps us understand ourselves as agents. To borrow his words:

> Causal implications are built into our intentional concepts and intentional modes of explanation. We think that we make things happen because we make decisions or will to do things. We think that we make assertions, form theories, and create cultures, because we think certain thoughts and have certain goals - and we express and fulfill them. In this context, we identify ourselves primarily in terms of our intentional mental aspects - our wants, our thoughts, our values. Our agency consists in our wants', willings', thoughts', values' as such (under these aspects) having some sort of efficacy in the world. Our mental events' having the intentional characters that they have is, in individual instances, what we define our agency in terms of.[4]

Burge goes on to say that if psychological explanation merely rationalized or made sense of what we did but did not provide any insight into the nature of any causal efficacy, it would lose much of its point since "it would provide no insight into the various forms of agency that give life its meaning and purpose, and psychology its special interest."

The rest of this study will be devoted to exploring the possibility of an account of commonsense psychological explanation that gives a causal-explanatory role to psychological states or events, or to the properties of such states or events, but which does not appeal to any putative commonsense psychological causal covering laws. Before we begin this task

---

[4] Tyler Burge, "Mind-Body Causation and Explanatory Practice," in Mental Causation, ed. John Heil and Alfred Mele (Oxford: Clarendon Press, 1993), 119.

however it is necessary to remind ourselves of the far-reaching implications of the conclusions we have drawn so far about the theory theory and the simulation theory. The purpose in raising the wider issues in relation to these theories in this place is twofold. First, in doing so we intend to acknowledge that in putting forward these views pertaining to the theory theory we are raising questions about well established philosophical positions both in relation to the nature of folk psychology and to the nature of non-basic explanation in general. And secondly it serves to remind us what exactly is at stake in rejecting these theories and to highlight what is required of a new account.

It is clear that if the conclusion that psychological explanation is not a covering law explanation is correct then this conclusion will have serious implications for the theory theory of folk psychology. If indeed it is the assumption that commonsense psychological explanations fit the covering law model that is the driving motivation for theory theorists to posit tacit knowledge of a theory of commonsense psychology in the folk, then if these explanations do not fit that model, theory theorists are left without a motivation for saying that folk psychology consists of an internally-represented tacitly-known set of laws which constitute a theory. This position is thus challenging the view of the nature of folk psychology that is the most widely accepted today.

The correctness of our conclusion about psychological explanation and the covering law model would also challenge the widely-held view about the nature of other types of non-basic explanation. Proponents of the theory theory see folk psychology as a special science and explanations in the special sciences as subsumed by the covering law model. The theorist who argues that commonsense psychological explanations cannot follow the

covering law model because it is likely that there are no psychological cp laws, has to face the fact that the same arguments might apply to the existence of cp laws in the special sciences. The view that there are no special science laws is bound to be unpopular since special sciences are, after all, sciences and the general view of scientific theories is that they are collections of laws. Also, in arguing that there are no special science cp laws one is arguing that special science explanations across the board cannot be of the covering law type.

When we turn to the conclusions we drew above about the simulation theory, we see that they too have some philosophical implications, although these seem to be more limited in scope than those derived from our conclusion about the theory theory. Let us suppose that it turns out that simulation is not involved in an account of the nature of explanation, but that simulation is the fundamental heuristic we actually use in giving our commonsense explanations, and not law-deployment. In this case a satisfactory philosophical theory of the nature of mind, including the strand that deals with psychological explanation, will have to be compatible with the simulation heuristic. It seems likely that different constitutive philosophical views about the nature of mind, including the theory theory view, will be compatible with the simulation view of the procedure we follow when we perform our folk psychological practices of explanation and prediction and that the simulation theory of folk psychological practices is unlikely to offer any decisive conceptual arguments in favor of any one of them. This speculation does not lead one to predict an interesting philosophical career for the simulation theory. It would seem that it will be mainly of interest to cognitive psychologists.

4:1 The Quausal Relevance of Properties

Let us now turn to the discussion we proposed above. Our aim is to find an account of psychological explanation that gives a causal-explanatory role to psychological properties but which does not presuppose that those properties are qualified to play that role because they are subsumed by causal laws or that such explanations are of the covering law model. We shall first focus on the question of what it is to be a property that is causally relevant to a particular causal transaction between two events.

Terence Horgan, in a paper whose main purpose is to answer the question "(D)o the mental types (properties) tokened by mental events and states have the kind of relevance to individual causal transactions which allows these properties to figure in genuine causal explanations?"[5] first sets out to elaborate the general view of the explanatory relevance of properties which he has already sketched elsewhere.[6] In his words his aim is "to account for what it is for two events c and e, together with two properties F and G, to jointly instantiate the four-place relation expressed by the locution 'c qua F causes e qua G'," a relation he calls "the relation of *quausation*.."

Horgan's position is that subsumption by strict laws is too stringent a requirement for quausal relevance, and that subsumption by heteronomic or cp laws, is simply not sufficient. To substantiate the first claim, Horgan gives us examples of 'garden-variety causal explanations': "The bridge collapsed because a truck drove onto it which exceeded its weight-bearing capacity," "The fire was caused by a short-circuit." It is hard to deny that

---

[5] Terence Horgan, "Mental Quausation," Philosophical Perspectives 3, Philosophy of Mind and Action Theory (1989): 47-76

[6] Horgan, "The Explanatory Role of Content," 91.

the properties of the cause events referred to in these explanations are explanatorily relevant and equally hard to argue convincingly that such properties will appear in strict laws, he says.

Horgan provides a vivid illustration of how appearing in a strict law, besides being not necessary, is not sufficient to qualify a property as quausally relevant by citing Wesley Salmon's birth-control pill example. "Suppose that Mr. Jones, a man who has been taking birth control pills regularly, fails to get pregnant. Even though the generalization 'All men who take birth control pills regularly fail to get pregnant' satisfies standard criteria for being lawlike, this generalization will not sustain a genuine explanation of why Mr. Jones failed to get pregnant. In his case, the property *taking birth control pills regularly* is explanatorily irrelevant. ..... On the other hand, suppose that Ms Jones, a female who has been taking birth control pills regularly, fails to get pregnant. Now the very same property becomes explanatorily relevant; and a very similar lawlike generalization (pertaining to women) underpins a genuine explanation (albeit a crude one, too crude to count as scientific)."[7]

When he turns his attention to cp laws, he remarks that there are three reasons to think that subsumption by cp laws will not qualify a property for quausal relevance. To illustrate the first one he gives the example of a case where two events, although subsumable by a robust counterfactual-supporting generalization, are not causally related at all. Suppose, he says, that a certain disease causes progressive liver failure in humans and also causes a desire to eat worms. A respectable heteronomic generalization will be 'Cp, anyone who eats worms will die of liver

---

[7] Ibid., 90.

failure.' Yet, he says, eating worms does not cause liver failure, the disease does.[8] So heteronomic law subsumption is not sufficient.

His second argument is against the necessity of heteronomic law subsumption for the causal relevance of properties. It concerns cases of causal transactions in which the cause and effect have properties which are evidently not connected in a lawlike way, but which seem explanatorily relevant anyway. He cites Barry Loewer's example: "Suppose for instance that Barry's noticing a flower shop causes him to remember that tomorrow is his wife's birthday. The properties *being a noticing that there is a flower shop yonder* and *being a remembering that tomorrow is one's wife's birthday*, certainly appear explanatorily relevant to the causal transaction : yet generalizations like

Ceteris paribus, a (married) man who notices that there is a flower shop yonder will remember that the following day is his wife's birthday

seem just false." Subsumption by heteronomic laws thus does not appear to be a necessary condition on quausally relevant properties.

Horgan's third reason for finding that subsumption of properties by heteronomic laws does not qualify them for causal relevance appeals to our pretheoretic intuitions. Intuitively, he says, "even when the proposed criterion yields the intuitively right verdict, one doesn't feel intuitively that the rightness rests on satisfaction of *this* putative criterion. ... [I]ntuitively, the existence of (cp generalizations linking reason-types to the act-types they rationalize) just doesn't seem good enough to guarantee that actions are caused by reasons *qua* reasons." So, again, the subsumption of such

---

[8] Horgan, "Mental Quausation," 55.

properties by such cp laws seems insufficient to qualify them as quausally relevant.

Horgan has a suggestion for a condition that a property satisfies when it is quausally relevant. He says; "It seems evident that part of what is required, in order for properties F and G of the respective cause-event c and effect-event e to be explanatorily relevant to the causal transaction between c and e, is that F's being instantiated is a counterfactually necessary condition for the occurrence of an event that falls under the event-type G."[9] So the effect phenomenon described as instantiating one kind of property, is shown to depend in a certain way on the cause phenomenon, described as instantiating another kind of property. This dependence, he says, largely involves the way the aforementioned properties fit into a suitably rich pattern of counterfactual relations among properties.

In "Mental Quausation" Horgan makes it clear that he is talking about a different kind of counterfactual relation among properties than those that are sometimes thought to obtain, counterfactuals of the type:

(N) If there had not occurred an event with property F, then there would not have occurred an event with property G.

It is often assumed that a counterfactual of this type can be used to weed out the causally relevant properties in a given situation from those that are not. Horgan demonstrates how this counterfactual actually fails to accomplish this task by reference to Dretske's example of the apparently

---

[9] Ibid., 56.

causally superfluous property of a sound's meaning, say, "shatter" to a glass's being shattered by that sound.

