THE JOURNAL OF PHILOSOPHY

VOLUME CIII, NO. 9, SEPTEMBER 2006

CONCEPTUAL ANALYSIS NATURALIZED: A METAPHILOSOPHICAL CASE STUDY*

Once upon a time, possibly for a brief moment in interwar Vienna or postwar Oxford, but more likely in a fairy-tale kingdom, there was a clear understanding of the goal of philosophy. The subject matter of philosophy was a cluster of concepts, typically concepts that are presupposed in science or in other aspects of human endeavor; such concepts include knowledge, causation, freedom, personal identity, goodness, existence, and so on. The goal was to analyze these concepts—to give necessary and sufficient conditions for their application—in terms of concepts that were deemed less problematic (such as empirical regularities). The phrase 'conceptual analysis' is often used to denote this very specific enterprise. But it would not be unreasonable to broaden the definition of 'analysis' to include the kind of careful scrutiny to which philosophers continue to subject concepts like knowledge, freedom, and so on. Conceptual analysis in this broader sense still comprises much of philosophy today. But under this broadened definition of 'analysis', conceptual analysis is no longer the exclusive provenance of philosophy. Empirical psychology attempts to explore the nature of our concepts in a systematic way; indeed, it is hard to imagine how the investigation into the nature of our concepts could fail to have an empirical dimension. What is the relationship between these two very different modes of inquiry? Can the results of empirical psychology inform the philosophical project of conceptual analysis? I attempt to

* For comments and suggestions I would like to thank Alex Byrne, David Chalmers, Richard Corry, Clark Glymour, Alison Gopnik, Frank Jackson, Joshua Knobe, Sarah-Jane Leslie, Sarah McGrath, Ruth Millikan, Shaun Nichols, Dan Osherson, Hugh Price, Kim Sterelny, Michael Weisberg, Jim Woodward, and audience members at Australian National University, Florida State University, Princeton University, and the University of Sydney.

0022-362X/06/0309/427–51

explore these issues using my favorite philosophical concept, causation. Of course I will be happy if my explorations make a positive contribution to the understanding of causation, but my main purpose is to treat the concept of causation as a metaphilosophical case study.

In the past few years, there has been a small but growing movement in empirical philosophy. Much philosophical theorizing is built upon our untutored "intuitions" about various hypothetical situations. All too often, these intuitions are simply those of one person, or perhaps a handful of people polled informally. So one natural line of research is to determine whether people do in fact have the sorts of intuitions that are taken to support or condemn various philosophical analyses. Even if we are not willing to accept the intuitions of experimental subjects as probative in philosophical theorizing, psychology might teach us something about the processes whereby intuitions are generated. It would be interesting to know, for example, whether there are strong *framing effects*—whether our intuitions about a case depend upon just how it is presented to us. Or perhaps there are cultural differences in making certain kinds of judgments. For example, E. Machery and collaborators recently found that many of the intuitions that are taken to support direct theories of reference were widely shared among American respondents, but rejected by those from Hong Kong.[1] This gives rise to the possibility that our notion of reference is culturally specific.[2] This use of empirical methods to provide a reality check on the intuitions used to defend philosophical analyses is certainly to be lauded. While I suspect that such an empirical test of the intuitions relied upon in the causation literature might yield some surprising results, this will not be my focus in the present paper. Rather, I am searching for a deeper connection.

There are many different ways in which philosophers try to analyze concepts. As noted above, the canonical approach is to offer necessary and sufficient conditions for the application of a concept in terms of other concepts deemed to be more basic. I will instead focus on a "theoretical role" approach to the analysis of concepts. This approach aims to analyze concepts by articulating their role in facilitating inferences of various kinds. I will argue that the results of empirical

[1] Machery, R. Mallon, S. Nichols, and S. Stich, "Semantics, Cross-cultural Style," *Cognition*, XCII (2004): B1–B12.

[2] M.W. Morris and colleagues suggest that there may be cultural differences in notions of causation, although their discussion does not take place at the level of intuitions about specific cases: Morris, R.E. Nisbett, and K. Peng, "Causal Attribution Across Domains and Cultures," in Dan Sperber, David Premack, and Ann James Premack, eds., *Causal Cognition: A Multi-Disciplinary Debate* (New York: Oxford, 1995), pp. 577–612.

psychology can enrich our understanding of the theoretical role of the concept of causation. By the same token, an adequate philosophical account of the theoretical role of the concept of causation can equip us to better interpret a variety of psychological findings.

My program is as follows: In section I, I will provide a brief sketch of some of the traditional philosophical approaches to understanding causation. I will not expound upon the theories in any detail, but just rehearse their basic elements. Each account naturally suggests certain views about the epistemology of causation, and about the uses to which causal knowledge is put; I will draw particular attention to these. In section II, I will review the theoretical role approach to analyzing concepts, and offer the beginnings of a theoretical role analysis of causation. In section III, I will outline some recent trends in the empirical study of causal reasoning. In the final sections, I will argue that the theoretical role analysis and the empirical results serve to illuminate one another.

I. THEORIES OF CAUSATION

1.1. Regularity Theories. According to regularity theories of causation, causal relations are underwritten by empirical regularities. In the simplest formulation, associated with David Hume, an event *c* is a cause of a later event *e* just in case events similar to *c* are always followed by events similar to *e*. This simple formulation is universally recognized to be inadequate for a number of reasons. For one thing, there are infinitely many ways in which one event might be similar to, or different from, another. Moreover, many causes fail to be invariably followed by their effects; smoking causes lung cancer, but not all smokers develop lung cancer. These and other difficulties have led regularity theorists to posit more sophisticated patterns of co-occurrence as the foundation of causal relations. According to John Mackie,[3] for example, a property or event-type *C* is said to be an *inus* condition, a kind of minimal cause, for *E* when there is a cluster of properties or types *ABDF*...which sometimes fails to be conjoined with *E*, but is invariably followed by *E* when conjoined with *C*. I also intend for probabilistic theories such as those put forward by Patrick Suppes,[4] Nancy Cartwright,[5] and Ellery Eells[6] to count as regularity theories in a broad sense. These theories do not require that *C* be invariably followed by *E*, even when a complete constellation of ad-

[3] Mackie, *The Cement of the Universe* (New York: Oxford, 1974).

[4] Suppes, *A Probabilistic Theory of Causality* (Amsterdam: North-Holland, 1970).

[5] Cartwright, "Causal Laws and Effective Strategies," *Noûs*, XIII (1979): 419–37.

[6] Eells, *Probabilistic Causality* (New York: Cambridge, 1991).

ditional factors is held fixed: they only require that *E* occur more frequently, or that it be more probable, in the presence of *C*.

