Cheng, P.W. (2000). Causal reasoning. In R. Wilson & F. Keil (Eds.), *The MIT encyclopedia of cognitive sciences* (pp. 106-108). Cambridge, MA: MIT Press.

Rips, L. J. (1990). Reasoning. *Annual Review in Psychology* 41: 321–353.

Rosch, E., and C. B. Mervis. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605.

Rosch, E., C. B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. (1976). Basic objects in natural categories. *Cognitive Psychology* 8: 573–605.

Smith, E., and D. L. Medin. (1981). *Categories and Concepts.* Cambridge, MA: Harvard University Press.

Tanaka, J. W., and M. Taylor. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cognitive Psychology* 23: 457–482.

Tversky, A. (1977). Features of similarity. *Psychological Review* 84: 327–352.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: integrating information in object and social categorization. *Cognitive Psychology* 28: 274–328.

Wisniewski, E. J., and D. L. Medin. (1994). On the interaction of theory and data in concept learning. *Cognitive Science* 18: 221–281.

Causal Reasoning

Knowing that all pieces of butter have always melted when heated to 150°F, one would probably be willing to conclude that if the next solid piece of butter is heated to 150°F, it will melt. In contrast, knowing that all coins in Nelson Goodman's pockets, up to this point, were silver, one would be reluctant to conclude that if a copper coin were put in his pocket, it would become silver (examples adapted from Goodman 1954/1983). Why is it that one is willing to believe that heating causes butter to melt, but unwilling to believe that Goodman's pocket causes coins to be silver? These contrasting examples point to the kinds of questions that psychologists who study causal reasoning have asked and to the approaches taken in their answers. The central question is: What makes some sequences of events causal, thus licensing inference involving similar events, and other sequences noncausal?

The problem of causal INDUCTION as posed by David HUME (1739/1987) began with the observation that causal relations are neither deducible nor explicit in the reasoner's sensory input (where such input includes introspection as well as external sensory input). Given that sensory input is the ultimate source of all information that a reasoner has, it follows that all acquired causal relations must have been computed from (noncausal) sensory input in some way. A fundamental question therefore arises for such relations: How does a reasoner come to know that one thing, or type of thing, causes another? In other words, what is the mapping from observable events as input to causal relations as output?

The solution Hume (1739/1987) proposed is that causal relations are inferred from the spatial and temporal contiguity of the candidate cause c and the effect e, the temporal priority of c, and the constant conjunction between c and e. For the butter example, Hume's regularity approach might explain that one concludes that heating causes butter to melt from the fact that the heat is close to the butter, melting follows soon after heating, and whenever butter is heated to 150°F, its melting follows. Similarly, this approach might explain that one is reluctant to believe that Goodman's pocket causes coins to be silver, because his pocket probably did not come into existence before the coins did. If his pocket did predate all the coins in it, however, Hume's solution would fail: the coins were close to his pocket and they were silver whenever they were in his pocket, including soon after they were in his pocket.

One approach to the psychology of causal inference inherited the problem posed by Hume and extended his *regularity* solution. A branch of this approach was adopted by Kelley's (1973) ANOVA model and subsequent variants of it in social psychology (e.g., Hilton and Slugoski 1986). To illustrate contemporary statistical variants of Hume's solution, consider the contrasting examples again. Cheng and Holyoak's (1995) model would explain that a reasoner concludes that heating causes butter to melt because heating occurs before melting and melting occurs more often when the butter is heated to 150°F than when it is not, when other plausible influences on melting such as the purity of the butter are controlled. In contrast, a reasoner does not conclude that Goodman's pocket causes coins to be silver because one knows of alternative causes of coins being silver that might be uncontrolled. For example, Goodman might have selectively kept only silver coins in his pocket, whereas there is no such selection for coins outside his pocket.

