Research Article

WHY CAUSATION NEED NOT FOLLOW FROM STATISTICAL ASSOCIATION: Boundary Conditions for the Evaluation of Generative and Preventive Causal Powers

Melissa Wu and Patricia W. Cheng

University of California, Los Angeles

Abstract—*In experimental design, a tacit principle is that to test whether a candidate cause* c *(i.e., a manipulation) prevents an effect* e, e *must occur at least some of the time without the introduction of* c. *This principle is the preventive analogue of the explicit principle of avoiding a ceiling effect in tests of whether* c *produces* e. *Psychological models of causal inference that adopt either the covariation approach or the power approach, among their other problems, fail to explain these principles. The present article reports an experiment that demonstrates the operation of these principles in untutored reasoning. The results support an explanation of these principles according to the power PC theory, a theory that integrates the previous approaches to overcome the problems that cripple each.*

As a young child, one of the authors once dropped a vase right before a power outage occurred. She thought that by dropping the vase, she caused the outage, and so began to cry. One might simply chuckle at this anecdote; but upon further thought, it raises some important issues. Why do adults, and not the child, know that dropping a vase cannot cause a power outage? We return to the case of the child later; but first, how do adults come to know what they know?

One may have two intuitions about how adults reach their interpretation of the vase-dropping episode. First, they may know that the act of dropping an object such as a vase does not involve any mechanism that causes outages. Second, they may know that power outages do not follow more often when objects are dropped than when they are not. In other words, "dropping objects" does not *covary* with subsequent power outages. These two related lines of thought, in fact, respectively illustrate two divergent dominant approaches to the psychology of causal induction, the mechanism, or *power,* view (e.g., Ahn, Kalish, Medin, & Gelman, 1995; Koslowski, 1996; Shultz, 1982; White, 1989, 1995) and the *covariation* view (e.g., Allan, 1980; Cheng & Novick, 1990; Kelley, 1967; Pearce, 1994; Rescorla & Wagner, 1972; Shanks & Dickinson, 1987).

Power theorists have typically cast the two views in opposition to each other. However, the debate is bogus (Cheng, 1993; Cheng & Lien, 1995; Glymour & Cheng, 1998): The power view concerns the application of causal knowledge, whereas the covariation view concerns the discovery of such knowledge. Pitting explanations of disparate psychological processes against each other cannot be fruitful. Worse yet, the debate obscures some problems common to both views. One of these concerns the boundary conditions for evaluating causal power, as illustrated in the following scenario:

Address correspondence to Patricia Cheng, Department of Psychology, Franz Hall, University of California, Los Angeles, CA 90095-1563; e-mail: cheng@lifesci.ucla.edu.

Suppose you are a researcher who has developed a new drug. According to your theory, when this drug is administered to young cats of a dangerous breed, it will prevent the cats from becoming fertile when they reach maturity. You ask your research assistant to conduct a test of this drug. He randomly assigns young cats of the breed in question to two groups. None of the cats in either group starts out fertile. He administers the drug to one group but not the other, taking steps to ensure that all other factors are held constant and that the drug is administered properly. At the end of the experiment, your assistant informs you that there was no difference between the two groups in the proportion of fertile cats. This result is reliable because he used a large sample of cats. He concludes that the drug is ineffective. You examine the results closely and find that at the end of the experiment, none of the cats in his two groups were fertile.

Would you agree with your assistant that the drug is ineffective? More likely, you would think that he has conducted an uninformative test of the drug; the test violates a boundary condition for causal inference. This hypothetical anecdote illustrates that testing whether a *candidate cause c* (the drug) prevents an effect *e* (reduces the frequency of fertility) requires that *e* occurs at least some of the time without *c*. This principle is so ingrained that it is followed without exception in experimental design even though it is never explicitly stated in any textbooks. It is not the typical floor effect, which concerns situations in which some enabling conditions of a candidate *c* that produces *e* are missing so that *c* does not increase the frequency of *e*. Why is there this unspoken principle in experimental design?