(N1) If there had not occurred sounds which meant "shatter," then the glass would not have shattered.

The general assumption is that if this counterfactual fails to be nonvacuously true, then the property of having the meaning 'shatter' fails to satisfy a necessary condition for quausal relevance. However, Horgan points out, it is doubtful that the counterfactual expressed at (N1) does fail to be true. He explains how, given the right sort of background story, this counterfactual will, in fact, be non-vacuously true. For example it would be true if we suppose that the glass-shattering sounds were sung by a soprano who was either strongly determined to sing "shatter" high and loud or to sing nothing at all. He concludes that although satisfaction of the counterfactual at (N) is a necessary condition for the quausal relevance of properties F and G, it is not a sufficient condition.

Although one might think that the solution is to make the counterfactual at (N) stronger, viz.:

(N') If c had not instantiated property F, then there would not have occurred an event with property G,

where c is the event which was the cause in the actual world, one would be mistaken, says Horgan. For, he continues, in Dretske's example, when we take the group of properties that clearly are quausally relevant to the glass's breaking, those of being of a certain pitch, amplitude, etc. and

substitute them for property F in (N'), it is still not clear that it will be *non-vacuously* true that if the sound had not instantiated these properties the glass would not have shattered. One might plausibly argue, he says, that in a possible world where the sound lacks *these* properties, the sound occurring in that world will not be identical with the sound that actually occurred. And so the counterfactual will be true, although vacuously so.

This problem can be fixed, he says, if the pattern of counterfactual dependence required for quausal relevance is wider. He suggests that the pattern of dependence include what would have happened under various pertinently similar circumstances had there occurred or not occurred an event which both instantiated property F and was pertinently similar to the actual cause event c. He asks us to imagine that associated with the causal transaction between c and e in the actual world there is a set of pertinently similar worlds, each of which contains a situation pertinently similar to the situation in which c caused e in the actual world, and exactly one event c* which is pertinently similar to c, and which might or might not contain an event e* which is pertinently similar to e. For an event in one of these worlds to be pertinently similar to c or e, it must be similar not only in pertinent intrinsic respects but it must occur in pertinently similar circumstances to those in which the actual world event occurred. The transworld pertinent similarity relation holds between two events only if the event in the possible world is a good prima facie candidate for being identical with the event in the actual world and if it is a better such candidate than any other event in that world. This relation, however, will be weaker than the transworld identity relation, says Horgan.

He suggests that the quausal relevance of certain properties, say F and G, of a cause event c and an effect event e depends upon what happens

across the range of pertinently similar worlds associated with the causal transaction between c and e.

If
(i) event c causes event e,
(ii) c and e respectively instantiate properties F and G,
(iii) F and G are logically and metaphysically independent and
(iv) the causal transaction between c and e does not involve pre-emption, overdetermination, or the like,
then the fact that c and e instantiate F and G, respectively, is explanatorily relevant to the fact that c causes e iff for any world w in the set of pertinently similar worlds associated with that causal transaction, if c* is the event in w that is pertinently similar to c in the actual world, then
(i) if c* instantiates F in w, then c* causes in w an event e* which both instantiates G in w and is pertinently similar to the actual world event e; and
(ii) if c* does not instantiate F in w, then c* does not cause in w an event which is pertinently similar to the actual world event e.[10]

Horgan calls his view the "the counterfactual pattern conception of explanatory relevance."[11] He adds that this account of quausal relevance requires saying something systematic about what determines the class of pertinently similar worlds relative to a given causal transaction and about what standards govern the transworld pertinent similarity relation. However, he says, we should allow our pretheoretic intuitions to guide us

---

[10] Ibid., 59.

[11] Horgan, "The Explanatory Role of Content," 91.

here since any theoretic account would have to accommodate such intuitions in any case.

To return to Dretske's soprano case; we can now see why the meaning of the sounds is irrelevant to the sounds' causing the glass to shatter. Among the pertinently similar worlds associated with the causal transaction between the sounds and the shattering are worlds in which pertinently similar sounds occur but which either mean something different or mean nothing at all and the glass's shattering or not shattering is not correlated, across the pertinently similar worlds, with the presence or absence of sounds which mean 'shatter.'

Horgan turns his attention next to the quausal relevance of mental properties. Mental properties do qualify as quausally relevant on his account even though things are somewhat different with regard to these properties, he says. Typically, in cases of mental causation, the causal transaction between a mental event c and another event e obtains when a principle of Supervenient Relevance holds for some salient mental property F of c and some salient property G of e. Horgan states this principle as follows:

> For any world w in the set of pertinently similar worlds associated with the causal transaction between mental event c, and e in the actual world, if mental event c* is the event in w that is pertinently similar to the mental event c in the actual world, then mental event c* instantiates mental property F in w, and mental event c* causes in w an event e* that both instantiates property G and is pertinently similar to e of the actual world.[12]

This principle asserts a form of dependence, or supervenience, of the properties F and G upon other properties of the c* and e* events of the

---

[12] Horgan, "Mental Quausation," 61.

various pertinently similar worlds associated with the causal transaction between c and e, says Horgan. That is, in each pertinently similar world the mental event c* has F and an e* event occurs caused by c* which has the property G. The above principle expresses the supervenient relevance of mental property F and property G to the causal transaction between c and e.

Horgan illustrates the plausibility of his view of the quausal relevance of mental properties with an example:

> Consider a case where a mental event causes an action: Fred experiences a desire for a beer, and this desire causes him to walk to the refrigerator. The mental cause c has the property *being a desire for a beer*. Now, there are pertinently similar worlds in which Fred experiences a desire for a beer in a situation which is modestly different from the actual situation - for instance, worlds in which Fred is located in a slightly different position in his house than his actual-world position at the time he experiences the desire for a beer. In these pertinently similar worlds, Fred's desire causes a sequence of bodily motions that is somewhat different from his actual-world motions; however, the otherworldly sequence still qualifies as an action of walking to the refrigerator, and hence counts as pertinently similar to Fred's actual-world action.[13]

The problems that arise here with regard to the quausal relevance of mental properties, as in the generic case, are to do with what is to count as a pertinently similar world. Horgan himself raises one such problem for his account : are there pertinently similar worlds in which a mental event occurs that is pertinently similar, say, to Fred's mental event c and yet does not have the F property of being a desire for a beer? No, says Horgan. Although there might be *possible* worlds in which an event occurs that is physically just like the physical event in the *actual* world which is the

---

[13] Ibid., 62.

mental event with the property of being the desire for a beer and in these possible worlds these physical events might cause motions just like the motions of going to the fridge although they lack the property F of being a desire for a beer, these *possible* worlds will not count as *pertinently similar* worlds. They are just too different from the actual world.

Horgan points out that the changes one would have to make to the actual world in order to get such a world are so great that what happens in them would have no relevance to what happens in this our actual world. For example in order to get such a possible world one might have to make changes in the world external to Fred's brain, make it into Twin Earth so that although physically Fred is the same the content of his desire is different. It is a desire for schmeer composed partly of XYZ rather than a desire for beer composed partly of water. But, asks Horgan, why should one consider what happens in this far away possible world relevant to what happens in the actual world? Other changes one might make to the actual world in an attempt to get a relevant possible world are ones that are head internal. In this case one might leave intact the physical event that realized the desiring and alter the neural wiring so the physical event figures so differently in the internal causal nexus that it no longer tokens the property of being the desire for a beer. But again, why count this as a *pertinently* similar world? The actual world neural wiring, being an important background condition, should surely remain the same within all the worlds that are pertinently similar relative to the causal transaction in question.[14]

Horgan's position, then, so far, as we understand it, is that neither subsumption by strict laws nor subsumption by cp laws is a necessary or a sufficient condition for the quausal relevance of properties. He argues

---

[14] Horgan, "Mental Quausation," 63.

persuasively that properties, when they are explanatorily relevant to a given causal transaction between two events, fit into suitably rich patterns of counterfactual relations. This is a necessary condition that properties in general, and psychological properties in particular, satisfy when they are explanatorily relevant to a particular event.

4:2 Quausal Relevance, Soft Laws, and Explanation

In two later papers Horgan clarifies his position on explanation.[15] It turns out that although fitting into rich patterns of counterfactual dependence is a necessary condition on quausally relevant properties, it is not sufficient to qualify them to play a genuine explanatory role in correct explanations. Properties play this role only when they satisfy another condition. Fortunately it turns out that the properties that fit these patterns satisfy this second condition in virtue of satisfying the first. Horgan argues that the properties that fit into these patterns of counterfactual relations are the very same properties that are subsumed by laws, and that it is in virtue of this latter fact that they figure in correct explanations.