As these examples illustrate, the regularity approach requires specific knowledge about alternative causes. But how does such knowledge come about in the first place? Unless one first knows *all* alternative causes, normative inference regarding a candidate cause seems impossible. Yet such inferences occur everyday in science and daily life. For example, without assuming that one knows all the alternative causes of butter melting, it is nonetheless possible to feel convinced, after observing the heating and subsequent melting of butter, that heating causes butter to melt.

An independent branch of the regularity approach begun in Pavlovian CONDITIONING, culminating in Rescorla and Wagner's (1972) connectionist model and its variants, and has been adopted to apply to human causal reasoning (e.g., Dickinson, Shanks, and Evenden 1984). This branch modifies Hume's solution in a manner similar to the statistical approach, but in addition provides algorithms for computing causal output from observational input. These connectionist variants of the regularity approach explain a wide range of empirical findings but have their shortcomings, such as a failure to explain the causal analog of the extinction of conditioned inhibition (for reviews, see Cheng 1997; Miller, Barnet, and Grahame 1995). A common source of these shortcomings may be the inability of these variants to represent causal power as an explicit variable existing independently of its value (*see* BINDING PROBLEM).

A second approach rejects all regularity solutions, and claims to offer an alternative solution to causal inference: one infers a relation to be causal when one perceives or knows of a causal mechanism or causal power underlying the relation (e.g., Koslowski 1996; Michotte 1963; Shultz 1982; White 1995). Because *power* theorists do not explicitly define causal "power" or causal "mechanism," it is unclear whether heating, for example, qualifies as a causal mechanism for substances melting. Assuming that it does, then power theorists would predict that heating should be

Cheng, P.W. (2000). Causal reasoning. In R. Wilson & F. Keil (Eds.), *The MIT encyclopedia of cognitive sciences* (pp. 106-108). Cambridge, MA: MIT Press.

understood to cause butter to melt. In contrast, reasoners do not know of any mechanism involving Goodman's pocket that would cause coins to be silver, and therefore would not believe that his pocket causes coins to be silver. Power theorists attempt to refute the regularity view by demonstrating that knowledge regarding specific causal powers influence causal judgments.

To regularity theorists, it is unclear what question the power approach seeks to answer; that question, however, is definitely not the one posed by Hume (Cheng 1993). If it is "What kind of causal inference do people, including infants, typically make in their everyday life?" then the answer is that they often make inferences based on prior causal knowledge (e.g., previously acquired knowledge that heating causes substances to melt). Regularity theorists, however, have no objection to the use of prior causal knowledge, as long as not all of that knowledge is innate; the kind of evidence offered by power theorists is therefore compatible with the regularity view (Cheng 1993; see Morris and Larrick 1995 for an example of an application of prior causal knowledge using a statistical approach). If the power solution were to be regarded as an answer to Hume's problem, then it begs the question: How does acquired knowledge about the *causal* nature of mechanisms (e.g., heating as a cause of melting) come about? That is, how does a reasoner infer a *causal* mechanism from *noncausal* observations? The answer to this question (the same question that the regularity view attempts but fails to answer) is what ultimately explains why one believes that one relation is causal and another not.

In addition to their other problems, neither the regularity nor the power approach can explain the boundary conditions for causal inference (see Cheng 1997 for a review). For example, neither explains why controlling for alternative causes allows a regularity to imply causality.

A third approach to the psychology of causal inference inherited Hume's problem, but modified his regularity solution radically by adding a Kantian framework that assumes an a priori notion of causal power. This notion differs critically from the causal knowledge presupposed by traditional power theorists in that it is general rather than specific (see INFANT COGNITION for assumptions regarding specific causal knowledge). According to this approach, the reasoner innately postulates that there exist such things as causes that have the power to produce an effect and causes that have the power to prevent an effect, and determines whether a regularity is causal by attempting to generate it with such general possible powers. By integrating the two previous approaches, this new power approach claims to explain a wide range of findings regarding causal inference, overcoming many problems that cripple earlier approaches (Cheng 1997). The same basic approach has been adopted by computer scientists and philosophers in the last decade to study how it is possible in principle to draw inferences about causal networks from patterns of probabilities (BAYESIAN NETWORKS; Pearl 1995; Spirtes, Glymour, and Scheines 1993). Although psychological work has begun on aspects of causal networks (Busemeyer, McDaniel, and Byun 1997; Spellman 1997), how humans and other animal species infer causal networks remains to be investigated.