Regularity theories suggest that we learn about causal relationships by observing regular patterns of coinstantiation among properties or types of events. This observation may take many forms. It may be completely unconscious, as in the case of classical conditioning in animals. Pavlov's dogs, famously, learned to associate the ringing of a bell with the immanent provision of food. The observation may be conscious and informal, as when we note that a colleague is usually in an irritable mood after meeting with the dean. A number of psychologists have proposed that we learn about causal relationships from discrepancies in the conditional frequencies of observed outcomes.[7] Or we may keep very precise records of our observations and examine the patterns using various statistical methods such as regression and analysis of variance.

Knowledge of regularities is particularly useful in the context of prediction. Knowing that certain antecedents are invariably or frequently attended by particular effects, we can, upon observation of those antecedents, predict the occurrence of those effects.

I.2. Process/Mechanical Theories. Process theories focus on the processes that transmit causal influence across space and time, as well as on the interactions between them. Causal processes, such as rocks, electrons, photons, and sound waves, must be distinguished from mere pseudo-processes such as shadows and spots of light. The most prominent proposal for marking this distinction cites the ability of causal processes to transmit momentum, energy, and other conserved quantities.[8] A pulse of light fired from a laser will carry energy from its source to a point on a wall that is temporarily illuminated. If the source is rotated so that the spot of light moves along the wall, energy will be present at the various points along the wall occupied by the spot. Nonetheless, the spot of light is not *transmitting* the energy, which is instead being freshly supplied by the light source. Process theories are sometimes called *mechanical* theories of causation. They emphasize action by contact and the production of motion, and more generally the importance of the physical mechanism whereby the cause exerts its influence over the effect. I will include in this broad category of causal theory other approaches that are "mechanistic" in the sense that they lay emphasis on the processes and mechanisms that

[7] See, for example, P. Cheng and L. Novick, "A Probabilistic Contrast Model of Causal Induction," *Journal of Personality and Social Psychology*, LVIII (1990): 545–67.

[8] See especially Wesley Salmon, "Causality without Counterfactuals," *Philosophy of Science*, LXI (1994): 297–312; and Phil Dowe, *Physical Causation* (New York: Cambridge, 2000).

connect cause and effect, rather than on the patterns of correlation or dependence between the cause and effect.[9]

We can acquire knowledge of processes and mechanisms through observation, which can be enhanced by all manner of detection devices (microscopes, Geiger counters, cloud chambers, and so on). Distinguishing causal processes from pseudo-processes requires the application of physics, either naïve or sophisticated. A certain amount of naïve physics may be innate; the famous "launching" experiments of A. Michotte[10] seem to suggest that we directly perceive physically possible collisions to be different from perceptually similar stimuli. Using a dishabituation technique to measure changes in attention, A. Leslie and S. Keeble[11] showed that even infants discriminate between the exchange of motion produced in collisions and other random sequences of moving objects.

Knowledge of causal processes and interactions is particularly useful for understanding the workings of complex systems, such as machines comprised of gears, levers, and pulleys that interact with one another. By way of illustration, here is a common sort of experience that happened to me recently. My toilet would not stop refilling, so I opened up the tank. I had no particular knowledge of the workings of toilets or of plumbing more generally; nonetheless I was able to observe the handle, the lever, the chain, and the ball, and fathom the workings of the device. It turned out that the chain had become tangled, so that the ball could not be lowered far enough to stop completely the hole through which water flows from the tank into the bowl.

I.3. Manipulability Theories. Manipulability theories emphasize that causal relations (in contrast to spurious correlations, for example) can be used by agents to manipulate and control their environment. Causes are "handles" that can be "wiggled" in order to effect changes elsewhere in the world. Critics have charged that manipulation is itself a causal notion, and hence cannot be used to analyze causation. But this does not rule out the possibility of interesting interdefinitions of causation in terms of manipulation, and manipulation in terms of causation, and these interdefinitions may impose a set of highly nontrivial constraints on the possible structures of causal systems.[12]

[9] See, for example, the discussions of the notion of mechanism in S. Glennan, "Mechanisms and the Nature of Causation," *Erkenntnis*, XLIV (1996): 49–71, and P. Machamer, L. Darden, and C. Craver, "Thinking about Mechanisms," *Philosophy of Science*, LXVII (2000): 1–25.

[10] Michotte, *The Perception of Causality* (New York: Basic Books, 1963).

[11] Leslie and Keeble, "Do Six-month Old Infants Perceive Causality?" *Cognition*, XXV (1987): 265–88.

[12] The most detailed account is James Woodward's *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford, 2003).

Manipulability theories suggest that we learn about cause and effect through our own interventions on the world. Again, this learning can come at many different levels. At the most basic level is operant conditioning. When an action is regularly followed by a reward or punishment, we adjust our behavior (often unconsciously) so as to perform that action more or less frequently. We also engage in more conscious trial and error learning. At the most sophisticated level, we perform controlled experiments, such as double blind clinical trials. When we wish to test whether a new drug is effective at reducing hypertension, for example, we do not simply allow subjects to choose for themselves whether to take the drug, and then observe the results. Rather, we intervene to determine who takes the drug and who does not.

Manipulability theories focus on our role as agents in the world. Causal knowledge allows us to intervene effectively in the world in order to achieve our ends. This is true at the level of individual decision-making, as well as at the level of national policy.

I.4. Counterfactual Theories. Counterfactual theories attempt to analyze causation in terms of relations of counterfactual dependence. An event *e* counterfactually depends upon event *c* just in case *e* would not have occurred if *c* had not occurred.[13] The simplest counterfactual theory would identify causation with counterfactual dependence. This theory will not work because of problems involving cases of *preemption*, but we need not detain ourselves with the epicycles.[14]

At the most immediate level, we learn about what would have happened under various counterfactual assumptions through a process of imagination or mental simulation similar to that involved in thought experiments. Simulation can also be facilitated by the use of mathematical models or computers. Ultimately, of course, our ability to run such simulations must stem from an understanding of the workings of the world gained from previous experience.[15] Our knowledge about counterfactuals will thus also depend upon the kind of knowledge gained from observation and experimentation described above.

Counterfactual reasoning is particularly apt for retrospective evaluations of responsibility. If one wants to know the extent to which a certain agent's action contributed to an outcome, compare that outcome with what would have happened had the agent behaved differently. This "but for" or "sine qua non" test is often used as a criterion for causal responsibility in the law.

[13] I have omitted an explicit statement of the standard qualifiers: *c* and *e* both occur, and they are distinct from one another.

[14] For the classic statement, see David Lewis's "Causation," this JOURNAL, LXX, 17 (October 11, 1973): 556–67.

[15] Ernst Mach famously made this point about thought experiments in *The Science of Mechanics* (London: Open Court, 1883).

