

# The Meaning and Computation of Causal Power: Comment on Cheng (1997) and Novick and Cheng (2004)

Christian C. Luhmann  
Vanderbilt University

Woo-kyoung Ahn  
Yale University

D. Hume (1739/1987) argued that causality is not observable. P. W. Cheng (1997) claimed to present “a theoretical solution to the problem of causal induction first posed by Hume more than two and a half centuries ago” (p. 398) in the form of the power PC theory (L. R. Novick & P. W. Cheng, 2004). This theory claims that people’s goal in causal induction is to estimate causal powers from observable covariation and outlines how this can be done in specific conditions. The authors first demonstrate that if the necessary assumptions were ever met, causal powers would be self-evident to a reasoner—they are either 0 or 1—making the theory unnecessary. The authors further argue that the assumptions the power PC theory requires to compute causal power are unobtainable in the real world and, furthermore, people are aware that requisite assumptions are violated. Therefore, the authors argue that people do not attempt to compute causal power.

*Keywords:* causal induction, causal power, power PC theory, reasoning

Hume (1739/1987) argued that causality is not present in our experience and that the conception of causality results from inductive inferences based on several observable cues such as covariation. Covariation is obviously a fallible indicator of causal relationships. For example, the number of drownings and sales of ice cream presumably covary but do not cause each other. In response, Cheng (1997) suggested that if covariation is explicitly treated as a consequence of unobservable causal relationships, then inferences may be made that accurately reflect the causal strength of those relationships. The power PC theory (Cheng, 1997; Novick & Cheng, 2004) claims to correctly extract these causal powers under certain circumstances. Indeed, Cheng (1997) wrote that in many situations the power PC theory “presents a theoretical solution to the problem of causal induction first posed by Hume more than two and a half centuries ago” (p. 398). On the basis of the claim that reasoners seek to infer causal powers, the power PC theory is also presented as a descriptive model of human causal induction. In what follows, we evaluate these claims.

## The Power PC Theory

Cheng (1997) defined causal power as the probability with which a cause produces an effect when the cause is present. This quantity differs from the probability of the effect given the presence of the cause (i.e.,  $P(e|c)$ ), because the latter includes those occasions when the cause brought about the effect as well as

occasions on which the cause was present but failed to bring about the effect and an alternative cause brought about the effect. Causal power aims to capture the probability with which the cause actually causes the effect.

The feature of the power PC theory that sets it apart from other covariation accounts (e.g., Cheng & Novick, 1990; Rescorla & Wagner, 1972) is that it seeks to compute a context-free description of the causal relationship. Most previous covariation accounts provide an estimate of causal strength that is inherently bound to the context in which the learning occurred. Thus, if causal strength is computed from a set of observations in one setting (the learning context), this quantity may be of no value in any other context. In contrast, the causal power of a given relationship does not vary with context (e.g., the number or effectiveness of other causes). The power PC theory is proposed as a both normative and descriptive account of causal induction. Thus, it is argued that a reasoner’s goal is “to induce the unobservable causal power of a candidate cause in the distal world from observable events represented in the proximal stimulus” (Buehner, Cheng, & Clifford, 2003, p. 1120).

To compute causal power, Cheng (1997) began by assuming causal situations<sup>1</sup> contain two causes and an effect. One cause is the candidate cause—the cause we are interested in evaluating (called  $i$ ). The second cause is a composite that represents all other causal factors in the world (called  $a$ ). These two causes operate to bring about the effect of interest (called  $e$ ). Here are the assumptions as Cheng (1997) laid them out.

---

Christian C. Luhmann, Department of Psychology, Vanderbilt University; Woo-kyoung Ahn, Department of Psychology, Yale University.

Support for this research was provided in part by National Institute of Mental Health Grant R01-MH57737 to Woo-kyoung Ahn.

Correspondence concerning this article should be addressed to Christian C. Luhmann, 2 Hillhouse Avenue, New Haven, CT 06511. E-mail: christian.luhmann@vanderbilt.edu

---

<sup>1</sup> Here we describe the power PC theory’s process for computing the simple causal power of a generative cause (Cheng, 1997). This is done for the sake of clarity and brevity. Despite this focus, the following analysis can be similarly applied to the preventative causes as well as interactive causes.

1. when  $i$  occurs, it produces  $e$  with probability  $p_i$ ; when  $a$  occurs, it produces  $e$  with probability  $p_a$ ; and nothing else influences the occurrence of  $e$ ;
2.  $i$  and  $a$  influence the occurrence of  $e$  independently; and
3.  $i$  and  $a$  influence the occurrence of  $e$  with causal powers that are independent of how often  $i$  and  $a$  occur. (p. 373)

An additional assumption that is only implied by Cheng (1997)<sup>2</sup> and more explicitly stated by Cheng (2000) is that  $a$  can produce  $e$  but not prevent it (this is hereafter referred to as Assumption 4 for brevity).