This article focuses on situations in which there is no difference between the frequency of *e* when *c* occurs and when it does not (i.e., $P(e|c) = P(e|\bar{c}) = P(e)$, where $P(e|c)$ is the probability of *e* given the presence of *c*, and $P(e|\bar{c})$ is that probability given the absence of *c*). Now, consider another situation: What if at least some, but still an equal proportion, of the cats in the research assistant's two groups were fertile at the end of the experiment? You should now be more likely to agree that the drug is ineffective.

The covariation view cannot explain why the ability of reasoners to reach a conclusion about the causal status of *c* depends on the overall frequency of *e*. Rescorla and Wagner's (1972) model, for example, predicts a causal strength of 0 for both situations regardless of the values of its rate parameters. Even for a covariation model that allows more flexible weighting of frequency information, such as Schustack and Sternberg's (1981) model, some definite output should result in both situations (e.g., intermediate causal strengths), rather than an output resulting in the latter situation but not in the former. The lack of output would correspond to judging that the test is uninformative. The difference in the intuitive answers for these scenarios confounds all variants of the covariation view, whether the variant is statistical (e.g., Cheng & Novick, 1990) or associationist (e.g., Rescorla & Wagner, 1972). For both kinds of variants, the failure is traceable to their lack of representation of causal power as a variable existing independently of the value of covariation (see Lien & Cheng, 1998). Thus, purely covariational

models are inherently incapable of having an unbound causal-power variable while the covariation in the situation has a clear-cut value.

The power view does not fare any better. All knowledge about specific causal powers or mechanisms is constant across these situations. In fact, this view cannot explain the intuitive answer in either situation: The drug could have produced a difference in fertility between groups, and a plausible mechanism must therefore exist in both situations. Failing to incorporate the use of covariation, this view cannot explain why the drug is considered effective in neither situation, nor can it explain why you would think the test was uninformative in one situation but not the other.

An analogous problem arises for both traditional views when one evaluates whether *c* produces *e*. To continue our example, suppose your goal is to evaluate whether a drug causes cats of an endangered breed that otherwise would have been infertile at maturity to become fertile. Now, if all cats in both groups became fertile at the end of the test, the test would be uninformative. This situation corresponds to a ceiling effect in experimental design. In contrast, if not all cats, but still an equal proportion of them, in the two groups became fertile, you would likely infer that the drug is ineffective. The difference in causal judgments between these situations confounds each traditional view.

THE POWER PC THEORY

An integration of these views overcomes these and other problems common to the two views (Cheng, 1997). The relationship between causal power and covariation is analogous to that between a scientist's theory (e.g., the kinetic theory of gases) and the laws that it explains (e.g., Boyle's Law). Just as Boyle's Law is defined in terms of observable entities, covariation is defined by probabilities that are estimable by observable frequencies. Just as the kinetic theory postulates unobservable entities, the *power PC theory* (a causal-power theory of the probabilistic-contrast model) postulates unobservable causal powers. The *generative power* of candidate cause *c* with respect to effect *e* is the probability with which *c* produces *e* (Cartwright, 1989). The *preventive power* of *c* is the probability with which *c* prevents an otherwise-occurring *e* from occurring. To estimate these unobservable powers, reasoners bootstrap by using them to explain their covariation model, the probabilistic-contrast model (Cheng & Novick, 1990). According to this model, given that *c* is perceived to occur before *e*, reasoners assess the *probabilistic contrast* for *c* with respect to *e*, ΔP_c over a *focal set*, where

$$\Delta P_c = P(e|c) - P(e|\bar{c}) \quad (1)$$

and the focal set is the set of events that the reasoner uses as input to the covariation process, a set that is not necessarily the universal set assumed by most previous psychologists studying causal inference. A mental construct in the reasoner's explanation is a distinction between *c* and the composite of (known and unknown) causes alternative to *c*, which we label *a*. For example, when reasoners evaluate whether *c* produces *e*, they explain $P(e|c)$ in Equation 1 by the union of two events: (a) *e* produced by *c* and (b) *e* produced by *a* if *a* occurs in the presence of *c*. That is, they reason that when *c* is present, *e* can be produced by *c*, or by *a* if *a* occurs in the presence of *c*. Likewise, reasoners explain $P(e|\bar{c})$ in that equation by *e* being produced by *a* alone when *a* occurs in the absence of *c* (because when *c* is absent, it cannot produce *e*).