We have argued at some length in the preceding chapters that since none of the usual candidates for psychological laws appears to qualify as a covering law it seems unlikely that our commonsense psychological explanations do appeal to laws. In this chapter we have laid out Horgan's account of what it is that makes a property quausally relevant carefully because it seems to us a plausible account with regard to properties in general and psychological properties in particular. Since it turns out that

---

[15] Terence Horgan and John Tienson, "Soft Laws," Midwest Studies in Philosophy XV (1990): 256-279. Also Terence Horgan, "Nonreductive Materialism and the Explanatory Autonomy of Psychology," in Naturalism: A Critical Appraisal, ed. Steven J. Wagner and Richard Warner (University of Notre Dame Press, Indiana, 1993): 295-320.

although subsumption by laws of any kind is neither a necessary nor a sufficient condition on quausal relevance, it *is* a condition that properties satisfy when the role they play in correct explanations is a genuine explanatory role, it seems essential at this point to consider the nature of the laws that Horgan suggests that this counterfactual feature of psychological properties is linked to, the nature of the link itself, and whether or not the properties' being linked to such laws is a necessary condition on the correctness of the explanations in which these properties figure, as Horgan maintains that it is. Only then will we be in a position to decide whether we should reconsider our view on the existence of psychological laws and the need for them in psychological explanation or whether we should turn our attention to looking for an account of explanation that not only makes no appeal to laws in order to account for the quausal relevance of properties, but neither does it invoke them in order to account for the correctness of explanations in terms of those properties.

We shall start by examining Horgan's view about the nature of the link between these two features of properties, viz., their being subsumed by laws and their entering into patterns of counterfactual dependence, both in the general case and in the particular case of psychology. In his opinion the nature of the connection between the laws of the natural sciences and the patterns of counterfactual dependence among the properties they subsume is that the counterfactual dependence relation between properties is related to the very structure of scientific laws. The generality of the fundamental laws of the natural sciences consists not merely in their having the logical form All A's are B's, he says, but rather in the fact that

they are systematic in scope and structure, so that a wide range of phenomena are subsumable under relatively few laws. One major source of their systematicity is that (i) the laws cite *parameterized* properties - viz., quantitative magnitudes, where the parameters are numerical values that these magnitudes can take on when instantiated; and that (ii) the laws contain universal quantifiers ranging over the values of these parameters (in addition to the universal quantifiers ranging over the non-numerical entities in a law's domain) ... The resultant generality of a physical law consists largely in the existence of a whole (typically infinite) set of specific nomically true principles, each of which is a specific instantiation of the law with specific numeric values "plugged in" for the quantitative parameters. Rich patterns of counterfactual dependence, of the sort that are a crucial feature of successful scientific explanation ... are reflected by the truth of such sets of specific law instantiations.[16]

The same applies to the laws of intentional psychology, both folk and theoretical, he says. Horgan's position with regard to commonsense psychology is that the robust patterns of counterfactual dependence among the state types and act types posited by folk psychology are systematizable via generalizations containing universal quantifiers ranging over *intentional* or *propositional* parameters (i.e. they are the kinds typically specified by 'that'-clauses) in place of the *quantitative* parameters over which the universal quantifiers range in the physical sciences. For example, the rich pattern of counterfactual dependence of action types upon intentional mental state types is undergirded by the following generalization:

(G): (S)(D)(A) (If S wants D and S believes that doing A will bring about D, then ceteris paribus, S will do A)

---

[16] Horgan, "Nonreductive Materialism and the Explanatory Autonomy of Psychology," 297.

And the patterns of counterfactual dependence among mental states themselves are undergirded by generalizations like the following:

(x)(p) [If Wants(p)x & Discovers(p)x, then *ceteris paribus*, Pleased(p)x]
(x) (p) [If Fears(p)x, then *ceteris paribus*, Wants(— p)x]
etc.

Horgan adds: "[W]anting, believing, etc., are seen in these laws as vast (possibly infinite) highly structured arrays of properties, a different specific property for each value of *p*. Correlatively, each of these candidate laws implies numerous more specific (putative) nomic principles that are instances of the (putative) law."[17]

The assumption is that if the counterfactual relations between properties are undergirded by such laws, then it will be clear that the counterfactuals hold in more than one case, that they will not merely relate one specific thing to another specific thing. In the case of psychology, Horgan says. "imagine that the only true psychological generalizations mentioned specific propositional attitudes, perhaps in a form like; when a cognizer is in S1, he goes into S2, where S1 and S2 are lengthy specifications of the specific beliefs, desires and so forth in total psychological states. And suppose there were no interesting structure or systematicity to the set of such transition principles, so that, for example, they could not be specified recursively. [S]uch a hodgepodge would not capture generalizations in a way that would give point to seeking out laws at the psychological level in the first place. They would not constitute a

[17] Horgan and Tienson, "Soft Laws," 260.

theory and would not give us any understanding of what they permitted us (piecemeal) to predict."[18]

So Horgan's view seems to be that it is intrinsic to the very nature of psychological properties both to enter into rich patterns of counterfactual dependence and to be subsumed by laws and that were it not the case that the patterns of counterfactual dependence among psychological properties are systematizable by psychological laws, our explanations of behavior would lack generality and our predictions would not be accompanied by scientific understanding.

In a paper written in collaboration with John Tienson, Horgan describes the kind of psychological cp law that supposedly undergirds the counterfactuals and provides the generality or systematicity required for correct explanation.[19] He refers to them as "soft laws." One essential characteristic of soft laws is that the cp clause they contain is ineliminable. This clause is essential both because there are limitless same level properties that could constitute exceptions - deflectors or defeaters - and because laws with cp clauses represent the type of law that is fundamental in certain domains.

Horgan and Tienson explain that although the generally accepted view is that scientific laws are either strictly exceptionless or at least quasi-exceptionless in that they have no exceptions statable in the vocabulary of that science, this view is mistaken. There are domains in which same-level exceptions are to be expected, and psychology is one of them. What such a

---

[18] Ibid., 260.

[19] Horgan and Tienson, "Soft Laws," Midwest Studies in Philosophy XV (1990): 256-279.

law expresses, they say, is a defeasible causal tendency. Or rather, each instance of, say, (G) above, expresses a defeasible causal tendency.

> The cp clause in (G) adverts to the fact that in each instance, other psychological states might interact with S's beliefs and desires in a way that would bring it about that she does not do A. Each of these possible defeaters is itself an instance of (G) or a similar law expressing a different array of content-determined defeasible causal relations. Thus what can defeat a defeasible causal tendency is another defeasible causal tendency — a tendency to an incompatible state or act. [T]he fact that a cognitive system is subject to certain defeasible causal tendencies is a basic or fundamental fact about that system. A psychological law such as (G) is a generalization about such defeasible causal tendencies, where the generality comes from quantifying over intentional-propositional parameters. As such it is no part of the laws' role to delimit all the possible ways that this tendency might get defeated when interacting with other causal tendencies.[20]

Besides having the three essential characteristics mentioned here - intentional parameters, *ceteris paribus* clauses and essentially limitless same-level exceptions - the soft laws of psychology are likely to exhibit three other attributes, say Horgan and Tienson.

First, soft laws are likely to be indeterminate in that they will not predict a determinate outcome for every situation to which they apply.[21] This is so because the cognitive systems that conform to them are likely to be indeterminate. And the reason for this indeterminacy, that is, for the likelihood of there being no fact of the matter about which defeasible causal tendencies will win out, is that psychological states are multiply realizable and that the particular psychology-level outcome will often

---

[20] Ibid., 267.

[21] Ibid., 270.

depend on details of realization which are left open by any psychology-level description.[22]

Secondly, the concepts that figure in soft laws in general and in soft psychological laws in particular, are likely to be vague, that is there are likely to be situations for which there is no definite fact of the matter about whether or not the concept applies since the conditions for applying these concepts are likely to include soft laws.

Thirdly, the concepts employed in soft psychological laws are likely to be non-quantitative, their vagueness being of the kind likely to preclude the precise measurability that quantitative magnitudes usually require but not of the kind that will preclude their use in comparisons. The generality that is conferred by quantitative parameters is already a feature of these laws by virtue of their containing quantifiers that range over intentional parameters.

Horgan and Tienson hold that these soft laws function as covering laws in explanations. "Causal explanations that rest upon soft laws presuppose that the law's *ceteris paribus* clause was satisfied in the given situation. That *ceteris* is *paribus* is an implicit premise in the explanation. ..... [D]eductive explanations of this kind are faithful to an intuitive idea that largely motivates the D-N model in the first place, viz., 'like causes - like effects.' .... In any other case *exactly like that one*, *ceteris* would again have been *paribus*, and a like action would have occurred. Indeed, in any case *relevantly* like this one, *ceteris* would have been *paribus*. We cannot - in practice, and probably in principle - spell out all of the ways that things

---

[22] Ibid., 269.

could fail to be relevantly like the present case. But that does not alter the fact that where a soft law applies, same causes do lead to same effects."[23]

Above we have described why Horgan and Tienson think that laws are needed in commonsense psychological explanation and the nature of the link between these laws and the patterns of counterfactual dependence into which properties which are causally related enter. We have also summarized some of the main points that Horgan and Tienson make about what they refer to as the "soft laws" of psychology.

4:3 The Role of Soft Laws

Let us examine Horgan's and Tienson's view about soft laws and explanation. A putative psychological cp law, which, we have suggested previously, one might think of as having the form "Ms cause Bs, cp," is understood by them in terms of a universally quantified conditional statement:

(G) (S)(D)(A) (If S wants D and S believes that doing A will bring about D, then ceteris paribus, S will do A).