Given this situation, the presence and absence of  $i$  and  $a$  operate to generate the covariation we ultimately observe. Under these assumptions, the observable probabilities can be related to the unobserved causal powers. The target event,  $e$ , occurs in the presence of  $i$  when  $e$  is caused by  $i$  or  $a$ :

$$P(e|i) = p_i + P(a|i) \times p_a - p_i \times P(a|i) \times p_a. \quad (1)$$

Similarly,  $e$  occurs in the absence of  $i$  because of  $a$ :

$$P(e|\sim i) = P(a|\sim i) \times p_a. \quad (2)$$

The quantity of interest,  $p_i$ , can then be solved for, resulting in Equation 3:

$$p_i = \frac{[P(e|i) - P(e|\sim i)] - \{[P(a|i) - P(a|\sim i)] \times p_a\}}{1 - P(a|i) \times p_a}. \quad (3)$$

This equation computes the causal power of the candidate, but we are still left needing quantities that are unavailable. For example,  $p_a$ , the causal power of the composite alternative cause,  $a$ , is, like all causal powers, unobservable. Because of this, Equation 3 cannot be used. To resolve this dilemma, Cheng (1997) noted when the occurrence of the candidate cause is independent of the occurrence of the alternative cause (i.e.,  $P(a|i) = P(a|\sim i)$ , referred to hereafter as Assumption 5), Equation 3 simplifies as follows:

$$p_i = \frac{P(e|i) - P(e|\sim i)}{1 - P(e|\sim i)}. \quad (4)$$

Equation 4 relates causal power to probabilities that are observable, allowing us to estimate causal power. Although there have been interesting debates about whether people's causal judgments match the quantitative predictions of Equation 4 (e.g., Buehner et al., 2003; Lober & Shanks, 2000), we focus on the assumptions behind Equation 4, an aspect of the theory that has rarely been questioned.

#### When All Requisite Assumptions Are Satisfied

Let us, for a moment, assume all these assumptions are met. This situation contains a single generative cause,  $i$ ; a single effect,  $e$ ; and an alternative, generative cause,  $a$ . In this situation,  $i$  and  $a$  do not interact, and there are no preventative alternative causes. Because all of these assumptions are necessary in deriving Equation 4, this is the only possible world where it is safe to apply Equation 4.

Note that in this situation, if  $i$  causes  $e$  at all (i.e.,  $p_i > 0$ ),  $e$  must follow whenever  $i$  occurs. Because there are no preventative

causes in this world (Assumption 4), it cannot be the case that  $e$  was prevented from occurring in this situation. Similarly, because there are no causes that interact with  $i$  to produce  $e$  (Assumption 2), it cannot be the case that some necessary interacting factor was absent. Hence,  $e$  must always follow from its generative cause  $i$  (when  $i$  is present)—unless for some inexplicable reason, the causal link between  $i$  and  $e$  is intrinsically indeterminate (see below). In summary, if all requisite assumptions are satisfied in the power PC theory, causal power should always be either 0 or 1. There is no need to use Equation 4 to derive causal power.

Let us now consider more complex situations in which multiple causes, say  $i$  and  $j$ , interact (Novick & Cheng, 2004). When we estimate conjunctive causal power, assumptions corresponding to Assumptions 2 and 4 are still required. Novick and Cheng (2004) stated, “. . . alternative causes influence  $e$  independently of  $i$ ,  $j$ , and their conjunction” (p. 471) and “unobserved causes in  $a$  could produce  $e$ , but not prevent it” (p. 463). Thus, our argument still holds in these situations. For example, if we have a complete, elaborate specification of how smoking leads to lung cancer (including what other factors interact with or prevent this process), then whenever this entire set of interacting causal factors is appropriately configured, the likelihood that lung cancer will occur should be 1. Again, we do not need to rely on equations derived by Novick and Cheng (2004) to estimate conjunctive causal power.

In the above argument, we specified one possibility in which causal power could be probabilistic—the situation in which a candidate cause (or conjunctive cause) fails to produce  $e$  for no reason at all (see the next section for apparent counterexamples). To concretely illustrate this, imagine that  $P(e|i) = 0.8$  and  $P(e|\sim i) = 0$ . In this situation, Equation 4 predicts that causal power should be estimated as 0.8. Now suppose a person were asked to estimate causal power. To do so, the person is asked to estimate  $P(e|i)$  in a counterfactual world where no alternative causes exist (see Buehner et al., 2003, for discussion of causal queries). Suppose this person gives a response of 0.8, matching the predictions of Equation 4. Such an estimate indicates that in a vacuum devoid of any other causal influences, the cause will lead to the effect 80% of the time. What happens on the other 20% occasions on which the cause fails to produce the effect? One might think that an alternative cause—whether known or unknown to the reasoner—prevented the effect from occurring on these occasions. However, because there are no alternative causes in this counterfactual world, this possibility is not allowed. Another interpretation might be that some factor interacted with the candidate cause so as to not produce the effect. This too relies on alternative causes and thus will not work as an explanation. Indeed, a causal power of 0.8 implies that in a vacuum, free of alternative causes, the candidate cause will fail to produce the effect for no reason at all. Could this be what Equation 4 is capturing? More critically, is it what people are estimating?

People have generally found the idea of inherently probabilistic causality to be implausible. Einstein famously denied the proposal of quantum mechanics because it required that the world, at its core, be probabilistic (Einstein, Podolsky, & Rosen, 1935). He

<sup>2</sup> For instance, in the first assumption, it is noted that when  $a$  occurs, it “produces”  $e$ .

claimed that quantum mechanics must be incomplete because god “does not play dice” (Hoffman, 1972, p. 193). If it was not until the development of modern physics that the notion of inherent randomness was taken seriously, it seems quite unlikely that this is part of laypeople’s conceptualization of causality.