II. THEORETICAL ROLE ANALYSES

The sort of approach to the analysis of concepts that I will focus on traces back to the logical empiricists. One classic statement is found in Carl Hempel's paper "The Theoretician's Dilemma."[16] Hempel is concerned with the status of theoretical concepts in scientific theories (although he formulates the problem in terms of theoretical *terms*). He argues that if theoretical concepts can be defined in purely observational terms, then the use of theoretical concepts in science is in principle unnecessary (although theoretical vocabulary may still be useful as a kind of shorthand). However, if theoretical concepts can only be partially interpreted in observational terms (by means of Rudolf Carnap's *reduction sentences*[17]) but this can be done in multiple ways, then theoretical concepts can play a role in what Hempel calls *inductive systematization*. For example (Hempel's),[18] suppose that our theoretical vocabulary contains the term 'white phosphorus', and observational terms 'smells like garlic', 'dissolves in turpentine', and 'ignites at a temperature of 30° C'.[19] 'White phosphorus' is not defined, but is given a partial interpretation:

> White phosphorus smells like garlic.
>
> White phosphorus will dissolve if placed in turpentine.

Our theory will include as a postulate:

> White phosphorus ignites at a temperature of 30° C.

Now if we find that a particular substance smells like garlic, and dissolves in turpentine, then we have inductive evidence that it will ignite at a temperature of 30° C. This inductive relationship is something that we could not capture in a purely observational language. The concept 'white phosphorus' thus plays a certain role in our theorizing, and an understanding of this role provides us with an understanding of the concept. Note that while I talk about the *theoretical* role of a concept, I do not intend to deny that the concept will also be useful in facilitating certain practical judgments. The concept 'white

[16] Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 173–226.

[17] See, for example, Carnap, "Testability and Meaning," *Philosophy of Science*, III (1936): 420–68; IV (1937): 1–40.

[18] Hempel, "The Theoretician's Dilemma," in *Aspects of Scientific Explanation*, pp. 214–15.

[19] Unlike Hempel, I here ignore the problem that temperature is also a theoretical concept in need of interpretation.

phosphorus' might well play a role in dissuading us from performing a certain procedure requiring temperatures in excess of 30° C.

For another illustration, consider the account of objective chance offered by David Lewis's "Principal Principle."[20] This principle says that:

$$Cr(A|E \;\&\; Ch(A) = x) = x,$$

where *Cr* is the credence function, or subjective probability, of a rational agent; *A* is an arbitrary proposition; *E* is an "admissible" proposition; and *Ch* is an objective chance function. For simplicity, I have omitted time indices, and I will not pursue the issue of what makes a proposition admissible. In words, the Principal Principle says that a rational agent's degree of belief in *A*, given that the objective chance of *A* is *x*, is equal to *x* (assuming that the agent is not also conditionalizing upon some inadmissible proposition). This principle does not tell us what kind of thing chance is in the world, but it does tell us a great deal about inferences involving chances. By plugging the principle into a Bayesian framework, it shows us how various pieces of evidence can bear upon hypotheses about the values of chances. Moreover, using some version of expected utility theory, the principle can tell us how information about objective chances should affect our choices. Thus, for example, the Principal Principle can mediate an inference from observed frequencies to the conclusion that one action is to be preferred to another.

My suggestion is that we might try to provide an analogous account of the theoretical role played by the concept of causation, and thereby gain a deeper understanding of that concept. How would such an account go? We may think of the different approaches to causation sketched in section I as identifying different causal *modalities*. (The word 'modalities' here has nothing to do with *alethic* modalities—necessity, possibility, and so on—although these are also obviously connected with causation.) Attempts to analyze causation in terms of these modalities will then be reinterpreted as Carnapian reduction sentences. The role of the concept of causation is to mediate intermodal inferences. To give an oversimplified example in the model of Hempel, suppose that we have the following 2 principles:

Reg: If *C* causes *E* in background circumstances *B*, then in those circumstances *E* will occur more frequently when *C* is present than when it is absent.

[20] Lewis, "A Subjectivist's Guide to Objective Chance," in Richard Jeffrey, ed., *Studies in Inductive Logic and Probability, Volume II* (Berkeley: California UP, 1980), pp. 263–94; for a similar account, see also Brian Skyrms, *Causal Necessity* (New Haven: Yale, 1980).

Man: If, in circumstances *B*, intervening to produce *C* is a reliable means for producing *E*, then *C* causes *E* in those circumstances.

Neither is an analysis: Reg states only a necessary condition for *C*'s causing *E*; and Man states only a sufficient condition. We know that the converse of Reg can fail in cases of spurious correlations, and we know that the converse of Man can fail when practical limitations make it impossible for us to intervene so as to produce *C*. Thus there is no logically valid inference from

In circumstances *B*, *E* occurs more frequently when *C* is present than when it is not

to

In circumstances *B*, intervening to produce *C* is a reliable means for producing *E*.

Nonetheless the former may still provide inductive evidence for the latter, and the ability to draw such inferences may be extremely useful. Note, moreover, that Reg and Man can serve this purpose even if they are not universally true, that is, even if they do not state genuinely necessary and sufficient conditions, respectively.

For a more sophisticated example, consider the use of Bayes nets to represent causal relationships.[21] I will not attempt a detailed introduction to Bayes nets for the uninitiated, but the basic idea is that one represents a causal system using a graph with arrows connecting variables. The arrows represent relationships of "direct" causation. These relationships are not further defined (except in informal exposition). Rather, there are a set of principles describing the consequences of a causal graph for probabilities, interventions, and counterfactuals. (The connection with causal processes or mechanisms is much less clear.) This permits various kinds of inferences to be drawn: for example, one may be able to infer whether or not a counterfactual is true, or the probability that it is true, on the basis of information about probabilistic correlations and the results of various manipulations. Thus Bayes nets can be used to mediate inferences between these different modalities.

Note that my proposal would still leave a role to play for traditional philosophical analyses. A new and improved regularity theory of causation, for example, would provide a better account of the sorts of

[21] See, for example, Judea Pearl, *Causality: Models, Reasoning, and Inference* (New York: Cambridge, 2000); and Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, Prediction, and Search* (Cambridge: MIT, 2000, second edition).

inferences that could be made between regularities and causal relationships. But traditional philosophical analysis would no longer occupy a privileged position: developments in statistics and causal modeling could also inform us about inferential relationships between causal relationships and regularities. The principal difference is that philosophical analyses have always aimed at establishing *deductive* inferential connections.

So far, I can only claim to have given a skeleton of an account. My goal in the rest of the paper will not be so much to provide the flesh, as to argue that empirical psychology provides a useful resource for adding flesh to the bones.

III. TWO TRENDS IN PSYCHOLOGY

Like philosophy, psychology has only recently (in the past twenty-five years or so) begun to emerge from a positivistic influence that made causation a taboo subject.[22] Nonetheless, the psychological literature on causal reasoning in humans and other animals is already vast, and I cannot hope to provide a systematic overview of even the little that I do know in a paper such as this. What I will attempt, instead, is to sketch two diverging trends that merit attention. On the one hand, recent developments in developmental psychology indicate that children develop a highly sophisticated capacity for causal reasoning by about the age of four, many pieces of which are already present at a much younger age. By contrast, recent primate studies strongly suggest that even our closest relatives, chimpanzees and other nonhuman primates, lack a wide variety of capacities that appear to require some form of causal understanding. This contrast between young human children and nonhuman primates is particularly interesting since it makes it less likely that the differences in the causal reasoning abilities of humans and other primates is just due to differences in general intelligence. It seems that our capacity for causal reasoning rests neither in our mature scientific understanding, nor in our animal origins. It would take us too far afield to address the issue of whether our capacity for causal reasoning is *innate*, but it does seem to be a basic feature of our uniquely human endowment.