This theory explains a diverse range of findings that are inexplicable by previous accounts (see Cheng, 1997), such as the boundary conditions for causal inference and the distinctions among a novel or unknown factor, an irrelevant factor, an enabling condition, a noncausal but covarying factor, and a cause. It concerns simple direct causes (see Glymour, 1998, and Glymour & Cheng, 1998, for extensions of the theory to indirect causes; see Novick & Cheng, 1998, for an extension to conjunctive causes—causes that typically produce an effect only in combination, the way both talent and hard work in combination lead to success).

Three Boundary Conditions for Causal Inference

A mathematical consequence of these explanations of the components of contrast is that if and only if *a* occurs independently of *c* (i.e., $P(a \mid c) = P(a \mid \bar{c}) = P(a)$), then for nonnegative contrasts ($\Delta P_c \geq 0$),

$$p_c = \frac{\Delta P_c}{1 - P(e \mid \bar{c})}. \quad (2)$$

This derivation (Cheng, 1997) assumes that *c* and *a* influence *e* independently, in which case p_c in Equation 2 is the generative power of *c*. Without this assumption, as long as *a* occurs independently of *c*, p_c in this equation is the generative power of a conjunction of *c* and whatever part of *a* interacts with *c* (e.g., its enabling conditions; see Cheng, in press). Notice that to infer p_c, the theory does not require knowledge of *a* or its power, even though the derivations make use of the assumption that *e* can be produced by *a*. If *a* does not occur independently of *c*, p_c is unestimable; the independent occurrence of *a* and *c* is hence a boundary condition for estimating p_c. (Such independence can be achieved by selecting a focal set in which *a* is held constant; see Cheng, 1997, for a review of evidence showing the spontaneous selection of such focal sets by humans and other species.) This condition is satisfied by all scenarios in the experiment we report in this article, and is not our focus.

A second boundary condition for the evaluation of generative causal power is that $P(e|\bar{c})$ must be less than 1. This condition follows from Equation 2: When $P(e|\bar{c}) = 1$, p_c is undefined. The violation of this condition explains why a reasoner cannot judge whether the drug in the scenario increases fertility if all cats became fertile, implying that $P(e|\bar{c}) = 1$. When this second boundary condition is met, Equation 2 implies that $p_c = 0$ (i.e., *c* is noncausal) when $\Delta P_c = 0$.

There is an analogous specification of the relation between nonpositive contrasts ($\Delta P_c \leq 0$) and preventive causal power (Cheng, 1997). The only assumption that differs in the preventive case is that reasoners assume that *c* may prevent *e*, rather than produce it, with some probability. This difference implies that when reasoners evaluate whether *c* prevents *e*, they explain $P(e|c)$ by the intersection of two events: (a) *e* produced by *a* if *a* occurs in the presence of *c* and (b) *e* not stopped by *c*. That is, they reason that when *c* is present, *e* occurs only if it is both produced by *a* and not stopped by *c*. $P(e|\bar{c})$ is explained just as before (because only *a* is relevant). These explanations imply that if *a* occurs independently of *c*, then for nonpositive contrasts,

$$p_c = \frac{-\Delta P_c}{P(e \mid \bar{c})}. \quad (3)$$

Equation 3 is the preventive analogue of Equation 2; p_c here is the preventive power of *c*. One critical difference is that whereas Equation 2 is undefined when $P(e|\bar{c}) = 1$, Equation 3 is undefined when $P(e|\bar{c}) = 0$. This difference leads to a third boundary condition: $P(e|\bar{c}) > 0$ for the evaluation of preventive power. When this condition is met, Equation 3 implies that $p_c = 0$ when $\Delta P_c = 0$. But, when $P(e|\bar{c}) = 0$, as when no cat in the scenario becomes fertile, a reasoner cannot judge whether *c* prevents *e*.

Summary

When alternative causes occur independently of candidate *c* and $\Delta P_c = 0$, the power PC theory predicts that the type of causal power assessed (generative vs. preventive) and $P(e|\bar{c})$ should interact to influence causal judgments. In particular, in such situations, reasoners should be unable to judge whether *c* produces *e* when $P(e) = 1$. When $P(e) < 1$, however, they should infer that *c* does not produce *e*. Analogously, reasoners should be unable to judge whether *c* prevents *e* when $P(e) = 0$. But when $P(e) > 0$, they should infer that *c* does not prevent *e*. This mundane set of predictions eludes all psychological accounts of causal inference that fail to integrate causal power with covariation.