Indeed, many psychologists have argued that indeterminism at the level of individual, direct causal links (like those evaluated by the power PC theory) is incompatible with most people’s worldview (e.g., Goldvarg & Johnson-Laird, 2001; Koslowski, 1996). Instead, people may treat seemingly indeterminate relationships as resulting from complex unobserved causal interactions. Smoking may cause lung cancer, but not everyone who smokes gets lung cancer. People may accept this by assuming that there is an array of unobserved and/or unknown causal influences in operation (e.g., genetics) that produces what looks like an indeterminate causal relationship. Some of these unknown and/or unobserved causal influences might be preventative (i.e., in violation of Assumption 4) and/or interact with the candidate cause (i.e., in violation of Assumption 2). Harré and Madden (1975) similarly argued that it is irrational to state that A causes B while agreeing to the possibility of observing A in the absence of B. They suggested two possibilities: (a) that such a situation should necessarily lead us to abandon the belief in the causal relation or (b) that “alternatively we can preserve the conceptual relation between the predicates by the claim that something had gone wrong in the aberrant case” (Harré & Madden, 1975, p. 9).

There are additional reasons why probabilistic causality is psychologically implausible. If this inherent randomness is what probabilistic causal powers capture, the power PC theory implies that the amount of this randomness present in causal relationships varies in a systematic way, such that the randomness results in a causal power of 0.8 for one causal link but 0.15 for another causal link, for instance. Moreover, this randomness inherent in a given causal relationship stays constant; it is an enduring property of the relationship. Finally, this randomness is not the result of other causal influences; probabilistic causal power implies that such indeterminism would occur in a vacuum. It is difficult to believe how such random behaviors could be enduring characteristics inherent to causal relations.

In summary, we have argued that if the assumptions required for Equation 4 to compute causal power are satisfied, then the result of Equation 4 should always be either 0 or 1 unless one subscribes to the notion of inherent indeterminism in the world. In this section, we argue that this kind of pure randomness is psychologically implausible. Thus, there is no need to derive causal powers using Equation 4 (or the even more complex equations needed for conjunctive causal powers; Novick & Cheng, 2004). They are always either 0 or 1.

### Apparent Counterexamples

Let us consider apparent counterexamples one might argue for in the situation we are considering. We discuss three cases in which causality appears to be indeterminate when all requisite assumptions are seemingly satisfied. Ultimately, each of these examples violates at least one of the conditions required to compute causal power.

The first possibility for probabilistic causality has to do with levels of activation. Consider an example of leaves being blown

off a tree. In this situation, different quantities of leaves may be blown off given a gust of wind, thus leading to a probabilistic outcome. This example can be treated in at least two different ways. First, the effect variable can be represented as a continuous variable (i.e., percentage of leaves blown off). Because the power PC theory does not currently handle continuous variables, this situation would fall outside the scope of the model. A second possible way to conceptualize this situation would be to consider each individual leaf as a possible effect. In this situation, the presence or absence of the effect and the cause are easier to classify. Each leaf could be observed and classified as to whether it had fallen. The question still remains as to why some leaves would fall and others would not, but, as we have argued, one would be forced to conclude that this variation would be due to preventative or interactive causes such as variations in the strength of attachment of leaves to the tree. That is, this situation, in fact, fails to meet the conditions required to compute causal power.

Second, if the relationship between the candidate cause and the target effect consists of multiple intermediate steps, the causal relation can look indeterminate when the intermediate steps fail. Suppose one hypothesizes that X causes Z. Nothing interacts with X in bringing about Z, and nothing prevents Z; Assumptions 2 and 4 are satisfied. Further suppose that the truth is that X causes Z only indirectly via Y (which might be thought of as the mechanism by which X influences Z). The relation between X and Z becomes indeterminate if there is a preventative cause of Y or some interacting cause necessary for Y to bring about Z. Thus, if X occurs but Y does not because of a preventative cause or interacting factors, Z would not occur either, making the relation between X and Z appear indeterminate. The problem with this situation is that X is not a cause of Z. Causal power is a quantity describing direct causal relations and thus cannot characterize the indirect influence of X on Z (Glymour, 2000). X appears to be a cause of Z because it is confounded with Y, which is a cause of Z (in violation of Assumption 5; i.e.,  $P(Y|X) \neq P(Y|\sim X)$ ). This fact should become evident when conditionalizing on the absence of Y (X and Z will no longer co-occur). Therefore, this apparent indeterminism occurs because of the mistaken belief about the true cause of Z and the subsequent violation of one of the required assumptions.

The third possibility concerns an incorrect categorization of the candidate cause (Cheng, personal communication, August 31, 2004; Lien & Cheng, 2000). Take the following example: Suppose 50 healthy participants in a hypothetical experiment drank from a glass of liquid that the experimenter prepared, and 50 healthy participants did not drink anything during this experiment. Of the 50 participants who drank, only 25 died immediately, and none of the participants who did not drink died immediately. In reality, of the 50 glasses the experimenter prepared, 25 contained plain water and 25 contained a clear odorless poison that is completely fatal without any antidote (i.e., Assumptions 2 and 4 are met). Yet, an outside observer, who has no way of detecting the different liquids (i.e., water vs. poison), might attempt to evaluate the causal power of drinking the contents of the glass with respect to death. If he or she was to do so, the result of Equation 4 would be 0.5, a seemingly indeterminate estimate of causal power in a fully deterministic situation.