III.1. Causal Reasoning in Primates. Some of the most interesting work to come out of the recent primatology literature has focused on

[22] The situation is of course more complicated than this. There has been a strong tradition in psychology, epitomized by the work of Michotte (see note 10), resisting the positivist influence. As noted above, this tradition stresses the apparent observability of causal relations, particularly the causation of motion in collisions between objects. I would argue, however, that this work has nonetheless suffered indirectly from the influence of positivism, particularly from a lack of sympathetic criticism.

two topics: tool use and social cognition. This is hardly surprising: along with language use, tool use and the complexity of our social arrangements are among the most distinctive features of human culture. Let us begin with tool use. Most of us are familiar with the observations of Jane Goodall[23] and others of chimpanzees stripping the leaves off pliant twigs and inserting them into the channels of termite mounds in order to draw out termites. It is very tempting to attribute to the chimpanzees a great deal of causal understanding on the basis of this behavior. Surely the chimpanzees are selecting pliant twigs so that they can follow the contours of the termite channels as they are inserted; surely the animals are deliberately modifying the shape of the twigs so that they can be inserted easily into the holes. A recent set of very careful experiments conducted by Daniel Povinelli and his collaborators[24] has cast doubt on this interpretation. For example, in one experiment, seven chimpanzees were trained to insert a straight stick through a hole in order to dislodge a piece of fruit that was otherwise inaccessible. The chimpanzees were then presented with a novel tool: a stick that had a cross-piece at one end. The cross-piece was designed to prevent that end of the stick from passing through the hole, and was large enough that its inability to pass through the hole was obvious to humans upon quick inspection. The apes would nonetheless attempt to insert the end of the stick with the cross-piece through the hole, rather than employing the narrow end of the tool.[25] In a related experiment, the chimpanzees were presented a stick containing a removable cross-piece in each end. The animals had previously been given an opportunity to play with these tools, and had learned how to remove and insert the cross-pieces. Instead of simply removing one or both cross-pieces and inserting the modified stick through the hole, the animals would engage in various types of ineffective behavior. In some cases, the animals would remove a cross-piece and then insert it again before trying to dislodge the fruit; in others, they would remove the cross-piece from one end and then attempt to insert the other end through the hole.[26] When the apes did succeed in dislodging the fruit, this success appeared to result from random trial and error rather than deliberate planning.

[23] Goodall, "The Behavior of Free-ranging Chimpanzees in the Gombe Stream Reserve," *Animal Behavior Monographs*, I (1968): 161–311.

[24] Povinelli, with individual chapters co-authored with J.E. Reaux, L.A. Theall, and S. Giambrone, *Folk Physics for Apes: The Chimpanzee's Theory of How the World Works* (New York: Oxford, 2000).

[25] This is, in fact, a rather perfunctory anecdote summarizing the results of three different experiments. See chapter 8 of *Folk Physics for Apes* for details.

[26] See chapter 11 of *Folk Physics for Apes* for a more detailed presentation.

On these and related tasks, a common pattern emerged. When presented with a novel problem, the subjects would gradually learn how to solve it by trial and error, and then internalize a specific procedural rule. When presented with a variant of the original problem, they would initiate a previously learned procedure based upon superficial perceptual characteristics (such as the straightness of the stick) rather than upon the relevant structural features of the problem (the size of the hole in comparison to the end of the stick). This strategy might serve well enough in a natural environment where the superficial perceptual features of a problem are well correlated with the relevant structural features. It is, in fact, quite difficult to distinguish the sort of deliberate behavior that belies an understanding of the causal parameters of a problem from routines learned through trial and error; hence the need for carefully controlled experiments. The limitations of the chimps' strategies were manifest only in a laboratory setting where the perceptual and structural features of a problem could be independently varied.

Another set of results bears upon the much vaunted imitation skills of apes and other primates. Primates do mimic the bodily movements of others, but they seem to lack more sophisticated imitative skills. For example, J. Call and M. Tomasello[27] conducted an experiment in which a stick could be manipulated in a certain way to dislodge a piece of food. The food, and the end of the stick that made contact with it, could not be seen. Orang-utans observed while human demonstrators manipulated the stick to obtain food. When the apes were given their turn, they would manipulate the stick at random, discovering the solution (if at all) by trial and error. The Orang-utans did no better when they observed a conspecific who had mastered the task, rather than a human demonstrator. Whiten and collaborators[28] presented chimpanzees with food in a transparent container that could be opened in one of two ways. One group of chimpanzees was allowed to observe an experimenter open the container in one manner, while a second group observed the container being opened in the other manner. Again, when the chimps had their go, they attacked the box by trial and error. Those in the first group were no more likely to employ the first method of opening the container than those in the second. In each case, the apes seemed to learn an

[27] Call and Tomasello, "The Use of Social Information in the Problem-solving of Orang-utans and Human Children," *Journal of Comparative Psychology*, CIX (1995): 308–20.

[28] A. Whiten, D.M. Custance, J.C. Gómez, P. Teixidor, and K.A. Bard, "Imitative Learning of Artificial Fruit Processing in Children and Chimpanzees," *Journal of Comparative Psychology*, CX (1996): 3–14.

affordance of the environment, that food could be obtained in a certain place, and certain objects were brought to their attention, such as the stick in the experiment of Tomasello and Call, but the apes were not able to learn the appropriate means-end relationships by observation. This may be due in part to a deficit in social cognition: the apes could not appreciate that the demonstrators were deliberately employing a known method to obtain food (and similarly in the case where the apes were allowed to observe conspecifics).

We noted above that one context in which we apply causal reasoning is in the assessment of responsibility. It would be interesting to know whether chimpanzees and other primates engage in an analog of responsibility ascribing behavior. I know of no studies addressing this issue in chimpanzees, but there is some evidence that other primates adhere to simple heuristics in situations that would seem to call for responsibility ascription. For example, dominant long-tailed macaques will respond to the cries of a juvenile by indiscriminately attacking subordinates that are closest to the juvenile.[29]

III.2. Causal Reasoning in Human Children. In striking contrast to the results described above, even very young children perform very differently on similar tests. Ann Brown[30] conducted an experiment in which children were given a choice of tools for reaching a toy. The children were allowed to attempt the task as often as necessary to master the skill, receiving help in the form of verbal instructions and pantomime if necessary. The same children were then given the same task with a new set of tools. Even children as young as two years old could select the tool appropriate to the task, in preference to tools superficially resembling (in color, shape, and/or size) the tool originally used to solve the problem. These children, unlike Povinelli's apes, were able to attend to the relevant structural features (such as length and rigidity) of the tools used to solve the original problem, and not merely to perceptual characteristics (such as color and shape). In a similar vein, young children perform very well at imitation tasks. For example, most three and four year olds were able to imitate the actions of a human model in the imitation test of Call and Tomasello described above (*op. cit.*).