In addition we have seen that they hold that "[w]hat (G) expresses, ... , is a *defeasible causal tendency.* More exactly, each instance of (G) expresses a defeasible causal tendency."[24] It does not seem unreasonable to interpret the suggestion that psychological laws express defeasible causal tendencies in terms of defeasible sufficient conditions, to understand it to be the suggestion that M is defeasibly nomologically sufficient for B in the

---

[23] Ibid., 271.

[24] Ibid., 267.

absence of any defeating conditions. However if one does accept this as a reasonable interpretation, a problem arises. There is a standing objection in the literature to an analysis of psychological cp laws in terms of defeasible sufficient conditions. Schiffer points out that it is not plausible to suggest that an intentional condition, such as M, is a defeasible sufficient condition for B because it cannot be the case that M is nomologically sufficient for B *in the absence of some defeater.* The only plausible suggestion, he says, is that M is *never* nomologically sufficient for B, but that it can be *supplemented* in a way that would yield a nomologically sufficient condition for B.[25] This, however, does not appear to be what Horgan and Tienson are suggesting.

A different problem with the suggestion of soft laws for psychological cp laws is to do with the question of the multiple realization of intentional states. It may be the case that although M, under some realizations, will win out over other defeasible causal tendencies, and cause a B, it is likely that M, under other realizations, will not. Since on these latter occasions M will be defeated by other causal tendencies, even in cases in which no defeating conditions obtain, those in which ceteris is paribus, or paribus enough, M will not cause a B. So "Ms cause Bs, ceteris paribus," cannot be a law, since even when ceteris is paribus, Ms do not always cause Bs.

There are other grounds for doubting whether or not "Ms cause Bs, cp" warrants talk of a psychological law on the soft law interpretation. It would seem that the notion of "defeasible causal tendency" would need to be explicated before it would be clear whether or not this suggestion provides a genuine explanatory role for M. It seems likely that when this is

[25] Chapter Two was devoted to a discussion of this suggestion.

done the problems incurred by the multiple realization of M that arose for cp laws with lower-level completions will arise here. The difficulty, we may recall, was that once the multiple realization of psychological states had been taken into account in a formulation of the conditions on explanation, it turned out that M did not appear in a law that would qualify as a psychological covering law.[26]

And, finally, it seems that Horgan himself has an equivocal attitude with regard to the suggestion of a soft law condition on commonsense psychological explanation. Let us grant for a moment that soft laws are laws and that they do play the role of covering law in commonsense psychological explanations. Let us take as an example of such an explanation Barry Loewer's flower shop example, viz., Barry remembered that the following day was his wife's birthday because he noticed a flower shop yonder. According to Horgan's and Tienson's view of psychological explanations as deductive-nomological explanations with soft psychological cp laws as covering laws, the full explanation will be:

Ceteris paribus, a (married) man who notices that there is a flower shop yonder will remember that the following day is his wife's birthday.
Barry, (a married man), noticed that there was a flower shop yonder.
Ceteris was paribus

Barry remembered that his wife's birthday was the following day.

We saw earlier that Horgan thinks that cp law subsumption is not a necessary condition on a property being explanatorily relevant to a causal

---

[26] See Chapter Two, Section 2: 3.

transaction between two events. In fact we may recall that he used this very same example to demonstrate this point: "Suppose for instance that Barry's noticing a flower shop causes him to remember that tomorrow is his wife's birthday. The properties *being a noticing that there is a flower shop yonder* and *being a remembering that tomorrow is one's wife's birthday*, certainly appear explanatorily relevant to the causal transaction : yet generalizations like

Ceteris paribus, a (married) man who notices that there is a flower shop yonder will remember that the following day is his wife's birthday

seem just false."[27]

Now if Horgan were to acknowledge, as he surely would, that the explanation that Barry remembered his wife's birthday because he noticed a flower shop yonder is likely to be correct, and if he were to maintain his view about the properties that enter into correct explanations being subsumed by soft laws, it seems that it would be incumbent upon him to come up with a suggestion for a soft law that subsumes the properties cited in this explanation and plays the role of covering law in it, but which does not seem just false. But what could that be? It does seem that the cp generalization cited might be not merely the most likely candidate to play the role of covering law in this explanation but maybe the only one that *could* do so. However, the fact remains that the "law" is clearly false and that the explanation is just as clearly a satisfactory explanation. We conclude that Horgan's own reasoning seems to indicate that subsumption

[27] Horgan, "Mental Quausation," 55.

by cp laws, soft or otherwise, does not appear to be necessary in order for properties to play a genuine explanatory role in correct explanations.

In this section we have attempted to show that it is unlikely that the suggested soft laws actually do qualify as laws and that Horgan's rejection of cp generalizations in his discussion of quausal relevance in virtue of their apparent falsity seems to be at variance with his later re-introduction of them as laws capable of playing the role of covering law in commonsense psychological explanation.

Although these difficulties with soft laws lead us to reject Horgan's view of explanation, this rejection does not affect our endorsement of his view of the quausal relevance of properties. What we are in search of now is a view of explanation that accommodates the latter view.

4:4 Explanation and Pragmatic Relevance

Stephen Schiffer has recently outlined an account of explanation whose conditions do not include the subsumption of psychological properties by laws.

In his view the explanatory role or "causal relevance" of psychological properties is nothing over and above their ability to occur in true propositional-attitude 'because'-statements. The question he sets out to answer is: What is it that will make an instance of "*x* Fed because *x* Ged," true?[28] The first essential condition is (i) that *x*'s Ging caused *x*'s Fing. However, Schiffer remarks, as Davidson pointed out, since the satisfaction of this condition alone implies nothing about the relevance of G to the actual causal transaction there is need for a second condition, viz., the counterfactual condition (ii) that *x*'s Ging would not have caused *x*'s Fing if

[28] Stephen Schiffer, "Ceteris Paribus Laws," Mind 100 (January 1991): 10.

it had not been a Ging, that is, if that event had not had the property of being a G event. (Schiffer says that he prefers to refer to the conditions (i) and (ii) as the essential core of explanation rather than as necessary conditions since in some rare cases of causal overdetermination, the counterfactual condition will not play a part in selecting the property that is relevant to the causal transaction between events, a fact which he chooses to ignore in this initial sketch of his view).

We can see from these causal-counterfactual conditions on true commonsense propositional attitude 'because'-statements that Schiffer agrees with Horgan's view that quausally relevant properties are those that enter into rich patterns of counterfactual dependence. Schiffer also agrees with Horgan that satisfying these causal-counterfactual conditions alone does not ensure that the properties cited in propositional attitude 'because'-statements play a genuine explanatory role in correct explanations. He points out that even if the essential-core conditions (i) and (ii) are satisfied, that will still not be sufficient for the truth of a 'because'-statement since:

> [I]t is *not* the case that
>
> > *x F*ed because *x G*ed *if* (E*e*)(E*e'*)(*e* caused *e'* & *e* is a *G*ing by *x* & *e'* is an *F*ing by *x* & *e* wouldn't have caused *e'* if *e* hadn't been a *G*ing by *x*).

For it may be that *e* caused *e'*, that *e* was something that Hugo did in the presence of Regina, that *e'* was Hugo's becoming embarrassed, and that *e* wouldn't have caused *e'* if *e* hadn't been something that Hugo did in the presence of Regina. But we shouldn't on that account say that Hugo became embarrassed because he did something in the presence of Regina - not when, as it happens, he became embarrassed because he burped in her presence.[29]

---

[29] Ibid., 13.

Schiffer holds that the reason that the 'because'-statement - "because he did something in the presence of Regina" - does not constitute a genuine explanation of why Hugo was embarrassed is not because the property of being something that is done in the presence of Regina is not subsumed by a law (soft or otherwise), but simply because this property of the event that caused Hugo's embarrassment is not pragmatically relevant to the 'why'-question to which it is cited as an answer, viz., "Why was Hugo embarrassed?" The person who asks *this* question is expecting the answer to cite a certain kind of property, a kind that satisfies certain conditions of pragmatic relevance, conditions that the kind of property, of which being something that is done in the presence of Regina is an instance, does not satisfy. The inquirer's question concerns which properties of that kind are operative on the occasion in question.

So according to Schiffer, whether or not a substitution instance of '*x Fed because x Ged*', yields a bona fide explanation, depends not only on the satisfaction of the essential causal-counterfactual conditions but also on the satisfaction by the properties cited in that 'because'-statement of a certain pragmatic condition. The pragmatic condition on explanation concerns the nature of the concepts instantiating F and especially G. It is to do with the interests of the person who asks the why-question to which the 'because'-statement is an answer.

Schiffer's suggestion for a third condition that is satisfied by a true 'because'-statement is that the properties cited in such a statement will satisfy the following condition of pragmatic relevance:

(iii) Suppose there were kinds of properties K and K' such that (a) events having a property of kind K' are typically caused by events having a property of the kind K, (b) the latter events wouldn't have caused the

former events without their kind K properties, (c) there is a reliable practice of predicting events having properties of kind K' on the basis of events having properties of kind K , and (d) properties of kind K are relatively manipulable and epistemically accessible. Then, (he suggests), properties of kind K will have satisfied the pragmatic and other concerns necessary to earn them a place in true because-statements that explain the occurrence of events having properties of kind K'.[30]

It is clear that the property of being something done in the presence of Regina is not an instance of a kind of property that satisfies this condition on the pragmatic relevance of properties. Propositional attitude properties, on the other hand, do satisfy the pragmatic conditions mentioned here as well as satisfying the causal and counterfactual conditions, Schiffer says.