The problem in this situation is that the miscategorization of the candidate cause has led to a confound. The true cause (drinking

poison) is confounded with the reasoner's candidate cause (drinking the contents of the glass). This is a violation of Assumption 5.<sup>3</sup> It should be evident that the result of Equation 4 is not an accurate estimate of causal power in this case. First, the causal power of drinking from a glass in this situation is intuitively 0 (drinking from the glass per se is not the cause of any of the 25 deaths). The reason it looks as though drinking from the glass is (probabilistically) sufficient for death is because it is confounded with the true cause (i.e., drinking poison). Second, the result of Equation 4 in this situation is not context-free (as causal power is; e.g., Cheng, 1997). If our reasoner was to go to a poison-free context, estimates of causal power of drinking from glasses would drop inexplicably to 0. Third, if the reasoner was given information about the presence and absence of poison in glasses, this factor would appear to interact with the candidate cause; death would occur after drinking when poison was present but not when poison was absent. Each of these points indicates that although miscategorizing the candidate cause can lead Equation 4 to produce probabilistic results, such results are normatively incorrect estimates of causal power (i.e., assumptions have been violated).

In summary, in each of these scenarios, assumptions required by the power PC theory are violated and lead to the appearance of indeterminism. Thus, these cases do not contradict the claim that determinism should necessarily lead Equation 4 to produce estimates of 0 or 1 when the necessary assumptions are met. Instead, each of these situations further illustrates how incomplete or incorrect knowledge can lead to the appearance of indeterminism.

### Are These Assumptions Met in the Real World?

So far, we have argued that if a reasoner's goal is to estimate causal powers (as claimed by the power PC theory), there is no need to derive causal powers through Equation 4 because they are either 0 or 1 when all the assumptions required for Equation 4 are satisfied. Let us now consider the possibility that causal powers can be probabilistic—for reasons that we cannot grasp—and there is a need to quantify causal powers and, hence, a need for an equation like Equation 4. This possibility brings up a question of whether Equation 4 computes causal power in the real world, that is, whether the assumptions necessary to apply Equation 4 are met in the real world. In this article, we focus on Assumptions 2 and 4 above.<sup>4</sup>

Accurate assessment of Assumptions 2 and 4 requires the reasoner to know something about the nature of *a* (i.e., how the composite of alternative causes influences the target effect, *e*) a priori. Yet, because *a* can be unobserved, it is unclear how this might occur. How would one know whether an unobserved factor had the ability to prevent the effect or that an unobserved factor interacted with the candidate cause to bring about the effect? If it is difficult to check whether these assumptions are satisfied in practice, is it the case that these assumptions are safe to make in the real world? Orthogonally, even if these assumptions are of questionable validity in the real world, is it the case that people make analogous assumptions of the world? We would argue that these assumptions are almost always violated in the real world and that people do not normally make these assumptions.

Consider Assumption 4, for example. Equation 4 requires that alternative causes (even those that are unobserved) be generative in nature. Otherwise, "there is no unique solution [for causal

power] in general" (Cheng, 2000, p. 239).<sup>5</sup> This is true even though one can apply these equations to either generative causes or inhibitory causes. This is also required when estimating conjunctive causal power (Novick & Cheng, 2004). Yet, it is not clear how one would guarantee the validity of this assumption. For instance, drinking coffee makes one alert unless there is some unknown pollutant in the air that causes people to feel drowsy, unless the person has a currently unspecified genetic predisposition to be resistant to the effect of caffeine, and so on.

Similar arguments can be made for Assumption 2. The validity of this assumption has been already disputed by Novick and Cheng (2004): "Most causes in the real world . . . are complex, involving a conjunction of factors acting in concert, rather than simple, involving a single factor acting alone" (p. 455). Because of this problem, Novick and Cheng (2004) provided six additional equations to compute the causal power of causes that work to jointly influence an effect (i.e., causes that interact, say *i* and *j*).

These equations may appear to solve the problems with Assumption 2 raised above, but they only push the dilemma back one step. Even when we compute conjunctive causal power of candidate causes *i* and *j*, assumptions that correspond to Assumptions 2 and 4 are still needed; it must be assumed that they do not interact with *a* (" . . . alternative causes influence *e* independently of *i*, *j*, and their conjunction"; Novick & Cheng, 2004, p. 471)<sup>6</sup> and *a* should not be preventative. Again, this requires a priori beliefs about (potentially unobserved) alternative causes. If we cannot be

<sup>3</sup> Cheng (personal communication, October 4, 2004) argues that because *i* and *a* are supposed to be disjoint sets, it would be incorrect to say that Assumption 5 is violated in our example. However, Novick and Cheng (2004) specifically stated that causes must be binary variables: "Candidate causes and effects are represented by binary variables or by other types of discrete variables that can be recoded into that form" (p. 455). "Drinking something" is one such binary variable. "Drinking poison" is yet another binary variable. For each of these variables there is a set of occurrences that makes the cause present. The sets of occurrences that render each of these causes present are certainly not disjoint, but this is not uncommon; some observations of weather render both "cold" and "windy" to be present, some observations of foods render "salty" and "rich," but that does not imply that these are not separable causal factors. Our example of drinking something versus drinking poison is a special case of overlap in which all the occurrences that render "drinking poison" present also render "drinking something" present (the former occurrences are a subset of the latter). This relationship implies that the former cause entails the latter cause (i.e.  $P(\text{drinking poison} | \text{drinking something}) < P(\text{drinking something} | \text{drinking poison}) = 1$ ), a violation of Assumption 5.