Experiments performed by Harris and others[31] show that children are adept at counterfactual reasoning. Children were presented with

[29] H. Kummer and M. Cords, "Cues of Ownership in Long-Tailed Macaques," *Animal Behavior*, XLII (1991): 357–62.

[30] Brown, "Domain-specific Principles Affect Learning and Transfer in Children," *Cognitive Science*, XIV (1990): 107–33.

[31] P.L. Harris, T. German, and P. Mills, "Children's Use of Counterfactual Thinking in Causal Reasoning," *Cognition*, LXI (1996): 233–59.

scenarios acted out with toys. For example, in one scenario, a person walks across a clean surface, getting it dirty. Children are able to reason that the floor would have remained clean if the person's shoes had been removed, and cite the failure to remove the shoes when asked to explain why the floor is dirty.

Perhaps the most impressive evidence of children's prowess at causal reasoning comes from a series of experiments performed by Alison Gopnik and her collaborators.[32] In one set of experiments, children are exposed to a device called a "blicket detector," and told that blickets "make it go." The children are then shown an object that is identified as a blicket, which is placed on the machine, whereupon the machine plays music and flashes lights. (The machine is in fact operated remotely by the experimenter.) Various objects are put on the machine, either singly or in combination, and children are asked to identify which objects are blickets. Children are able to disregard superficial characteristics of the objects such color and shape; rather, they attend to complex patterns of covariation between the presence of the objects and the operation of the machine in determining which objects are blickets. More specifically, they rely on information about conditional independence relations: if the machine goes off no more often when both objects *A* and *B* are placed on it, than when *A* alone is placed on it, they infer that *B* is not a blicket, even though the presence of *B* is frequently accompanied by the activation of the machine. In one version of the experiment, objects are left on the machine and children are asked to make it stop; children respond by removing the object that had been identified as a blicket.

In a related set of experiments by Laura Schulz, Gopnik, and Clark Glymour,[33] children are shown a machine with a visible crank that turns a gear. A switch can be used to turn the machine on and off. Children are then shown a machine with a switch and two interlocking gears; the mechanism is hidden. When asked to determine which gear(s) is (are) connected to a crank, children will experiment

[32] See, for example, Gopnik and Glymour, "Causal Maps and Bayes Nets: A Cognitive and Computational Account of Theory-formation," in Peter Carruthers, Stephen Stich, and Michael Siegal, eds., *The Cognitive Basis of Science* (New York: Cambridge, 2002), pp. 117–32; Gopnik, Glymour, D.M. Sobel, L.E. Schulz, T. Kushnir, and D. Danks, "A Theory of Causal Learning in Children: Causal Maps and Bayes Nets," *Psychological Review*, CXI (2004): 3–32; Gopnik and Sobel, "Detecting Blickets: How Young Children Use Information about Novel Causal Powers in Categorization and Induction," *Child Development*, LXXI (2000): 1205–22; Gopnik, Sobel, Schulz, and Glymour, "Causal Learning Mechanisms in Very Young Children: Two-, Three-, and Four-Year-Olds Infer Causal Relations from Patterns of Variation and Covariation," *Developmental Psychology*, XXXVII (2001): 620–709.

[33] Schulz, Gopnik, and Glymour, "Preschool Children Learn about Causal Structure from Conditional Interventions," *Developmental Science*, forthcoming.

with the switch and remove the gears one at a time to arrive at the correct answer.

IV. INTERPRETING THE RESULTS

It appears that there is a whole range of capacities that humans have, even from a very early age, that even our closest relatives do not. Moreover, these capacities are clearly related to causal reasoning. My concern is not so much to *explain* these differential abilities, but to suggest that the theoretical role account of causation sketched in section II above can provide a useful framework for interpreting these results. One important dimension of our capacity for causal reasoning is our ability to draw inferences across modalities such as regularities, mechanisms, and manipulation relations—to use evidence appropriate to one modality in an application appropriate to another. If the account of section II is correct, it is precisely our possession of a full-blown concept of causation that permits us to make these inferences.

The empirical results described in section III.2 establish quite convincingly that human children *are* able to make intermodal inferences. For example, the studies by Gopnik and her collaborators show that children can make use of evidence about regularities to infer the effects of their own actions, and that they can use information about the results of their manipulations to make inferences about unobserved mechanisms. The more problematic claim is that chimpanzees and other nonhuman animals are unable, or substantially less able, to make such intermodal inferences.

First let us note that the (alleged) inability of chimpanzees to make intermodal inferences would be considerably less interesting if they were also unable to make *intra*modal inferences. The astronomical literature on classical and operant conditioning establishes beyond a doubt that mammals are capable of learning about regularities in their environment, even fairly complex ones, and that they are capable of learning about the effects of their actions on the world. To what extent primates are capable of learning about causal processes and interactions is somewhat less clear. Given the sensitivity of their actions to perceptual features of the experimental situation, it is clear that the chimpanzees in Povinelli's experiments could attend to what we might call "geometrical facts," such as the shapes and configurations of objects, whether they are in physical contact with one another, and so on. It is also clear that chimpanzees are able to track moving objects, for example. But it is not clear that this adds up to any kind of understanding of causal processes or mechanisms. What would be desirable, from my point of view, would be something akin to the results of the various launching experiments in human adults and children. These experiments show that humans are able "passively"—without the need for manipulation—to recognize causal interactions between colliding objects. To my knowledge, analogous experiments have not been done with chimpanzees (although Hans Kummer voices the opinion that they would perform similarly to humans[34]). Nonetheless, there have been a number of studies using looking-time paradigms to study primates' mechanical understanding. In one experiment,[35] Laurie Santos and collaborators habituated cotton-top tamarins and rhesus macaques to an L-shaped tool being used to push a grape down a ramp. The primates were then shown two new tools performing the same task—one with the same shape but a different color from the original tool, and a second tool with a shape that made it inappropriate for the task.[36] The primates paid more attention (as evidenced by longer looking time) when the second tool was used, suggesting that this case violated their expectations.[37] Particularly interesting from my perspective are examples of *dissociations* between the performance of primates on looking-time tests and performance based tests. For example, Marc Hauser[38] showed rhesus macaques an arrangement with two boxes, one on top of a solid table, the other beneath it. The arrangement was then hidden behind a screen, and a slice of apple was dropped. The macaques searched for the apple in the box under the table. However, when Santos and Hauser performed a similar experiment,[39] then removed the screen to reveal the location of the apple in the upper or lower box,[40] there was a significant difference in looking times. The macaques looked longer when the apple appeared in the

[34] Kummer, "Causal Knowledge in Animals," in Sperber, Premack, and Premack, eds., *Causal Cognition: A Multi-Disciplinary Debate*, pp. 26–36.