EXPERIMENT

This experiment tested the interaction between the type of power assessed and $P(e)$, as implied by the last two boundary conditions derived in the power PC theory. We attempted to test the use of these conditions more directly and comprehensively than in previous studies (see Cheng, 1997, and Spellman, 1996, for reviews of previous support for their use). Participants were each given a scenario such as the cat example presented earlier, in which $P(e|c) = P(e|\bar{c})$, and were asked to judge whether they thought the status of *c* was causal, noncausal, or uncertain. The scenarios differed between participants in two respects: the type of causal power assessed and the relative frequency of *e* ($P(e) = 1$, 0, or a qualitative value in between).

All accounts other than the power PC theory predict that for each level of $P(e)$, regardless of parameter settings, as long as those settings are consistent, judgment should remain constant across the kinds of power. In particular, no competing account predicts that judgment should be withheld for some condition but not others. More specifically, some competing models predict that judgments should not differ at all among the six experimental conditions. For example, Rescorla and Wagner's (1972) model, Pearce's (1994) model (both regardless of their parameter settings), and the traditional ΔP model (e.g., Allan, 1980; Jenkins & Ward, 1965) all predict that participants will always judge *c* to be noncausal.[1] Linear combination models (e.g., Schustack & Sternberg, 1981) predict that *c* is noncausal, generative, or preventive, depending on their parameter settings. The weighted-ΔP model (e.g., Anderson & Sheu, 1995) makes two predictions: When $P(e) = 0$ (regardless of parameter settings), *c* is noncausal regardless of the kind of power assessed, and for other values of $P(e)$, depending on the parameter settings, *c* is noncausal, generative, or preventive.

Method

We conducted our experiment with 155 undergraduates at the University of California, Los Angeles. They were attending their first class on the first day of the fall quarter and therefore had not yet been exposed to college instructions on scientific methodology. Given the constraints of our subject pool, the study was conducted as part of a 50-min package of questionnaires.

The two primary manipulations were as explained. To increase the generality of the study, we constructed scenarios that concerned two domains for each of the six conditions (two kinds of causal power crossed with three levels of $P(e)$), resulting in 12 scenarios. Each participant was randomly assigned to one of these scenarios. We measured how a result reported in the scenario was interpreted.

The cat scenario presented earlier is an example of our materials for one domain. Both domains involved a researcher who wished to test whether a certain *c* produced a discrete characteristic (the effect). The researcher's assistant randomly assigned subjects[2] to two groups, introduced *c* to the experimental group but not the control group, held all other factors constant, made sure that none of the subjects began the test possessing the characteristic in question, and used a large sample of subjects. At the end of the test, there was no difference in the proportion of subjects who possessed the characteristic. The assistant concluded that *c* was noncausal. In one domain, *c* was a drug that purportedly affects the proportion of fertile cats in a sample, whereas in the other domain, *c* was a gene therapy that purportedly affects the proportion of plants that undergo a certain mutation.

Within each domain, the scenarios differed only as follows. In the *generative* conditions, *c* was intended to increase the proportion of subjects with the characteristic, whereas in the *preventive* conditions, *c* was intended to decrease that proportion. The justifications for these intentions were, respectively, that the characteristic was beneficial (e.g., fertility in an endangered breed of cats) and that it was harmful (e.g., fertility in a dangerous breed of cats). At the end of the test in the *all* conditions (where $P(e) = 1$), the researcher found that all subjects in both groups had the characteristic; in the *some* conditions (where $0 < P(e) < 1$), "some, but not all" subjects in both groups had the characteristic; and in the *none* conditions (where $P(e) = 0$), no subject had the characteristic.

A subsequent question asked whether participants (a) agreed with the assistant's conclusion that *c* was ineffective, (b) thought that his test was uninformative, or (c) disagreed with him and judged *c* to be effective. Participants indicated which statement they most agreed with. An explanation was requested to encourage thinking.