<sup>4</sup> Assumption 1 can be thought of as a description of the notations that the theory is using and thus is not discussed. Assumption 3 seems reasonable. Whether people accept Assumption 5 is an empirical question and is not discussed here (see Hagmayer & Waldmann, 2004; Luhmann & Ahn, 2003, for relevant findings).

<sup>5</sup> See the Contextual Power section for discussion of the solution provided by Cheng (2000) in such cases.

<sup>6</sup> ". . . an analogous assumption is logically required for some set of candidate causes and causes alternative to them as part of the procedure for inferring whether that set of candidates interact to influence *e*. Thus even departure from the independent influence assumption requires the use of this very type of assumption with regard to a different partitioning of the set of causes of *e*. In short, this assumption is necessary" (Cheng, 2000, p. 232).

assured that all necessary assumptions are met, then Novick and Cheng's (2004) equations cannot provide an estimate of causal power of the observed conjunctive causes. For instance, we might observe how  $i$  and  $j$  interact, but there might be other unobserved factors that interact with  $i$  and  $j$ .

To summarize this section, we discussed two requisite assumptions for the power PC theory; alternative causes may not interact with the cause being evaluated, and the alternative causes may not be preventative. We argued that these assumptions are problematic because they require a priori knowledge about the nature of alternative causes (both observable and unobservable ones) and they are unlikely to be valid in most (or arguably all) real-life situations. Take, for example, the extent to which a high-fat diet causes heart attacks. To use Equation 4 we must deny the possibility that factors such as general health, age, and genetics interact with diet to produce heart attacks. To the contrary, it seems more reasonable that a healthy 20-year-old's tendency to have a heart attack would be much less influenced by a high-fat diet than would be an overweight, older person's. Preventative causes such as exercise, medication, and genetics may also be operating. Thus, we cannot apply the power PC theory to compute causal power in this situation. It seems it would not be an exaggeration to state that all real-life situations are like this situation.

### Contextual Power

Cheng (2000) provided an analysis of what Equation 4 computes under various violations of the associated assumptions. When the assumptions associated with Equation 4 are violated, this equation computes what Cheng (2000) calls *contextual causal power*. Contextual causal power is not causal power. Instead, it is a measure of causal strength that is bound to the learning context. Unlike causal power, contextual causal power will be highly dependent on the frequency of occurrence and power of alternative causes in a given context. Cheng (2000; Novick & Cheng, 2004) argued that contextual causal power is still "useful" because it places constraints on the range of causal power. Let us call this position *contextual power theory*.

There are apparent contradictions between this contextual power theory and the power PC theory. Cheng (2000) stated,

In the reasoner's mind, causal powers are invariant properties of relations that allow the prediction of the consequences of actions regardless of the context in which the action occurs, with "context" being the background causes of an effect (those other than the candidate causes) that happen to occur in a situation. The reasoner's goal is to infer these powers. (p. 227)

However, the contextual power theory appears to state that a reasoner's goal is to simply compute something useful.

Furthermore, the power PC theory lays out several assumptions, each of which is a necessary condition for computing causal power from observations. Thus, if any one of them is violated, one should withhold his or her judgment (because the goal is to compute causal power). In contrast, the contextual power theory appears to state that when assumptions are violated, people would still compute something useful, namely, contextual power.

There is one possible solution to avoid these contradictions. One might assume that although a reasoner's goal is still to compute

causal power, people are not normally aware of violations of requisite assumptions. Such beliefs would lead reasoners attempting to compute causal power to unintentionally compute contextual power. Indeed, Cheng's (2000) descriptions of the contextual power theory strongly imply this unawareness-of-violations assumption. For instance, Cheng (2000) suggested that "reasoners would start with the simplest representation of the causal world, and add complications *only when necessitated by evidence*" (p. 232, italics added for emphasis). Similarly, Novick and Cheng (2004) stated, "reasoners would consider more complex models *when motivated by evidence* contradicting a simpler model" (p. 471, italics added for emphasis). The specific evidence they suggested using is experience in new contexts. If predictions stemming from causal power inferences fail in contexts other than that in which the learning took place, "their failure would suggest a violation of the independent influence assumption [or the assumption about preventative alternatives]" (Cheng, 2000, p. 238). All of these statements clearly imply that people cannot recognize violation of assumptions unless their predictions are tested in other contexts.

In contrast, we argue that people are very aware that violation of assumptions occurred.<sup>7</sup> We have maintained that reasoners believe in deterministic causal relations. This belief in deterministic causality necessarily entails awareness of violation of the assumptions whenever people observe indeterminism. That is, whenever a reasoner accidentally computes a contextual power value between 0 and 1, the reasoner acknowledges that there must have been an unobserved alternative cause that prevents the effect and/or an unobserved alternative cause that interacts with the candidate cause. The fact that a violation has occurred should become obvious.