Insofar as the satisfaction of (d) is concerned he points out that since we can often affect the beliefs and desires of others and are often in a position to have knowledge of their beliefs and desires, it would seem that propositional-attitude properties are both relatively manipulable and epistemically accessible. With regard to (c), Schiffer explains that we are often able to reliably predict someone's behavior on the basis of the propositional attitudes we ascribe to him and that this is largely because of the way the human belief mechanism works.

Schiffer's suggestion for what underlies the reliability of the practice of predicting behavior on the basis of the ascription of propositional attitudes does not involve tacit knowledge of a folk theory nor implicit deployment of laws, nor our working in accordance with rules or instructions. His view is that what we acquire when we acquire folk psychology is folk psychological concepts with complex functional roles. It is the functional roles of our psychological concepts that make possible the

---

[30] Ibid., 14.

ascription of propositional attitudes and on the basis of those ascriptions the prediction and explanation of action, he says.[31] He suggests that the complex ability that we gain when we acquire psychological concepts with complex functional roles is to be understood in terms of the acquisition of a subjective conditional probability function. This subjective conditional probability function has two aspects, he says. One of these is primarily of use in the formation of beliefs about the beliefs and desires of others and the other primarily of use in the formation of predictions and explanations of what others will do. As we understand it, both aspects come into play in the actual practice of prediction.

A prediction of behavior has two stages, says Schiffer. We take it that the first aspect of our subjective conditional probability function is used in the first stage which consists in the formation of beliefs about the probable beliefs and desires of another and in making an inference from those beliefs and desires to his intention to act. This inference will be a probabilistic version of the following generalization in which both the strength of the predictee's belief and desire and the probability with which the intention to act will occur, given beliefs and desires of those strengths, is specified in terms of the predictor's subjective probability:

(a) If *x* desires *p* and *x* believes [that *p* if *x* does A] & *x* believes that *x* knows how to do A & and *x* has no stronger competing desires, then *x* intends to do A.[32]

---

[31] Ibid., 15.

[32] Schiffer, Seminar The Natualization of Content, given at the Graduate Center of the City University of New York (Winter 1992).

With regard to the second aspect of our subjective conditional probability function Schiffer says: "Acquisition of our propositional-attitude concepts carries with it a disposition to move to a high subjective probability that *x* will Ø when we believe that *x* has such-and-such propositional attitudes and our subjective conditional probability function is such that, for any proposition *p* that we believe, our subjective probability that *x* will Ø given (p and that *x* has those propositional attitudes) is high."[33] So the second aspect of our subjective probability function is used in the second stage in making a prediction which consists of the predictor making an inference from the probable intentions of the predictee to his probable behavior, in making a probabilistic version of an inference such as the following:

(b) If *x* intends straightway to do A & A is a basic act type & *x* isn't prevented from doing A, then *x* will do A.[34]

What is going on in this latter type of inference, says Schiffer, is that the predictor is making certain assumptions about the predictee. Some of these are to do with the predictee's background beliefs and desires, - that it is unlikely that any of them will conflict strongly enough with the intention to act such that they will override it, and some of them have to do with the external circumstances in which the predictee finds himself at the time he forms his intention to act, - that it is unlikely, given these circumstances, that events will occur that will cause him to change his intention. There will doubtless be myriad other similar assumptions. What these different

---

[33] Schiffer, "Ceteris Paribus Laws," 14.

[34] Schiffer, Seminar The Natualization of Content, (Winter 1992).

assumptions amount to is the predictor's basic assumption that the person whose behavior he is predicting is reliable with regard to his numerous beliefs about his immediate environment and about how to successfully interact with it. The predictor is assuming that these beliefs are true, that, in fact, the interaction of the predictee's basic beliefs about the world and his commonplace desires will result most of the time in the satisfaction of those desires. In a word, he is assuming that the predictee is reliable insofar as his basic beliefs, desires, intentions to act and beliefs about basic acts are concerned.

The reliability of a predictor's subjective probability inferences then, stems not from the fact that he knows all of the relevant beliefs of the predictee, (nor from the fact that he knows the soft law under which the other's motivating beliefs and desires might fall), but rather from the fact that he knows that most of the other's beliefs in the area relevant to this prediction are likely to be reliable. He makes assumptions both about the other's overall head-world reliability, i.e. that most of the beliefs he forms about how to carry out an intention will be true, and he makes assumptions about the other's overall world-head reliability, i.e. that he is a reliable monitor of his relevant environment.

Schiffer acknowledges that it is hard to account for the reliability of those whose behavior we predict, to explain the overall reliability of, say, Bob, for each of the many instances of

$$P(p \text{ is true} \mid \text{Bob believes } p) = r$$
$$P(\text{Bob believes } p \mid p \text{ is true}) = r$$ [35]

[35] Schiffer, Seminar The Natualization of Content (Winter 1992).

A person's reliability supervenes on just about everything, he says: facts about the nature of his belief-forming mechanism, facts about his standing beliefs over time, facts about the sensory stimulations he experiences in the actual world and those he would experience in counterfactual circumstances, facts about the causes of these sensory stimulations both in the actual external environment and those in near-by possible worlds. All of this plays a role in determining a person's over-all reliability. In spite of the fact that it is hard to account for reliability, Schiffer maintains that this belief mechanism is reliable, that beliefs formed in this way about what people will do tend very often to be true.[36]

To summarize, Schiffer's suggestion is basically that a propositional-attitude 'because'-statement is true when it satisfies the causal-counterfactual conditions (i) and (ii), and when the properties referred to in the 'because'- statement satisfy the conditions for pragmatic relevance stated at (iii).

4:5 The Primacy of Practice

Tyler Burge has expressed a view recently in a paper discussing mental causation which appears to support Schiffer's basic position on psychological explanation, although Schiffer would not agree with all of Burge's intuitions.

Burge appeals directly to the fact that there exists an *explanatory* practice which uses the framework of mentalistic notions in order to support his view of the causal relevance of the mental. His position is that "[S]ystematic, informative, important explanatory schemes of events and states are also our strongest indication of causal relevance.... Our

---

[36] Schiffer, "Ceteris Paribus Laws," 14.

understanding of mental causation derives not primarily from re-descriptions in physical terms. It derives primarily from our understanding of mentalistic explanation. This understanding is largely independent of reference to the underlying processes."[37] In philosophical discussions about mental causation too much weight is usually given to materialist metaphysics, he says, and too little to explanatory practice.

What is important is that these explanations work, he says. "As long as mentalistic explanation yields knowledge and understanding, and as long as that explanation is (sometimes) causal, we can firmly believe that mind-body causation is a part of the world. The primary way of understanding such causation is by understanding mentalistic causal and explanatory statements in the ordinary, non-philosophical sense of 'understanding'. ..." "[W]hat matters is that our mentalistic explanations work and that they do not conflict with our physicalistic explanations."[38]

Burge accounts for the capacity of mentalistic explanations to yield knowledge and understanding, for their being genuine explanations of behavior, not in terms of the underlying reliability of people's basic beliefs, nor in terms of the existence or deployment of some sort of underlying psychological law, but in terms of the concept of human agency. He points out that most of our intellectual and practical norms and evaluations presuppose that we *are* agents. However, says Burge, "If our willing or deciding made something happen, but that event's being a willing or deciding were not causally efficacious (so that the efficacy resided in some underlying neural property), then the agency would not be ours. If our theoretical deliberations were not ours to control, we could not see

---

[37] Tyler Burge, "Mind-Body Causation and Explanatory Practice," 111.

[38] Ibid., 117.

ourselves as being the authors of our theories; nor could we criticize ourselves as deliberators. Most normative evaluations of our intellectual and practical activities would be empty."[39] He adds that if intentional explanation merely rationalized behavior, as opposed to citing what is causally relevant to it, then if, as he believes is the case, the raison d'etre of such explanations is to help us understand ourselves as human agents, it would lose its point. What seems to be hovering in the background of Burge's view is the suggestion that either mentalistic explanations are (sometimes) correct causal explanations or human beings never actually *do* anything. Things just take place inside of them.

Burge's view then is that cognitive practice, and in particular explanatory practice, is central to understanding mental causation. The causal powers of an event are to be understood in terms of the patterns of causation that events of that kind enter into and these patterns are those we see in causal explanation. Since such patterns are identified as explanatory because they constitute informative explanations the highly informative mentalistic pattern qualifies par excellence.

Although Schiffer would probably agree with Burge about the importance of the success of folk practices in relation to the legitimacy of explanations in terms of propositional attitudes and about the importance of the notion of human agency in the qualification of these explanations as genuine, it is unlikely that he would endorse the view that the existence of an informative explanatory practice of accounting for the occurrence of events with one type of property in terms of the occurrence of events with another type is sufficient evidence from which to conclude that the latter type of property is causally relevant to the former, and that the

---

[39] Ibid., 119.

explanations in which they figure are correct. For it seems that Burge's suggestion might be vulnerable to the following type of counterexample. Suppose there exists a possible world in which everything is the same as in the actual world except that there exists in that world a practice of explaining events with behavioral properties in terms of events with supernatural properties. Suppose also that the explanatory scheme of events and states to which these events belong in virtue of their having these two types of properties is an important, systematic, informative scheme in this world in that the supernatural causal and explanatory statements based on it are understood, in the ordinary, non-philosophical sense of 'understanding.' It would seem that explanations of behavioral events on the basis of events with supernatural properties would qualify as genuine, correct explanations on Burge's account. We can see from the robustness of the view of mental causation that we have described above, it is not the case that Burge is promoting any sort of deflationary view of what is to count as a correct psychological explanation.