[35] Reported in Santos, C. Miller, and M. Hauser, "Representing Tools: How Two Nonhuman Primate Species Distinguish between Functionally Relevant and Irrelevant Features of a Tool," *Animal Cognition*, VI (2003): 269–81.

[36] The grape and the bottom of the tool were occluded by a screen, so the primates saw only the top of the tool moving, and the grape rolling out from behind the screen. The grape was in fact moved by sleight of hand rather than pushed by the tool.

[37] The primates were also presented with a third condition, in which the original tool was used in a different, functionally inappropriate orientation. In this condition, the tamarins exhibited longer looking times, but the macaques did not. This suggests that the macaques may have been attending to the shape of the tool, rather than to its functionally relevant properties.

[38] Hauser, "Searching for Food in the Wild: A Nonhuman Primate's Expectations about Invisible Displacement," *Developmental Science*, IV (2001): 84–93.

[39] Reported in Santos and Hauser, "A Non-human Primate's Understanding of Solidity: Dissociations between Seeing and Acting," *Developmental Science*, V (2002): F1–F7.

[40] The apple slice was in fact placed in the box through a trap door.

lower box, indicating that this result violated their expectations. This suggests that the macaques understood that the solid table would prevent the apple from passing through, even though they could not make use of this information in guiding their search for the apple.[41]

What does the empirical evidence say about the abilities of chimpanzees to make intermodal inferences?[42] In the case of the experiments conducted by Povinelli and his collaborators, it strikes me as extremely natural to interpret the results as showing that the apes are unable to make use of the information they have about the shape, configuration, solidity, and so on, of the objects in order to draw conclusions about the effects of their own actions upon them. The dissociation between looking time and searching behavior in rhesus macaques also suggest that these primates are unable to make use of mechanical knowledge to guide their action. According to the interpretive framework that I suggest, this indicates that nonhuman primates lack the concept of causation needed to mediate these inferences.

Povinelli attributes his subjects' poor performance in his tool-using experiments to their positivism. Chimpanzees' attend only to the superficial features of the cases in which their actions succeed and fail: they do not attempt to explain those successes and failures in terms of unobservable causes or concepts such as "force." I find this interpretation somewhat puzzling: *I* would not invoke unobserved causes to explain the inability of the end of the stick with a cross-piece to pass through the hole. The hole and the end of the stick are both observable, and so, presumably, is the "smaller than" relation that holds between them. It does not seem necessary to posit any unobservable entities to explain one's inability to put the wide end of the stick through the hole.[43]

[41] Children under two and a half years old exhibit a similar dissociation. Note that I do not attempt to argue that chimpanzees are capable of engaging in the kind of hypothetical reasoning required by counterfactual theories of causation. I do not know how one would test their capacity for such reasoning, although their poor performance on the tool-use tasks set by Povinelli and his collaborators would seem to suggest that the chimps are unable to simulate their own actions in order to anticipate the results. Thus the counterfactual approach to causation is something of an outlier in my picture. One problem here is that counterfactual reasoning is not really intramodal to start with: it must make use of background assumptions acquired in other ways. So the ability to engage in counterfactual reasoning may itself be a manifestation of the capacity to make intermodal inferences.

[42] The following five paragraphs have been strongly influenced by lengthy discussions with Woodward. Woodward makes a number of similar points in section 8 of his "Interventionist Theories of Causation in Psychological Perspective," in Gopnik and Schulz, eds., *Causal Learning: Psychology, Philosophy and Computation* (New York: Oxford, forthcoming).

[43] To be fair, Povinelli seems to have something richer in mind by the notion of "observation" than is standard in the philosophical literature.

Note that the interpretation I favor is slightly different from one that Povinelli considers as an alternative to his own view:

> Perhaps they [chimpanzees] understand specific causal concepts but, for one reason or another, simply can't use them? For instance, perhaps there is only a weak connection between their causal knowledge and their action systems, or they just get stuck in some previous way of responding?[44]

As Povinelli formulates it, this proposal is that the apes have full-blown causal concepts that are not manifest in their manipulations of tools (but which may perhaps be demonstrated through looking-time experiments[45]). My proposal is not that the apes possess full-blown causal information that they are somehow unable to *use*; it is rather that they are unable to translate the broadly mechanical information that they do have into predictions about the effects of their actions; and that this failure is, in part, *constitutive* of their lacking a conception of causation.

What about inferences from information about regularities to the effects of manipulations (or vice versa)? Gopnik, a master of the apt slogan, recently suggested to me that the key question about causal cognition in animals is: "Why don't Pavlov's dogs try to ring the bell?"[46] This captures the spirit of my proposal in a nutshell. Despite the enormous literature on classical conditioning and operant conditioning, there is very little empirical research devoted to intermodal inferences from regularities to manipulations. In one interesting series of experiments,[47] A. Blaisdell and collaborators exposed two groups of rats to two different patterns of association between the sound of a tone and food delivery. One pattern had a temporal order suggesting that the tone was a cause of food delivery, while the other had a temporal order suggesting that the tone and the delivery of food were effects of a common cause (a flash of light). The rats then learned that they could cause the tone by pressing on a lever. The rats that had been exposed to the two different patterns of association differed in the frequency of nose pokes in the food receptacle after pressing the lever, indicating different expectations regarding the appearance of food. Those who had observed the pattern of association suggesting that the tone was a cause of food delivery seemed

[44] Povinelli, pp. 322–23.

[45] Povinelli objects that such experiments would not be able to demonstrate the apes' ability to recognize the unobservable phenomena that underlie the causal relations involved in tool use. Again, I find the emphasis on unobservables quite puzzling in this context.

[46] Gopnik, personal communication.

[47] Reported in Blaisdell, K. Sawa, K. Leising, and M. Waldmann, "Causal Reasoning in Rats," *Science*, CCCXI (2006): 1020–22.

to exhibit a greater expectancy for food consequent upon pushing the lever. Blaisdell and colleagues interpret this result as showing that the rats are engaging in causal reasoning. Interestingly, however, the two groups of rats did not exhibit any difference in the rate of lever-pressing.[48] Unfortunately, so far as I am aware, no one has attempted parallel experiments on primates. Tomasello and Call express their confidence that chimpanzees would fare no better than rats in the following passage:

> [S]uppose that an individual ape, who has never before observed such an event, for the first time observes the wind blowing a tree such that the fruit falls to the ground. If it understands the causal relations involved, that the movement of the limb is what caused the fruit to fall, it should be able to devise other ways to make the limb move and so make the fruit fall…[W]e believe that most primatologists would be astounded to see the ape, *just on the basis of having observed the wind make fruit fall*, proceed to shake a limb, or pull an attached vine, to create the same movement of the limb…[T]he problem is that the wind is completely independent of the observing individual and so causal analysis would have to proceed without reference to the organism's own behavior and the feedback it might receive from that….[49]

Their hypothetical example involves the observation of a single instance, and not a pattern or regularity, but that seems to be an orthogonal feature; their central claim is that an ape would not be able to infer from what it had passively observed to what it could do to achieve the same end.[50] The results of the imitation tests described above would seem to confirm this diagnosis: they suggest that apes do have difficulties translating the cause and effect sequences they have observed into effective manipulations.