Furthermore, we argue that people are well aware of the fact that they do not have complete knowledge of causal interactions in the world, and that is the reason why causal relations appear indeterminate. Indeed, careful examination of stimuli used in studies presented as supporting the power PC theory also makes us skeptical about the presupposition that people were not aware of violation of the assumptions even in controlled experimental situations. For instance, in Wu and Cheng (1999), one scenario involved testing the causal efficacy of a new drug in preventing cats from becoming fertile. It is difficult to believe that participants would have thought that nothing in this world could prevent the causal influence of this new drug or that there are no other events that would interact with causal efficacy of this new drug. Similarly, in Buehner et al. (2003), all stimulus materials involved the domain of medicine, where undergraduate participants almost certainly have preconceptions about alternative interacting and preventative causes. That is, it would be safe to argue that even in studies apparently supporting the power PC theory, participants may have been acutely aware of the violations of assumptions.

Therefore, we see tremendous difficulties in reconciling how the power PC theory and context power theory each suggest people deal with violations of Equation 4's assumptions. We have argued

<sup>7</sup> People may not be aware of specifically which assumptions are violated and how. All they may be aware of is that there must be some violation. This belief is sufficient for our argument that follows.

that the assumptions required for the power PC theory are almost always violated in the real world and, furthermore, that people must be aware that these violations take place. Therefore, if people's goal is to infer causal powers, they should say either 0 or 1 when they (accurately or inaccurately) believe the requisite assumptions are met, or they should withhold their judgments when they notice that the requisite assumptions are violated. Cheng (2000) argued that people can estimate contextual causal power, but the only way to make this argument compatible with the power PC theory's core argument is to assume that people mistakenly believed the requisite assumptions are met. We argued that people are actually aware that some assumptions are violated, which should lead them to withhold judgments according to the power PC theory.

### One Possible Alternative

There is a possibility that allows people to recognize violations of the assumptions associated with Equation 4 and still provide causal strength judgments. Instead of striving to compute causal power as the Power PC theory assumes (Cheng, 1997, 2000; see also Buehner et al., 2003), people might be attempting to intentionally compute a quantity like contextual causal power. If this were the case, not all assumptions behind Equation 4 might be irrelevant.

Pearl (2000) provided an account of how and why one might compute a quantity similar to contextual power.<sup>8</sup> He argued that when estimating the sufficiency of a cause, one can compute what he calls the *probability of sufficiency* (PS). The equation for PS is illustrated in Equation 5:

$$p_i = \frac{P(e|i) - P(e|\sim i)}{1 - P(e|\sim i)}. \quad (5)$$

This equation is identical to Equation 4. So, what sets PS apart from causal power? There are two important theoretical differences: the assumptions required and the quantities sought. PS requires only two assumptions. First, it requires that  $i$  be either generative or preventative with respect to  $e$  but not both. Second, it requires that  $i$  and  $e$  not be confounded. The first assumption seems relatively benign. Pearl (2000) stated that it "is often assumed in practice" (p. 291), citing the example of epidemiology in which it is assumed that exposure to a risk factor will not protect anyone from contraction. The second assumption is likely violated in the real world (but see Gopnik et al., 2004; Spellman, 1996), but we believe that people should agree to it when attempting to estimate causal strength. That is, most people should believe that a common cause behind two correlated factors would be reason enough not to attempt a causal strength estimate (Goodie, Williams, & Crooks, 2003; Spellman, 1996; Waldmann & Hagmayer, 2001). The second assumption is essentially Assumption 5 in the power PC theory. It is critical to note that PS does not require the two assumptions that we discussed in this article (Assumptions 2 and 4). Thus, as acknowledged by Cheng (2000) herself, PS requires, "weaker assumptions" (p. 247) than does the power PC theory.

The second difference between PS and the power PC theory is the quantity they are computing. The power PC theory attempts to

compute a context-free description of the sufficiency of a cause, whereas PS seeks to describe the sufficiency of a cause in the learning context. Counterfactually, PS can be described as the probability that intervening to make  $i$  present will lead to  $e$  being present given a situation in which both  $i$  and  $e$  are both initially absent. Another way of stating this is the probability with which  $i$  produces  $e$  (directly or indirectly, with or without help from alternative interactive factors) in the learning context. Thus, there is a major theoretical difference between what each of these theories suggests people are attempting to compute (i.e., differences at the computational level in Marr's, 1982, sense).

How does PS differ from causal power in practice? Under Pearl's (2000) proposal, one can speak of the PS of taking aspirin with respect to relieving headaches, the PS of laying off a portion of a company's workforce with respect to reaching profitability in 5 years, or the PS of smoking with respect to lung cancer. Because these cause-effect pairs appear to violate the assumptions required by Equation 4, causal power cannot be computed. PS, however (depending on whether you believe these events are not confounded), can be calculated. Thus, one advantage of computing a contextual estimate of causal strength such as PS is that it reduces the immediate requirements for computation. The downside is that the estimate is bound to the learning context. As Hume (1739/1987) argued, this may be the best that can be done.

Because the quantitative portions of PS and causal power are so similar, it remains possible that evidence previously thought to support the power PC theory may instead be evidence for Pearl's (2000) PS (or other models yet to be suggested). Just because people agree to give a judgment of causal strength in an experimental situation and their estimate happens to be consistent with Equation 4 does not mean that people believe that the power PC theory's requisite assumptions are satisfied or that they were attempting to estimate context-free causal power. They could be very much aware of violations of the assumptions (as we have argued in the earlier sections) and intentionally compute contextual causal power as in Pearl's PS.<sup>9</sup> Indeed, work exploring the conditions under which people withhold causal strength judgments is completely consistent with PS. Wu and Cheng (1999) found that people preferred not to interpret situations when the denominator of Equation 4 was 0. Both PS and the power PC theory predict such hesitance.