The problem seems to disappear if one takes predictive practice as prime, as Schiffer does, since it seems likely that supernatural properties will not satisfy the condition of there existing a reliable practice of predicting behavioral events on their basis. *Systematically successful* predictions of events with one kind of property on the basis of events with another kind do seem to provide better evidence for these types of property being causally connected in some way and our for explanations in terms of them being correct than does the mere existence of an explanatory practice involving an "informative" scheme of states and events.

4:6 A Weaker Counterfactual View of Causal Explanation

In a recently published paper defending a counterfactual theory of commonsense causal explanation David-Hillel Ruben offers an objection to Schiffer's account. Ruben says that Schiffer's conditions on the properties that figure in correct explanations of this type are not necessary conditions. Many more properties actually figure in such explanations than Schiffer's conditions allow, in Ruben's opinion.

Overgeneral properties, such as that of being something done in the presence of Regina, do not need to be ruled out because they do provide some kind of an explanation, says Ruben. His view is that there is a difference between the properties of an event that are totally irrelevant to a particular causal transaction and those that are explanatorily relevant but overgeneral. Suppose, he says, that the event that has the property of being a burping in the presence of Regina also has the property of being the disturbance of a spider in Room A212. He points out that whereas the latter property is totally irrelevant to Hugo's embarrassment,

> if that token event is a burping by Hugo in the presence of Regina, it must also be a doing of something by Hugo in the presence of someone. The event qua a doing of something by Hugo in the presence of someone is not strictly explanatorily irrelevant to Hugo's embarrassment, in the way in which its being the disturbance of a spider was. It is simply not specific enough to capture all of what is explanatorily important, viz., that it was a burping by him in the presence of Regina, even though it [will satisfy the causal-counterfactual conditions] since it is entailed by the property that does capture all of what is explanatorily important. I disagree with Schiffer about the explanatory force of overgeneral properties. He claims that they do not explain at all; I think that he should have said that they may explain only to a very small, perhaps vanishingly small, extent.[40]

---

[40] David-Hillel Ruben, "A Counterfactual Theory of Causal Explanation," Nous XXVIII (December 1994): 470.

According to Ruben, since overgeneral properties do not need to be ruled out, the condition on the pragmatic relevance of properties, the condition which accomplished this task in Schiffer's account, is not needed. In fact, for *partial* explanations such as "Hugo was embarrassed because he did something in the presence of Regina," the satisfaction of the causal-counterfactual conditions by the properties cited is sufficient, says Ruben.[41] When it comes to *complete* causal explanations, however, or, in his words, to explicating "... fully (or completely) causally explains ...," more is needed. What one needs, according to him, is a condition that selects the most determinate property in the hierarchy of logically related properties at least one of which satisfies the causal-counterfactual conditions. He suggests the following;

> Let 'h' be a hierarchy of logically related properties, at least one of whose members satisfies the causal-counterfactual conditions (a) and (b) below:
>
> (CE) the F fully causally explains the G iff (a) the F caused the G, and (b) if there had been another token event c* in place of the F, which failed to be an F, c* would not have caused the G, and (c) in the hierarchy of properties, h, to which F belongs, F is the most determinate property in h that still meets conditions (a) and (b).[42]

Ruben goes on to raise a different problem for Schiffer's account. Schiffer's counterfactual condition successfully rules out overspecific properties, but, he says, what if such properties are explanatorily relevant? What if the fact that the event that caused Hugo's embarrassment had the property of being a burping while wearing a blue shirt was explanatorily relevant to his embarrassment? Such a problem with the counterfactual

---

[41] Ibid., 471.

[42] Ibid.

condition on correct explanation will arise for those who think that the inclusion of overspecific properties do not ruin an otherwise acceptable explanation but merely render it less good, those who hold that it does explain although to a lesser degree than one that cites a specific property.

In order to accommodate this situation Ruben introduces two other categories of explanation, (in addition to the above-mentioned full and partial ones); succinct and excessive ones. He points out that only the properties cited in succinct explanations (full or partial) satisfy the counterfactual condition stated in his account at (b) of (CE). The properties cited in excessive explanations however can be accommodated by his account in virtue of the following corollary of (CE):

> Say that a property F provides a succinct explanation for the G (whether full or partial) if F is a property for which (b) is true. Then consider any property H which is more determinate than F, but for which (b) is false. Say that H provides an excessive explanation for the G (whether full or partial) just when there is some property F which provides a succinct explanation for the G and to which H is related in the way I have described.[43]

So Ruben's position is that overgeneral and overspecific properties of cause-events can and do enter into correct explanations of other events. The conditions stated at (CE) and in the corollary above on the properties that play a genuine explanatory role allow full succinct, partial succinct, full excessive, or partial excessive explanations. In contrast, Schiffer's conditions rule out all properties except those that take part in full succinct explanations.

---

[43] Ibid., 472.

Let us look at how a typical commonsense psychological explanation fares under Ruben's analysis. Let us take the example "Barry remembered that his wife's birthday was the following day because he noticed a flower shop yonder." This commonsense explanation cites the property of being a noticing of a flower shop yonder as being the property of the cause-event which causally explains Barry's remembering. Since this property appears to satisfy all of Ruben's conditions at (CE), the explanation seems to qualify as "full succinct." Let us now consider another possible explanation of Barry's remembering: Barry remembered his wife's birthday because he noticed a double-decker bus going to Knightsbridge on the evening of December 4th. The property of the cause-event that is being offered as causally explaining his remembering on this occasion is that of its being a noticing of a double-decker bus going to Knightsbridge on the evening of December 4th. Let us see whether or not, on Ruben's account, this property also qualifies to play a genuine explanatory role in a full succinct explanation of Barry's remembering.

Let us assume that the causal-counterfactual conditions at (a) and (b) of (CE) are satisfied by the property cited in the "explanation"; that it was indeed that event's having the property of being a noticing of a double-decker bus going to Knightsbridge on the evening of December 4th that caused the event, Barry's remembering his wife's birthday, and that the remembering would not have occurred had the event that was the noticing of the bus not been a noticing of a double-decker bus ... . In order for this explanation to qualify as "full succinct," however, the property in question must also satisfy (c) of (CE).

And indeed this property does appear to satisfy (c). In the hierarchy of logically related properties to which it belongs, say that of being

noticings of things, it is pretty determinate. In fact since it is difficult to see how one might make it more determinate in a way that might make the putative explanation fuller, it seems safe to assume that it is the most determinate property in that hierarachy that will satisfy the causal-counterfactual conditions (a) and (b). Thus the "explanation" - Barry remembered his wife's birthday because he noticed a double-decker bus going to Knightsbridge on the evening of December 4th - seems to satisfy all of the conditions at (CE), and thus qualify for a "full succinct" explanation of Barry's remembering.

Yet it is hard to see how this property of the cause-event could play a genuine explanatory role in *any* type of explanation of Barry's remembering his wife's birthday, let alone play that role in a "full succinct" one. It simply does not explain Barry's remembering, not even to a vanishingly small extent. Even if it is true that the noticing of this particular bus on this particular occasion played an important role in Barry's internal causal nexus of beliefs and memories, in particular those about double-decker buses, birthdays, his wife, Knightsbridge and the date, December 4th, and even if it is true that this property of the cause-event is the most determinate in the hierarchy of logically related properties to which it belongs, this still does not appear to be sufficient to guarantee that the property cited plays a genuine explanatory role in a correct commonsense explanation.

Schiffer's account handles cases like this well. The property of being a noticing of a flower shop yonder earns a place in a correct explanation of Barry's remembering his wife's birthday because, besides its being the case that the because-statement - "because he noticed a flower shop yonder" - satisfies the causal-counterfactual conditions on correct explanation, the

property referred to in that statement satisfies the conditions on the pragmatic relevance of properties. Events with a property of the type to which being a noticing of a flower shop belongs do typically cause such rememberings, and the former events would not cause the latter events were it not for their having the property of being a noticing of a flower shop, and there is a reliable practice of predicting rememberings of birthdays on the basis of noticings of flower shops and these properties are epistemically accessible and manipulable. When it comes to the putative explanation - Barry remembered his wife's birthday because he noticed a double-decker bus going to Knightsbridge on the evening of December 4th - things are different. Although the causal-counterfactual conditions on the truth of a because-statement are satisfied here (by stipulation), the condition on the pragmatic relevance of properties is not. It is not the case that typically noticings of double-decker buses going to Knightsbridge on the evening of December 4th cause rememberings of birthdays, and it is not the case that events with the property of being a noticing of a bus going to Knightsbridge would not cause events with the property of being a remembering if the former events were not noticings of buses.... . And finally, there is no reliable predictive practice involving these two types of properties. We may conclude from this that Schiffer's account does not allow such properties to qualify to play a genuine explanatory role in correct commonsense explanations.