The interpretation that I have been advocating is similar to one offered by James Woodward.[51] Woodward espouses a manipulationist conception of causation, according to which causal relations are those patterns or correlations that persist under (idealized) interventions; such correlations can be exploited for purposes of achieving desirable ends. Spurious correlations, by contrast, are disrupted by interventions, and cannot be exploited for practical ends. Woodward distinguishes three different levels of causal understanding. An agent that exhibits full causal understanding appreciates that the same kind

[48] Thanks to Gopnik for pointing this out.

[49] Tomasello and Call, *Primate Cognition* (New York: Oxford, 1997), p. 389.

[50] Somewhat surprisingly, Povinelli claims that he and his collaborators would not be surprised to see a chimpanzee behave in the manner described (p. 314).

[51] Woodward, "Interventionist Theories of Causation in Psychological Perspective."

of relationship that she exploits when intervening in the world can exist independently of the actions of any agents. Thus evidence gained by purely passive observation—such as the recognition of correlations or mechanisms—can be used to predict the consequences of her interventions. By contrast, an agent exhibiting purely egocentric understanding recognizes correlations between his actions and various consequences, but does not comprehend that this same relationship among events can exist even when he is not acting. Woodward interprets the failure of apes to solve various causal reasoning tasks as evidence that they have only an egocentric understanding of causal relationships. Woodward suggests that between these two levels of causal understanding is a third, wherein an agent recognizes that the relationship that exists when she intervenes can also be present when other agents interact with the world. This intermediate stage suggests, intriguingly, that our social understanding may play a role in getting us beyond the egocentric stage of causal understanding. Woodward's proposal differs from mine in its emphasis on manipulability relations; on my proposal, full causal understanding requires the ability to make all kinds of intermodal inferences, not merely inferences to and from the effects of one's own actions.

I conclude this section with a cautionary note. Many of the empirical results on which I rely are very recent and/or controversial, and may be overthrown or reinterpreted in light of new evidence. Of course, one cost of adding an empirical dimension to the enterprise of conceptual analysis is that one thereby subjects one's analyses to the risk of empirical refutation. I have presented the research to accentuate the differences between humans and nonhuman primates, but it is not unlikely that the differences may be less dramatic than the experiments just described seem to indicate. There may, for example, be a variety of causal reasoning tasks, some more complex than others, some of which can be solved easily by nonhuman primates, and some which cannot. Hauser and others[52] have found that cotton-top tamarins, which do not employ tools in the wild, could solve a number of causal reasoning problems that were simpler than those facing Povinelli's chimps.[53] For example, when faced with a choice between pulling a blanket that had a food item resting on it, and one with a food item next to it, the tamarins would pull on the former. Or, given a choice between pulling on a blanket that was

[52] Hauser, J. Kralik, and C. Botto-Mahan, "Problem Solving and Functional Design Features: Experiments on Cotton-top Tamarins, *Saguinus Oedipus Oedipus*," *Animal Behaviour*, LVII (1999): 565–82.

[53] Thanks to Knobe and Nichols for bringing these studies to my attention.

connected to a piece of blanket that had a food item on it, and pulling on a blanket that was not connected to the one with the food, the tamarins would pull the former. These results suggest that the tamarins have some conception of the principle of action by contact. This is consistent with the results of Povinelli's experiments; for example, the chimps in the experiment requiring the dislodgment of the apple all attempted to use the novel tool to make contact with the piece of fruit on the shelf. Another possibility is that there may be a variety of tasks where apes and other primates are successful at a rate that is significantly greater than chance, but which nonetheless falls short of the sort of reliability with which humans can succeed at the same tasks. It may also be that there are important developmental differences between humans and other primates. Povinelli's experiments were performed on adolescent chimps, while the field studies of T. Matsuzawa[54] suggest that only full adult chimpanzees use tools in the wild. By contrast, human children succeed in causal reasoning tasks involving tool use at a much earlier developmental stage.

It is clear that there remains empirical work to be done. A conceptual analysis that motivates interesting avenues of empirical investigation is worthwhile for that reason alone. Even if there is no clean all-or-nothing break between humans and the great apes, the sort of theoretical role account of causation that I have sketched could still provide a useful interpretive framework for understanding the cognitive capacities of different animals. However, since my goal is to argue that empirical results can inform the project of conceptual analysis, I will assume for the sake of argument that further empirical work does indeed support a fairly striking difference in the capacities of humans and other primates to make intermodal inferences, and draw a number of philosophical conclusions.

V. CONCEPTUAL ANALYSIS NATURALIZED

My central thesis is that the sorts of empirical results described in section III can substantially enrich the sort of theoretical role analysis that I have been considering. I do not mean to suggest that one could not formulate such an analysis in the absence of empirical data. Indeed, the discussion of section II suggests that extant accounts of causation can provide ample material for such a theoretical role analysis. My paradigm example of a theoretical role analysis, the Principal Principle, is purely normative in character: it aims to describe

[54] Matzuwa, "Chimpanzee Intelligence in Nature and in Captivity: Isomorphism of Symbol Use and Tool Use," in W.C. McGrew, L.F. Marchant, and T. Nishida, eds., *Great Ape Societies* (New York: Cambridge, 1996), pp. 196–209.

how an ideally rational agent would make inferences to and from claims about chances, with no suggestion that actual agents make such inferences. Nor am I suggesting that psychology *supplant* philosophy, that we should let psychology determine what concept of causation we actually deploy and leave it at that. For one thing, we should certainly allow that psychology may reveal ordinary human causal reasoning to be flawed in various respects, but that is only possible if there exist independent means for developing *norms* of causal reasoning. Some psychologists have attempted to develop such norms, but to a philosophical eye they are often very unsophisticated.[55] Moreover, while experimental psychology is well suited to studying the often unconscious causal reasoning used in everyday situations, causal reasoning also occurs in the sciences. In this domain, investigators consciously use specific techniques—experimental methods, statistical analyses, and so on—that will at least sometimes override their common-sense judgments. The examination of these techniques falls within the domain of philosophy of science, not psychology.

Nonetheless, there are many important ways in which psychological investigation can contribute to the philosophical project of analyzing a concept such as causation.

V.1. Theoretical Fruitfulness. It is a desideratum of conceptual analyses that they be theoretically fruitful; that they not merely capture the actual or idealized application of the concept, but also illuminate further theoretical endeavors. Thus, it is to be hoped that adequate accounts of freedom and personal identity will shed light on problems in moral philosophy. It is to be hoped that an adequate analysis of the concept of fitness will shed light on issues in evolutionary biology. And so on. I have argued in the previous section that a theoretical role account of causation can shed light on issues in empirical psychology.