However, if what a reasoner initially set out to compute is PS rather than causal powers, there is one interesting testable prediction. If PS is correct, people should be less bothered by the assumptions that are required only by the Power PC theory. That is, although the power PC theory would predict that people should withhold causal strength judgments when the power PC theory's other assumptions (e.g., Assumptions 2 and 4) are violated, PS

<sup>8</sup> Obviously, it is an empirical question whether Pearl's (2000) theory described below is psychologically valid. Our intent of introducing this theory here is not to advocate this as a psychological theory but rather to illustrate the problems with the attempt to estimate causal powers.

<sup>9</sup> Of course people's estimates might be better described by yet another alternative causal induction model whose quantitative predictions overlap with the power PC theory and PS.

would predict that it is unlikely people would do so. Such an empirical difference additionally illustrates the importance of closely examining assumptions and the computational goal of models.

### Summary

Our goal throughout this article has been to carefully think through the claims made by the power PC theory. Cheng (1997) claimed to have presented

a theoretical solution to the problem of causal induction first posed by Hume more than two and a half centuries ago. Moreover, the fact that this theory provides a simple explanation for a diverse set of phenomena regarding human reasoning . . . suggests that it is the solution adopted biologically by humans. (p. 398)

We have argued that this is not the case for a number of reasons.

Because the power PC theory seeks to compute a context-free estimate of causal strength for individual, direct causal links, it requires rather stringent assumptions to be made. We have raised two issues with respect to these requisite assumptions. First, if they are all satisfied, it seems as though causal power is easily computable (i.e., power is either 0 or 1) unless people believe in inherent indeterminism, which we find highly unlikely. Second, even if people accept such a notion, they must still believe that the world works in a very restricted manner for the power PC theory to work. Alternative causes (especially those that are unobserved) must work in very special ways. They must be generative and may not interact with the candidate cause (or candidate causes in the case of Novick & Cheng's, 2004, equations). We argue that the world is a messy place and that idealized situations such as this are few and far between even in purportedly well-controlled experimental situations.

If these assumptions are violated, then causal power cannot be computed (according to the power PC theory). If one attempts to compute causal power anyway, the resulting quantity is not causal power. Instead, it is what Cheng (2000) referred to as contextual causal power. Contextual causal power is not the context-free property of individual causal links that the power PC theory suggests people seek (Cheng, 1997, 2000; see also Buehner et al., 2003). Instead, it is a context-dependent quantity that varies depending on the composition of the alternative causal influences.

Furthermore, to make the contextual power theory compatible with the power PC theory, one needs to assume that contextual power is what reasoners accidentally compute while attempting to compute causal power. We have argued that such errors should be readily apparent because any probabilistic causal powers should imply violations of assumptions, and thus the power PC theory should predict that people withhold judgments. Instead, we argue that people may actually intend to compute contextual causal power (or something like it). Given the apparent difficulty in computing context-free causal power, this more modest goal seems plausible.

The power PC theory has made an important theoretical advance by specifying the conditions under which causal powers can be inferred (i.e., the five requisite assumptions above). Ironically, our analysis shows that such conditions are so restrictive that causal powers cannot be computed in the real world. Furthermore, if we

add in the relatively benign idea of deterministic causality, it becomes more apparent that the derivations in the power PC theory are not needed in computing causal powers.

### References

- Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1119–1140.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.
- Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive causal power. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227–253). Cambridge, MA: MIT Press.
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545–567.
- Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, *47*, 777–780.
- Glymour, C. (2000). Bayes nets as psychological models. In F. Keil & R. Wilson (Eds.), *Cognition and explanation* (pp. 169–197). Cambridge, MA: MIT Press.
- Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610.
- Goodie, A. S., Williams, C. C., & Crooks, C. L. (2003). Controlling for causally relevant third variables. *Journal of General Psychology*, *130*, 415–430.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal naps and Bayes nets. *Psychological Review*, *111*, 3–32.
- Hagmayer, Y., & Waldmann, M. R. (2004). Seeing the unobservable—Inferring the probability and impact of hidden causes. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 523–528). Mahwah, NJ: Erlbaum.
- Harré, R., & Madden, E. H. (1975). *Causal power: A theory of natural necessity*. Totowa, NJ: Rowman and Littlefield.
- Hoffman, B. (1972). *Albert Einstein, creator and rebel*. New York: Viking Press.
- Hume, D. (1987). *A treatise of human nature* (2nd ed.). Oxford, England: Clarendon Press. (Original work published 1739)
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, *40*, 87–137.
- Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, *107*, 195–212.
- Luhmann, C. C., & Ahn, W. (2003). Evaluating the causal role of unobserved variables. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 734–739). Mahwah, NJ: Erlbaum.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, *111*, 455–485.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.
- 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. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). New York: Appleton-Century-Crofts.
- Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judg-

ments are made while controlling for alternative potential causes. *Psychological Science*, 7, 337–342.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, 82, 27–58.

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from

statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, 10, 92–97.