Results and Discussion

Figure 1 presents the percentage of participants in each of the six conditions who judged the status of *c* to be noncausal, uncertain, or causal (i.e., effective). These percentages were collapsed over the two domains (fertility and mutation) because there was no reliable difference between domains in any condition. The participants' explanations are not in general reported because they were uninformative.

1. Our predictions for associationist models assume the inputs to these models to be sequences of trials that are consistent with the summary data presented in the corresponding conditions.

2. We use the term "subjects" to refer to the cats and plants mentioned in the scenarios and reserve the term "participants" to denote the humans who read the scenarios; likewise, we use "test" to refer to the experiment in the scenarios and "experiment" to refer to our study.

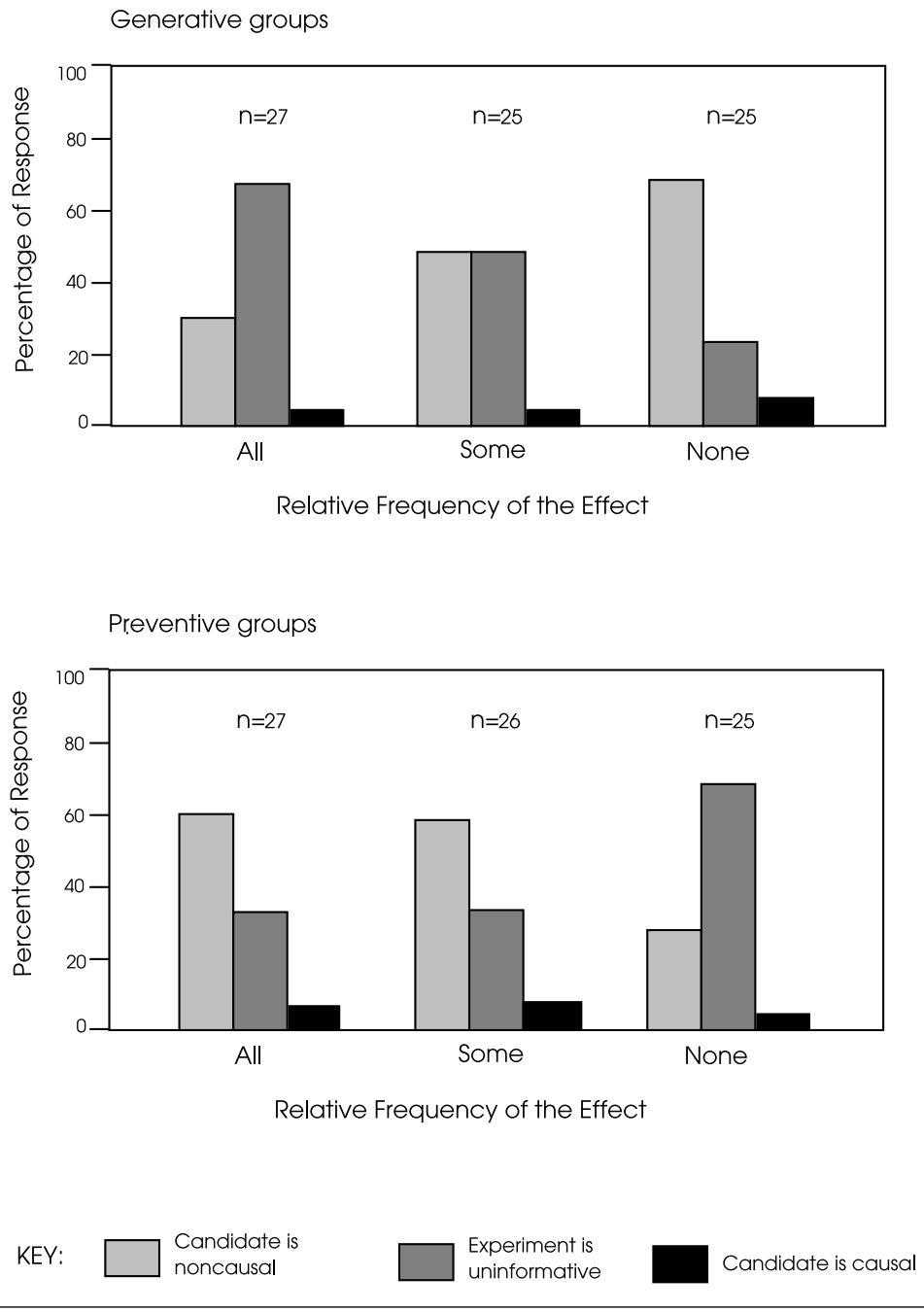

Fig. 1. Percentage of various responses for participants who evaluated generative and preventive causal powers when the effect in question occurred in all, some, or none of the subjects (*n* is the number of participants in each group).