Ruben's objections to Schiffer are made in the course of establishing his own counterfactual account of causal explanation. The main advantage of his account in his view is that it will accommodate the difficult cases of property pre-emption and overdetermination, properties whose position with regard to explanatory relevance is often left vague in counterfactual

theories of causal explanation. In attempting to account for these rare cases Ruben has relaxed the requirements on explanatorily relevant properties. However, we submit, the merit of his account with regard to these rare cases, will be undermined if it is true, as appears to be the case, that his theory does not provide sufficient conditions for the more common case of correct "full succinct" commonsense explanation.

4:7 The Present State of Affairs Reconsidered

In concluding we shall look briefly at how the view of commonsense psychological explanation defended here fares when considered in relation to issues which have been of concern to us throughout this study, those of the nature of folk psychology, the nature of other types of non-basic explanation, and the dominant explanatory strategy in cognitive science today.

First of all let us consider the pragmatic view of the nature of psychological explanation in relation to the nature of folk psychology. Rejecting the theory theory of psychological explanation, besides having the immediate consequence of necessitating a new and better account of explanation also has the far-reaching consequence that it necessitates a new and better account of the nature of folk psychology. For if it is indeed true, as suggested here, that what motivates the view that folk psychology is a theory is that such a view allows folk psychological explanation to fit the covering law model of explanation, then if the theory theory of psychological explanation is misguided the theory theory of folk psychology no longer has a raison d'etre.

We have seen that the pragmatic account of explanation defended above offers a different view of the nature of folk psychology, one that has

the advantage of not positing tacitly-known psychological covering laws whose existence is so hard to substantiate. The pragmatic view holds that "folk psychology" should be thought of as a name for certain folk psychological practices that human beings indulge in. What we acquire when we acquire knowledge of folk psychology is not knowledge of a theory, but mastery of folk psychological concepts with complex functional roles and a new subjective conditional probability function involving them.

Another consequence of rejecting the theory that psychological explanations are covering law explanations on the grounds that there are no cp laws that could play the role of covering law is that this undermines the widely-accepted view that other types of non-basic explanation are covering law explanations using cp laws as covering laws also. It remains then to say what kind of explanation other types of commonsense explanation and special science explanations can be if they are not of the covering law type using cp laws as covering laws. Insofar as commonsense explanations of the behavior of middle-sized objects are concerned, explanations such as "John broke his neck because he fell out of a window on the sixteenth floor," the pragmatic view holds that these satisfy the causal-counterfactual conditions and the pragmatic condition on explanatory relevant properties described above in the same way that commonsense psychological explanations do.

Special science explanations, however, belong to a different type from that which accommodates these commonsense explanations, according to Stephen Schiffer. He holds that the view that special science explanations need laws is based on a mistaken view of the nature of special science theories. Special science theories are theories about how mechanisms in a

particular domain work.[44] Schiffer's view is that theories about how mechanisms work do not require laws. What is used in a special science explanation are generalizations which describe what happens when things work, that, for example, at a certain point in a process a certain thing takes place. A large mechanism can be broken down into smaller mechanisms, says Schiffer. A fuller, more detailed, explanation will be given when one asks how the component mechanisms work. Schiffer points out that it is worth noting in this regard that there is an implicit pragmatic component in special science explanations, just as there is in commonsense explanations. How correct an explanation is, how well it explains, will depend on its appropriateness to the context in which the 'why'-question is asked. If one goes on in an explanatory chain long enough one will reach a level where laws will be cited, but these will be the laws of physics. In special science explanations laws are not needed, says Schiffer, and it is likely that none will be available at the level of the special science. The explanation of the mechanism stops when it gets to some lower level, to a more basic discipline. At this point, he says, one might feel that there is an explanation, even though one might not be able to give it.

Throughout this study the issue of the cognitive mechanism underlying our different abilities has cropped up again and again. Not only did it seem to be one of the motivating factors in the development of the theory theory but it was also the focus of the simulation theorists' objections to that theory. This being the case it seems appropriate to consider the relation between the view defended here, the view that the mechanism responsible for our ability to give commonsense explanations, the belief-desire mechanism, works in terms of the conceptual roles of our

---

[44] Schiffer, "Ceteris Paribus Laws," 16.

psychological states and of our subjective conditional probabilities, and the popular view that the cognitive mechanism underlying most of our cognitive abilities is the deployment of a tacitly-known theory. It is important to point out that the conceptual role view of the mechanism that is responsible for our ability to give commonsense explanations does not impugn the popular account with regard to the mechanism that underlies many of our other abilities. It does not deny, for example, that such a view might be able to account satisfactorily for our ability to produce and comprehend natural language or for our mathematical ability. The implication is that there is a difference between these abilities and our ability to give explanations. We have seen that with regard to our ability to give commonsense explanations of behavior, there just do not seem to *be* any psychological covering laws that we might tacitly know and whose deployment would account for our ability. What we need to know is what this implies with regard to our other explanation-giving capacities, our capacity to give other types of commonsense explanation and our capacity to give special science explanations.

It seems likely that a theorist who holds that there are no commonsense causal laws that subsume intentional properties will also hold that there are no commonsense causal laws that subsume the properties conferred on states and events by folk physics, properties such as being a breaking of a neck and being a falling from a window. Such a theorist is unlikely to hold that the capacity to give these commonsense explanations is satisfactorily accounted for by the strategy of positing tacit knowledge of a commonsense theory of folk physics. The question that arises is whether or not a version of the conceptual role-subjective probabilities view of our capacity to give commonsense explanations can accommodate commonsense

explanations of this type. It seems to us that it is likely that the ability to give such explanations can be accounted for in a similar fashion. Just as the concepts of belief and desire are acquired at the same time as is a complex ability (in terms of a conditional subjective probability function) to manipulate them according to the conceptual roles they play in our internal causal nexus, it could be the case that the acquisition of the concepts of middle-sized physical objects is accompanied by the acquisition of a subjective probability function with regard to the beliefs that are formed about the roles these objects are likely to play in the external causal nexus. Both views rely on the plausible assumption that human beings are reliable monitors of the world around them, that their beliefs about how it works are true most of the time.

But even if, on closer examination, this turns out to be a plausible suggestion for the mechanism that underlies other types of commonsense explanation, this still leaves our ability to give special science explanations unaccounted for. The problem presented by special science explanations seems to disappear however when one considers it from the perspective of someone who holds that special science explanations are explanations of how mechanisms work. Our ability to give such explanations will be measured in terms of how many of the generalizations that describe the mechanisms in question we have learned. There seems to be no need to appeal a specialized cognitive mechanism in order to account for our explanatory ability in this regard.

The conclusions that we draw from our discussion of commonsense psychological explanation in this chapter are the following. We find that Horgan's account of the quausal relevance of properties provides strong

grounds in support of the conclusion that the properties cited in correct explanations necessarily satisfy the causal-counterfactual condition on explanation. We find that these conditions are not sufficient to guarantee the correctness of such explanations, under any thick construal of "explanation," i.e. when it is taken to indicate "full succinct" ones. We have shown that Horgan's suggestion for the missing condition, viz., that the properties that satisfy the causal-counterfactual conditions are also subsumed by soft laws, not only meets what appear to be unsurmontable problems in the face of multiple realization, but that on Horgan's own account the laws appealed to on these occasions are likely to be just false. Ruben's suggestion for a third condition, as we have just seen, fails to give sufficient conditions for "full succinct" correct explanation, the discovery of which has been our aim throughout this study.

When the above conclusions are added to the conclusions we drew from our discusssions of the theory theory of commonsense psychological explanation and from our discussion of the simulation theory, we find that a serious case has been established in favor of the greater plausibility of a view of the nature of commonsense psychological explanation in terms of the pragmatic relevance of the properties cited over its current competitors. We rest our case then on the suggestion that commonsense psychological explanations qualify as correct explanations of behavior not only because the properties cited in them satisfy the causal and counterfactual conditions on the explanatory relevance of properties, but also because such properties satisfy conditions of pragmatic relevance, because true propositional-attitude 'because'-statements are pragmatically relevant answers to commonsense 'why'-questions.

Bibliography

Abel, T. "The Operation called *Verstehen.*" In Readings in the Philosophy of Science, ed. Herbert Feigl and May Brodbeck, 677-87. New York: Appleteon-Century-Crofts Inc., 1953.

Blackburn, S. "Theory, Observation and Drama." In Mind and Language 7 (Spring/Summer 1992): 187-203

Baron-Cohen, S., A. Leslie, and U. Frith. "Does the Autistic Child have a 'Theory of Mind'?" Cognition 21 (1985): 37-46.

Block, N. "What is Functionalism?" and "Troubles with Functionalism." In Readings in the Philosophy of Psychology 1, ed. Ned Block, 171-84 and 268-305. Cambridge: Harvard University Press, 1980.

Burge, T. "Mind-Body Causation and Explanatory Practice." In Mental Causation, ed. John Heil and Alfred Mele. Oxford: Clarendon Press, 1993.

Chomsky, N. Aspects of the Theory of Syntax. Cambridge: MIT Press, 1965.

_____. Rules and Representations. Oxford: Basil Blackwell, 1980.

Churchland, P. M. "The Logical Character of Action-Explanations." Philosophical Review 79 (1970): 214-36.