Received August 2, 2004

Revision received October 29, 2004

Accepted November 1, 2004 ■

### Postscript: Abandonment of Causal Power

Christian C. Luhmann  
Vanderbilt University

Woo-kyoung Ahn  
Yale University

What does it mean when causal power is greater than 0 but less than 1? Cheng and Novick (2005) argue that when a reasoner represents a potential cause at an inappropriately high level (e.g., citrus fruit instead of the true cause, oranges), true causal power can be between 0 and 1 even when all the required assumptions still hold. In our comment, we argued that such a case involves confounding (properties of oranges are confounded with properties of citrus fruit<sup>1</sup>) and thus causal power cannot be computed according to the power PC theory. Cheng and Novick (2005) disagree and include an illustration that they claim demonstrates the contradiction inherent in our claim. In their example,  $P(eli)$ , or  $P(\text{repelling bugs}|\text{citrus fruit})$ , is 0.5 (three oranges repel bugs, three lemons do not), but when re-expressed as  $q_i + P(ali) \times q_a - q_i \times P(ali) \times q_a$ , they claim  $P(eli)$  should be  $0.5 + 0.5 \times 1 - 0.5 \times 0.5 \times 1 = 0.75$ , which does not match the observed value. Note that this expression uses a value of 0.5 for  $q_i$  (i.e., causal power of citrus fruit) on the basis of application of Equation 5 (Cheng & Novick, 2005, p. 701). From our perspective, this estimate is erroneous because the situation is confounded, and thus Equation 5 cannot be used.<sup>2</sup>

Another way to explain that the value of 0.5 for  $q_i$  of citrus fruit is not a normatively accurate estimate of causal power is the following. Note that the probabilistic causal power estimate in this case is equivalent to the ratio between the frequency of true cause (e.g., oranges) and the frequency of the broader category (e.g., citrus fruits).<sup>3</sup> It should be clear then that these probabilities are not “invariant properties of relations” (Cheng, 2000, p. 227) because they could easily vary (e.g., there is no natural law that constrains the proportion of oranges in citrus fruits). If we travel to a new context in which citrus fruits are common but oranges are rare, the estimated causal power of citrus fruit will change. This test (suggested in Cheng, 2000) indicates that the estimate of the causal power of citrus fruit (i.e., 0.5 in the above illustration) is not context-free and thus an inaccurate estimate of causal power.<sup>4</sup> Thus, this example fails to show that “a probabilistic causal power can be obtained when all of the power PC assumptions are met if candidate cause  $c$  is an imperfect hypothesis, even for a reasoner who assumes causal determinism” (Cheng & Novick, 2005, p. 701). That is, our claim that incorrectly categorized causes violate the no-confounding assumption remains valid.

These difficulties associated with confounds imply that accurate computation of causal power requires a tremendous amount of

accurate knowledge, much of which reasoners are unlikely to possess (Cheng & Novick, 2005). We agree that this poses a problem for accurately judging causal power and that such situations are yet another obstacle to valid causal inferences (including the successful computation of causal power). Indeed, one could argue that “no causal inference should ever occur” (Cheng & Novick, 2005, p. 702). Therefore, the fact that people are not paralyzed in their causal inferences can actually be taken as evidence against the power PC theory itself (or any other model that requires equally stringent assumptions).

Beyond the problem of confounding, we have also argued that each of the assumptions required to compute causal power is difficult to obtain (Luhmann & Ahn, 2005). To deal with these difficulties, Cheng and Novick’s (2005) reply heavily emphasizes the claim that contextualized causal power (Cheng, 2000) is consistent with the power PC theory. Yet, Cheng and Novick still argue that during causal judgments, one possible goal of reasoners (perhaps a particularly important goal) is to compute causal power. They state, “aiming for causal power and accepting contextual power is as ‘contradictory’ as aiming for a gold medal and accepting silver” (Cheng & Novick, 2005, p. 703), implying that causal power (i.e., the gold medal) is the reasoner’s ultimate goal. How-

<sup>1</sup> To illustrate, we refer to the “conjunction of substances in dried orange peel [that] deterministically repels beetles” (Cheng & Novick, 2005, p. 701) as RB. The property RB is confounded with properties of the broader category, citrus rinds. For example, the probability of RB is higher in the presence of leathery, bitter, aromatic rinds than in the absence of these properties.

<sup>2</sup> Instead of using Equation 5,  $q_i$  can be derived as follows. Because  $P(eli) = 0.5$ ,  $P(eli) = q_i + 0.5 \times 1 - q_i \times 0.5 \times 1 = 0.5$ . Simplifying this expression,  $q_i = 0.0$ . This value makes sense given that this example is set up such that “some conjunction of substances in dried orange peel deterministically repels beetles, and no other fruit peel . . . has this effect” (Cheng & Novick, 2005, p. 701). Note that this derivation is only possible from the omniscient perspective in which we already know  $P(ali)$  and the true cause. Reasoners would not have such knowledge, but the derivation is presented simply to further illustrate why the value of  $q_i$  used in Cheng and Novick (2005) is incorrect.

<sup>3</sup> More generally speaking, when causes are inappropriately represented (e.g., true cause is a subset of the candidate cause or the true cause overlaps with the candidate cause), causal power estimates, if calculated as in Cheng and Novick (2005), will be equivalent to the conditional probability of true cause (i.e., oranges) given the hypothesized causal candidate (i.e., citrus fruits).

<sup>4</sup> Again, we are not implying that a reasoner must perform such a test. The fact that the estimated value is neither 0 nor 1 is sufficient to inform the reasoner that some error or errors have taken place even though they might not know the nature of the error(s). The test is presented here from an omniscient perspective to illustrate why this is not a context-free causal power.