See also CAUSATION; CONCEPTS; DEDUCTIVE REASONING; EXPLANATION-BASED LEARNING;

—*Patricia Cheng*

References

Busemeyer, J., M. A. McDaniel, and E. Byun. (1997). Multiple input-output causal environments. *Cognitive Psychology* 32: 1–48.

Cheng, P. W. (1993). Separating causal laws from casual facts: pressing the limits of statistical relevance. In D. L. Medin, Ed., *The Psychology of Learning and Motivation, 30.* New York: Academic Press, pp. 215–264.

Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review* 104: 367–405.

Cheng, P. W., and K. J. Holyoak. (1995). Complex adaptive systems as intuitive statisticians: causality, contingency, and prediction. In H. L. Roitblat and J.-A. Meyer, Eds., *Comparative Approaches to Cognitive Science.* Cambridge, MA: MIT Press, pp. 271–302.

Dickinson, A., D. R. Shanks, and J. L. Evenden. (1984). Judgment of act-outcome contingency: the role of selective attribution. *Quarterly Journal of Experimental Psychology* 36A: 29–50.

Goodman, N. (1954/1983). *Fact, Fiction, and Forecast.* Fourth edition. Cambridge, MA: Harvard University Press.

Hilton, D. J., and B.R. Slugoski. (1986). Knowledge-based causal attribution: the abnormal conditions focus model. *Psychological Review* 93: 75–88.

Hume, D. (1739/1987). *A Treatise of Human Nature.* Second edition. Oxford: Clarendon Press.

Kant, I. (1781/1965). *Critique of Pure Reason.* London: Macmillan and Co.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist* 28: 107–128.

Koslowski, B. (1996). *Theory and Evidence: The Development of Scientific Reasoning.* Cambridge, MA: MIT Press.

Michotte, A. E. (1946/1963). *The Perception of Causality.* New York: Basic Books.

Miller, R. R., R. C. Barnet, and N. J. Grahame. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin* 117: 363–386.

Morris, W. M., and R. Larrick. (1995). When one cause casts doubts on another: a normative analysis of discounting in causal attribution. *Psychological Review* 102: 331–355.

Pearl, J. (1995). Causal diagrams for experimental research, *Biometrika* 82(4): 669–710.

Rescorla, R. A., and A. R. Wagner. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy, Eds., *Classical Conditioning II: Current Theory and Research.* New York: Appleton-Century-Crofts, pp. 64–99.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development* 47 (1).

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General* 126: 1–26.

Spirtes, P., C. Glymour, and R. Scheines. (1993). *Causation, Prediction and Search.* New York: Springer-Verlag.

White, P. A. (1995). Use of prior beliefs in the assignment of causal roles: causal powers versus regularity-based accounts. *Memory and Cognition* 23: 243–254

Further Readings

Hart, H. L., and A. M. Honoré. (1959/1985). *Causation in the Law.* 2nd ed. Oxford: Oxford University Press.

Cheng, P.W. (2000). Causal reasoning. In R. Wilson & F. Keil (Eds.), *The MIT encyclopedia of cognitive sciences* (pp. 106-108). Cambridge, MA: MIT Press.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation.* Oxford: Clarendon Press.

Shanks, D. R., K. J. Holyoak, and D. L. Medin, Eds. (1996). *The Psychology of Learning and Motivation,* vol. 34: *Causal Learning.* New York: Academic Press.

Sperber, D., D. Premack, and A. J. Premack, Eds. (1995). *Causal Cognition: A Multidisciplinary Debate.* New York: Oxford University Press.