_____. Scientific Realism and the Plasticity of Mind. Cambridge: Cambridge University Press, 1979.

_____. Matter and Consciousness. Cambridge: MIT Press, 1988.

_____. A Neurocomputational Perspective. Cambridge: Cambridge University Press, 1989.

Coffa, A. "Hempel's Ambiguity." Synthese 28, 1974: 141-63.

Collingwood, R. G. The Idea of History. New York: Oxford University Press, 1946. Especially Part V. Section I.

Davidson, D. "Mental Events." Chap. in Essays on Actions and Events. Oxford: Oxford University Press, 1980.

Davies, M. "The Mental Simulation Debate." To appear in Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind. Proceedings of the British Academy 83. Ed. C. Peacocke.

Davies, M. and T. Stone. Mental Simulation: Philosophical and Psychological Essays. Oxford: Basil Blackwell, 1994.

Dennett, D. C. The Intentional Stance. Cambridge: MIT Press, 1987.

Dretske, F. Explaining Behavior. Cambridge: MIT Press, 1988.

Field, H. H. "Mental Representation." In Readings in the Philosophy of Psychology 2, ed. Ned Block, 78-114. Cambridge: Harvard University Press, 1980.

_____. "Physicalism." In Inference, Explanation and Other Philosophical Frustrations, ed. J. Earman, 271-91. Berkeley: University of California Press, 1992.

Flavell, J. H. "The Development of Children's Knowledge about the Mind: From Cognitive Connections to Mental Representation." In Developing Theories of Mind, ed. J.W. Astington, P.L. Harris, and D.R. Olson. New York: Cambridge University Press, 1988.

Fodor, J. A. "The Appeal to Tacit Knowledge in Psychological Explanation." Journal of Philosophy 65 (1968): 627-40.

_____. Psychological Explanation: An Introduction to the Philosophy of Psychology. New York: Random House, 1968.

_____. Psychosemantics. Cambridge: MIT Press, 1987.

_____. "Making Mind Matter More." Chap. in A Theory of Content and Other Essays. Cambridge: MIT Press, 1990.

_____. "Fodor's Guide to Mental Representation." In The Future of Folk Psychology, ed. John Greenwood, 22-50. Cambridge: Cambridge University Press, 1991.

_____. "Hedged laws and Psychological Explanation." Mind 100 (January 1991): 19-34.

Goldman, A. I. "Interpretation Psychologized." Mind and Language 4 (Autumn 1989): 165-82.

_____. "In Defense of the Simulation Theory." Mind and Language 7 (Spring/Summer 1992): 104-19.

_____. "The Psychology of Folk Psychology". Behavioral and Brain Sciences 16 (1993): 15-28.

_____. "Simulation and Interpersonal Utility". Unpublished.

Gordon, R. M. "Folk Psychology as Simulation." Mind and Language 1 (Summer 1986): 158-71.

_____. The Structure of Emotions. Cambridge: Cambridge University Press, 1987. Chapter 7.

_____. "The Simulation Theory: Objections and Misconceptions." Mind and Language 7 (Spring/Summer 1992): 11-34.

_____. "Reply to Stich and Nichols." Mind and Language 7 (Spring/Summer 1992): 87-97.

_____. "Reply to Perner and Howes." Mind and Language 7 (Spring/Summer 1992): 98-103.

_____. "Simulation without Introspection or Inference from Me to You." Mental Simulation: Philosophical and Psychological Essays, ed. M. Davies and T. Stone. Oxford: Basil Blackwell, 1994.

Greenwood, J. The Future of Folk Psychology. Cambridge: Cambridge University Press, 1991.

Harris, P. "From Simulation to Folk Psychology: The Case for Development." Mind and Language 7 (Spring/Summer 1992): 120-144.

Hanson, N. R. Patterns of Discovery. Cambridge: Cambridge University Press, 1958.

Hempel, C. G., and P. Oppenheim. "Studies in the Logic of Explanation." Philosophy of Science 15 (1948): 135-75.

Hempel, C. G. "Deductive-Nomological vs. Statistical Explanation." In Minnesota Studies in Philosophy of Science 3, ed. H. Feigl and G. Maxwell, 98-169. Minneapolis: University of Minnesota Press, 1962.

_____. "Aspects of Scientific Explanation." Chap. in Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. New York: The Free Press, 1965.

_____. Philosophy of Natural Science. Englewood Cliffs, N.J. : Prentice-Hall, Inc. 1966.

Horgan, T. "Mental Quausation." Philosophical Perspectives 3. Philosophy of Mind and Action Theory (1989): 47-76.

Horgan, T. and J. Tienson. "Soft Laws." Midwest Studies in Philosophy XV (1990): 256-79.

Horgan, T. "The Explanatory Role of Content." In Dretske and His Critics, ed. Brian McLaughlin, 71-101. Oxford: Blackwell, 1991.

_____. "Nonreductive Materialism and the Explanatory Autonomy of Psychology." In Naturalism: A Critical Appraisal, ed. Steven J. Wagner and Richard Warner, 295-320. University of Notre Dame Press, 1993.

Kahneman, D. and A.Tversky. "The Simulation Heuristic." In Judgment Under Uncertainty, ed. D. Kahneman, P. Slovic and A. Tversky, 201-08. Cambridge, M.A.: Cambridge University Press, 1982.

Lewis, D. "An Argument for the Identity Theory." Journal of Philosophy 63 (1966): 17-25.

_____. "How to Define Theoretical Terms." Journal of Philosophy 67 (1970): 427-446.

_____. "Psychophysical and Theoretical Identification." Australasian Journal of Philosophy 50 (1972): 249-58.

_____. "A Subjectivist's Guide to Objective Chance." Chap. in Philosophical Papers 2. Oxford: Oxford University Press, 1986.

Morton, A. Frames of Mind. Oxford: Clarendon Press, 1980.

Nisbett, R. and T. DeCamp Wilson. "Telling More Than We Can Know: Verbal Reports on Mental Processes." Psychological Review 84 (1977): 231-59.

Putnam, H. "The Nature of Mental States." In Art, Mind and Religion, ed. W.H. Capitan and D.D. Merrill, 37-48. Pittsburg: University of Pittsburg Press, 1967.

Ruben, D-H. "A Counterfactual Theory of Causal Explanation." Nous XXVIII (December 1994): 465-81.

Railton, P. "A Deductive-Nomological Model of Probabilistic Explanation." Philosophy of Science 45 (1978): 206-26.

_____. "Probability, Explanation, and Information." In Explanation, ed. David-Hillel Ruben, 161-181. Oxford: Oxford University Press,1993. First published in Synthese 48 (1981): 233- 56.

Rescher, N. Scientific Explanation. New York: Free Press, 1970.

Salmon, W. C. Four Decades of Scientific Explanation. Minneapolis: University of Minnesota Press, 1989.

Schiffer, S. "Truth and the Theory of Content." In Meaning and Understanding, ed. Parret and Bouveresse, 205-25. Berlin: New York: de Gruyter, 1981.

_____. Remnants of Meaning. Cambridge: MIT Press, 1987.

_____. "Physicalism." Philosophical Perspectives 4, Action Theory and Philosophy of Mind (1990): 153-85.

_____. "Ceteris Paribus Laws." Mind 100 (January 1991): 1-17.

Scriven, M. "Definitions, Explanations and Theories." Minnesota Studies in Philosophy of Science II, ed. H. Feigl, M. Scriven and G. Maxwell, 99-195. Minneapolis: University of Minnesota Press, 1958.

Sellars, W. "Empiricism and the Philosophy of Mind." Minnesota Studies in the Philosophy of Science I, ed. H. Feigl and M. Scriven.

Minneapolis: University of Minnesota Press, 1956. Reprinted in W. Sellars, Science, Perception and Reality.London: Routledge and Kegan Paul, 1963.

Sorenson, R. "Self-Strengthening Empathy: How Evolution Funnels us into a Solution to the Other Minds Problem". Unpublished.

Stich, S. "Beliefs and Subdoxastic States." in Philosophy of Science 45, (1978): 499-518.

_____. From Folk Psychology to Cognitive Science. Cambridge: MIT Press, 1985.

Stich, S. and S. Nichols. "Folk Psychology: Simulation or Tacit Theory?" Mind and Language 7 (Spring/Summer 1992): 35-71.

Stich, S. and S. Nichols. "Second Thoughts on Simulation." In Mental Simulation: Philosophical and Psychological Essays, ed. Martin Davies and Tony Stone, Oxford: Basil Blackwell, 1994.

Stich, S. and I.Ravenscroft. "What *is* folk psychology?" Cognition 50 (1994): 447-68.

Stich, S., S. Nichols, A. Leslie, and D. Klein. "Varieties of Off-Line Simulation." To appear in Theories of Theories of Mind, ed. P. Carruthers and P. Smith. Cambridge: Cambridge University Press.

Von Wright, G. H. Explanation and Understanding. Ithaca: Cornell University Press, 1971.

Wimmer, H. and J. Perner. "Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception." Cognition 13 (1983): 103-28.

Wimmer, H., J. Hogrefe, and B. Sodian. "A Second Stage in Children's Conception of Mental Life: Understanding Informational Accesses as Origins of Knowledge and Belief." In Developing Theories of Mind, ed. J. W. Astington, P. L. Harris, and D. R. Olson, 173-92. New York: Cambridge University Press, 1988.