Figure 1 shows that for most conditions, a moderate majority of the participants responded in accord with the power PC theory. One possible interpretation is that this theory does only moderately well, and the other accounts do worse. As in other psychological experiments, however, there existed extraneous variables that may have influenced participants' responses. Because this study was conducted as part of a long questionnaire, our procedure did not encourage attentiveness; if participants did not read, understand, remember, or accept one or another of the many critical aspects of the scenario given, they would not answer as predicted by a theory even if that theory does describe their causal reasoning. Evaluation should therefore be based on comparisons between conditions, in which the extraneous variables were presumably controlled.

The comparisons were evaluated using chi-square tests. Because very few participants judged *c* to be causal (see Fig. 1), this option is omitted from the analyses we report. (Including the omitted option

yields an identical pattern of reliability of the results.) That is, all chi-square analyses compared (a) the number of participants who indicated that *c* was noncausal and (b) the number who indicated that the test was uninformative.

Comparisons between the generative and preventive conditions

In this section, we compare the generative conditions with the preventive conditions with the same level of $P(e)$. The following pattern of results was predicted uniquely by the power PC theory:

- When $P(e) = 1$, participants who evaluated generative power were more likely than those who evaluated preventive power to indicate that the test was uninformative, $\chi^2(1, N = 51) = 5.64$, $p < .05$.
- When $0 < P(e) < 1$, the distribution of responses did not depend on whether the participants evaluated generative or preventive power, $\chi^2(1, N = 48) = 0.76$, $p > .30$.
- When $P(e) = 0$, participants who evaluated preventive power were more likely than those who evaluated generative power to indicate that the test was uninformative, $\chi^2(1, N = 47) = 9.41$, $p < .01$.

Comparisons between P(e) *conditions*

Participants who evaluated generative power were more likely to indicate that the test was uninformative when $P(e) = 1$ than when $0 < P(e) < 1$, as predicted by the power PC theory. This difference, however, was not reliable, $\chi^2(1, N = 50) = 1.92$, $p > .10$. An unexpectedly high proportion of participants in the generative-some condition indicated that the test was uninformative.[3]

All other comparisons between $P(e)$ conditions for the evaluation of the same type of power strongly support the power PC theory:

- Participants who evaluated generative power were more likely to indicate that the test was uninformative when $P(e) = 1$ than when $P(e) = 0$, $\chi^2(1, N = 49) = 9.09$, $p < .01$.
- When participants evaluated generative power, the distribution of responses did not depend on whether $0 < P(e) < 1$ or $P(e) = 0$, $\chi^2(1, N = 47) = 2.84$, $p > .05$.
- Participants who evaluated preventive power were more likely to indicate that the test was uninformative when $P(e) = 0$ than when $0 < P(e) < 1$, $\chi^2(1, N = 48) = 5.37$, $p < .05$, or when $P(e) = 1$, $\chi^2(1, N = 49) = 5.96$, $p < .02$.
- When participants evaluated preventive power, the distribution of responses did not depend on whether $0 < P(e) < 1$ or $P(e) = 1$, $\chi^2(1, N = 49) = 0.01$, $p > .50$.

This pattern of interaction contradicts Rescorla and Wagner's model (1972), Pearce's model (1994), and the ΔP model, all of which predict that these comparisons should show no differences. Our results also fail to support the weighted-ΔP model and linear combination models: *c* was hardly ever judged to be causal.

3. Of the 12 participants who chose this answer, 4 explained that the results were not clear-cut because only "some" subjects showed the effect.