V.2. Psychological Realism. Most theories of causation provide criteria (even if not in the form of necessary and sufficient conditions) for judging whether one thing is a cause of another. Psychology can tell us whether the actual procedures we use to judge causation are anything like those proposed by the theory. Such psychological realism is not a *sine qua non* for a successful philosophical analysis; no such analysis purports to be a literal account of how we actually make

[55] See, for example, Glymour's critique of A.G. Baker, P. Mercier, F. Vallée-Tourangeau, R. Frank, and M. Pan, "Selective Associations and Causality Judgments," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, XIX (1993): 414–32, in chapters 4 and 5 of *The Mind's Arrows* (Cambridge: MIT, 2001).

judgments about causation. Nonetheless, some measure of psychological realism is surely a *desideratum* of analysis—perhaps one that will need to be sacrificed in favor of others, but a desideratum nonetheless. Empirical investigations may be particularly useful in undermining analyses that are bolstered by false but superficially plausible accounts of the psychology of causal reasoning.

The empirical results described in section II.2 suggest human beings, from a very young age, make the sorts of intermodal inferences that would be licensed by the possession of an adequate account of causation. To that extent, the empirical data suggest that the type of theoretical role account of causation sketched in section II above does have a measure of psychological realism.

V.3. Theoretical Niche. As Hempel remarked, if it were possible to reduce theoretical concepts to observational ones, the former would be dispensable to theorizing, at least in principle. It is surely a prerequisite for the success of any theoretical role account that there be a need for something to play the role in question. To use an ecological metaphor, there must be a "theoretical niche" to be filled. The poor performance of apes in making intermodal inferences suggests that these inferences are cognitively nontrivial. This indicates that there is indeed a genuine need for something to play the sort of theoretical role that I have attributed to the concept of causation.

V.4. Expanded Theoretical Role. Although I have focused on the role of the concept of causation in mediating inferences between regularities, mechanisms, and manipulations, the concept of causation will surely play a much larger role within our life projects, and it would seem almost impossible to further articulate this role without leaving the armchair. For example, it is tempting to think that any creature, or at any rate any creature that engages in complex interactions with its social and physical environment, would need a concept of causation just to get by. Thus Michael Waldmann writes: "Our ability to acquire causal knowledge is central for our survival."[56] If recent developments in primatology really do show that even chimpanzees lack anything like our notion of causation, then this must be wrong: it is possible for an organism to fare quite well without engaging in anything like full-blown causal reasoning. The advantages that accrue to those who are able to employ a concept of causation come not at the level of avoiding predation and starvation, but perhaps at the level of tool construction and sophisticated social

[56] Waldmann, "Knowledge-based Causal Induction," *The Psychology of Learning and Motivation*, XXXIV (1996): 47–88, on p. 47.

interaction. Thus the role of specifically causal reasoning in our lives may be more subtle than has previously been appreciated.

V.5. A Metaphysical Interlude. A natural objection[57] to the line of argument I have been developing is that I am placing too much weight on the term 'concept'. It is handy to have an umbrella term like 'concept' to encompass the various things—causation, freedom, rationality, goodness, knowledge, and so on—that philosophers study. But it is a mistake to think that what philosophers are really interested in is the nature of our concepts. Rather, philosophers are interested in causation, freedom, rationality, goodness, knowledge, and so on.

In reply, I note that it is a common strategy for philosophers to attack their subject matter via our representations of it, be they "Ideas" (Locke and Hume), "Intuitions" (Kant), words (the logical positivists), or what have you. Even if what philosophers are really after is causation itself, it is hard to see how we can identify causation itself without being able to judge its conformity to our concept of causation. Thus conceptual analysis, in the very literal sense of gaining an understanding of the nature of our concepts, is at least a necessary prolegomenon to the analysis of causation, freedom, rationality, and the rest.

There is a fairly natural affinity between the sort of theoretical role analysis of concepts that I have been discussing and the sort of functionalist account that has been dubbed the "Canberra Plan." In the case of causation, such a functionalist account has been offered by Peter Menzies.[58] I have argued that the theoretical role of the concept of causation is to facilitate intermodal inferences; a natural further step would be to claim that causation itself is that feature of the world that renders such inferences legitimate. In other words, causation itself is that feature of the world that is typically present when there are characteristic regularities, which typically makes possible certain interventions, and so on. So far, my presentation has been agnostic on the question of whether there really is something there in the world that corresponds to the concept of causation. As with other theoretical concepts, I think that one could be a realist or an antirealist toward causation, and consistently maintain a theoretical role account of the concept of causation.[59] I think, however, that there is at least some reason to be skeptical of the full-blooded realist position; this will be the subject of the next subsection.

[57] Pressed upon me independently by Byrne and McGrath.

[58] Menzies, "Probabilistic Causation and the Pre-emption Problem," *Mind*, CV (1996): 85–117.

[59] I am especially indebted to audience members at the Australian National University—including Byrne, Chalmers, Jackson, Millikan, Sterelny, and Weisberg—for discussion of these issues.

V.6. Humean Modesty. It is worth remembering that Hume himself was as much a psychologist as a philosopher. One way of conceiving of the program of the *Treatise* is as an empirical investigation into the limits of human cognitive capacities. He used the results of this investigation to argue for a kind of modesty about the powers of human reason; we are more like ordinary animals than Aristotle's vaunted rational animals. The reason those tomes of metaphysics and theology contain nothing but sophistry and illusion is that they deal in subjects for which the human mind is just not fit.

Exploring the limits of the human mind is a tricky business: How does one discover limits without being able to see what lies beyond them? Reflecting on the limits of other species can provide an instructive analog. If a chimpanzee were to do philosophy, it seems a safe bet that she would not fare well in attempting to analyze causation. This failure would be due to a fundamental lacuna in her cognitive make-up that she would likely be unable to recognize. We might well be in a similar situation. In evolutionary time, we are mere beginners when it comes to causal reasoning. While nature has given us the capacity to make a heuristically useful family of intermodal inferences, there is no guarantee that we are tracking some specific feature of the world when we do so.

V.7. Self-Aggrandizing Conclusion. Recent developments in primatology suggest that we, meaning members of the species *homo sapiens*, are unique in possessing a full-blown concept of causation. This helps to put the enterprise of analyzing this concept in broader perspective. Philosophy is after all, part of the *human*ities. Nonetheless, there are certainly times, typically when I am constructing ever more convoluted counterexamples involving late preemption by double preventing omission, when I feel as though my work is disconnected from the bigger project of the humanities, to understand who we are, our culture, and our modes of expression. If, however, our concept of causation is part of what distinguishes us as a species, understanding it will indeed help us to better understand just who we are.

CHRISTOPHER HITCHCOCK
California Institute of Technology