SUMMARY AND IMPLICATIONS

Our pattern of results shows that, as predicted uniquely by the power PC theory, the kind of causal power assessed and the probability of the effect interact in their influence on the assessment of the causal power of a candidate *c* when $\Delta P_c = 0$ and alternative causes are assumed to occur independently of *c*. This pattern does not depend on reasoners receiving training in experimental design, as the participants in our experiment had at most minimal such training, yet adopted the design principle regarding the ceiling effect and its tacit preventive analogue. Our results add to the growing body of phenomena that confound psychological accounts of causal induction that exclusively adopt the power or the covariation approach. The relative success of the power PC theory suggests that both concepts are necessary: People explain their model of covariation according to a theory of causal power.

We have focused on how people discover causal mechanisms involving a single cause-to-effect link. We have not addressed the issue of how specific prior causal knowledge (innate or learned) influences subsequent reasoning (but see Lien & Cheng, 1998). Neither have we addressed the learning of causal chains and other multilayered causal networks. However, methods that do address these issues are consistent with our approach (Glymour, 1998; Pearl, 1988, 1995; Spirtes, Glymour, & Scheines, 1993; see Glymour & Cheng, 1998).

Finally, another important issue that we have not addressed is illustrated by the case of the child in our introduction: Why would a young child think that dropping the vase caused the power outage? Our conjecture is that the child's candidate cause, unlike an adult's, is the dropping of the vase rather than of an object. For the child, power outage follows the only episode of dropping a vase, whereas it never or rarely occurs otherwise. If the child assumes that other causes of power outage are constant before and after dropping the vase, she would conclude that dropping the vase caused the outage. An issue that awaits further investigation is, what determines the category or concept that serves as a candidate cause?

Acknowledgments—This article is based on an undergraduate honors thesis submitted by the first author, under the advisorship of the second author, to the Department of Psychology at the University of California, Los Angeles. We thank Keith Holyoak, John Kihlstrom, Ralph Miller, and two anonymous reviewers for their helpful comments on an earlier version of this article.

REFERENCES

Ahn, W., Kalish, C.W., Medin, D.L., & Gelman, S.A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54,* 299–352.

Allan, L.G. (1980). A note on measurements of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15,* 147–149.

Anderson, J.R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition, 23,* 510–524.

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, England: Clarendon Press.

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: Vol. 30* (pp. 215–264). New York: Academic Press.

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

Cheng, P.W. (in press). Inferring simple and conjunctive direct causes. In F. Keil & R. Wilson (Eds.), *Cognition and explanation.* Cambridge, MA: MIT Press.

Cheng, P.W., & Lien, Y. (1995). The role of coherence in differentiating genuine from spurious causes. In D. Sperber, D. Premack, & A.J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 463–494). New York: Oxford University Press.

Cheng, P.W., & Novick, L.R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545–567.

Glymour, C. (1998). Psychological and normative theories of causal power and the probabilities of causes. In G.F. Cooper & S. Moral (Eds.), *Uncertainty in artificial intelligence* (pp. 166–172). San Francisco: Morgan Kaufmann.

Glymour, C., & Cheng, P.W. (1998). Causal mechanism and probability: A normative approach (pp. 295–313). In M. Oaksford & N. Chater (Eds.), *Rational models of cognition.* Oxford, England: Oxford University Press.

Jenkins, H., & Ward, W. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs, 7,* 1–17.

Kelley, H.H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation: Vol. 15* (pp. 192–238). Lincoln: University of Nebraska Press.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Lien, Y., & Cheng, P.W. (1998). *Distinguishing genuine from spurious causes: A coherence hypothesis.* Unpublished manuscript, University of California, Los Angeles.

Novick, L.R., & Cheng, P.W. (1998). *A power theory of conjunctive causes.* Unpublished manuscript, University of California, Los Angeles.

Pearce, J.M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review, 101,* 587–607.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

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

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

Schustack, M.W., & Sternberg, R.J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120.

Shanks, D., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 229–261). San Diego: Academic Press.

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

Spellman, B.A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7,* 337–342.

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

White, P.A. (1989). A theory of causal processing. *British Journal of Psychology, 80,* 431–454.

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

(RECEIVED 6/20/97; REVISION ACCEPTED 9/15/98)