Logical Consistency and Objectivity in Causal Learning

Patricia W. Cheng (cheng@lifesci.ucla.edu)
Department of Psychology, UCLA

Mimi Liljeholm (mlil@caltech.edu)
Computation and Neural Systems Program, California Institute of Technology

Catherine M. Sandhofer (sandof@psych.ucla.edu)
Department of Psychology, UCLA

Abstract

Logical consistency and objectivity are cornerstones of science that distinguish it from cult and dogma. Scientists' concern with objectivity has led to the dominance of *associative* statistics, which define the basic concept of independence on observations. The same concern with avoiding subjective beliefs has led many scientific journals to favor frequentist over Bayesian statistics. Our analysis here reveals that to infer causes of a binary outcome, (1) the associative definition of independence results in a logical inconsistency -- even for data from an ideal experiment -- for both frequentist and Bayesian statistics, and (2) removing the logical error requires defining independence on counterfactual *causal* events. The logically coherent causal definition is the one intuitively adopted by humans. Our findings have direct implications for more consistent and generalizable causal discoveries in medicine and other sciences.

Keywords: Causal inference, rationality, cognition, statistics.

Introduction

Whenever we humans or other animals apply causal knowledge to achieve a desired outcome, we implicitly assume that the future resembles the past. Without the assumption that the course of nature remains invariant, all experience becomes useless (Hume, 1739). But what is the course of nature if not change (e.g., seeds sprout, species evolve, oceans warm, stars implode)? What we assume to remain invariant in nature are -- instead of events -- the forces of change, namely, causation (Kant, 1781; Kitcher, 1995). The fact that we routinely base actions on our causal knowledge (e.g., I strike this match because I expect it to ignite) is indubitable evidence that we hold the *causal invariance* assumption across the learning and application contexts. The present paper examines a previously unsuspected role that this assumption should play in scientific causal inference, leading to implications for more rational evaluations of hypotheses regarding causes of a binary outcome (e.g., a student graduating or not, an organism being dead or alive).

To test causal hypotheses based on data from experiments or quasi-experiments, the statistics in typical scientific use define invariance (often termed "independence" or "no interaction") on observations (Fienberg, 1980/2007; Jaynes, 2003; Wickens, 1989). Objectivity would seem to dictate this definition, given that the input necessarily consists of observations only. Because causal relations are inferred and inherently unobservable (Hume, 1739), defining invariance on causal relations seems objectionable.

Thus, for the respective purposes of scientific causal discovery and of justifying the application of causal knowledge, there are two distinct definitions of invariance: the associative and the causal. The *associative* conception defined on observable events traces its inspiration to the philosophical works of Hume (1739), who questioned the grounds for our compelling belief that causation exists in the world. The *causal* conception defined on causal influences rests on Kant's (1781) argument that an ontological commitment to causation is essential for a coherent interpretation of the world. We use "causal influence" here in the sense of *capacity* or *power*, which when realized explains the occurrence of the outcome.

There is a discrepancy between the two conceptions, but the discrepancy has not seemed problematic: The unspoken consensus is that while causal invariance justifies generalization, it plays no role in causal discovery. Accordingly, using associative statistics to test experimental data is standard practice, and is viewed as appropriate as long as the experimental manipulation, which disambiguates causal direction, succeeds in eliminating confounding.

The consensus opinion, however, is mistaken. Here we show that even in the ideal case in which there are no confounding variables, the definition of invariance incorporated in a measure can affect the statistical output. Moreover, with regard to causes of a binary outcome, a type of outcome prevalent in medical and business research (e.g., a tumor cell being malignant or not, a consumer buying an item or not), *only* the definition based on counterfactual causal events, the Kantian *causal power* definition, is logically consistent. Notably, the coherent definition is the seemingly less objective one.

To explain the inferential problem, we step back and examine the nature and definitions of causal invariance from a cognitive-science perspective, in particular, within the broader issue of how an intelligent agent with access to inherently limited information can construct a representation of the world that best enables desired outcomes. From this perspective we examine the implications of conceptions of causal invariance for the experimental sciences and everyday causal inference.

Causal Invariance and its Implications

Under the premise that all changes are caused, one way of stating causal invariance is: a cause *c* of an effect *e* retains the same capacity to affect *e* regardless of the temporal or spatial *context*, in which alternative (often unobserved) causes of *e* may occur with different probabilities. That is, the causal power of *c* is *independent* of the occurrence of alternative causes of *e,* as if those alternative causes were not there. A change in the capacity of a cause to produce its effect is an indication of the causal mechanism operating differently. As we show later, this interpretation of causal invariance applying the concept of *independence* (i.e., "no interaction") to causal powers enables logical implications of the assumption to emerge, by enabling a mathematical definition of causal invariance (Eq. 4). (Causal invariance is the simpler of two conceptions that are equivalent with respect to our conclusions; the other conception is that although *c* interacts with enabling conditions in the background, the enabling conditions occur just as frequently in the learning and application contexts [Cheng, 2000].)

The concept of causal invariance serves two distinct functions. First, as a working hypothesis, a defeasible default assumption, it justifies causal generalization and prediction. By rendering inference *compositional*, it enables the generation of logically consistent answers to an unlimited variety of questions regarding an outcome's occurrence in an unlimited range of application contexts (Cheng et al., 2007). Second, as a definition of what it *means* for the nature of a cause to remain the same (rather than as a description of a particular causal mechanism), causal invariance serves as a criterion for hypothesis revision. Thus, if a generalization proves wrong, as would often happen in the dynamic mental construction of our complex causal world, the deviation from expectation signals a need to better capture invariance. In this second role, causal invariance is a navigation device that orients hypothesis testing towards its goal of formulating the simplest explanation of a phenomenon that allows invariance to obtain (Carroll & Cheng, 2010).

Consider the alternative, the non-uniformity of nature, as the default. Not only would predictions and applications be impossible, so would hypothesis revision -- given no expectation, there is no deviation from expectation to guide revision towards causal invariance. Thus, the choice is a) inapplicable and stagnant causal knowledge or b) risky causal inference with the potential for effective generalization and hypothesis revision. In its two roles, as a default and a criterion for revision, causal invariance embodies the conviction that the world is knowable, that one can tease things apart, comprehend them, and mentally recompose them at will.

Defining Causal Invariance: Hume versus Kant

Assuming causal invariance requires two leaps of faith. The first is apparent: faith that the future resembles the past. The second is subtler: faith in the existence of causation, a faith Hume (1739) resisted. Here we show why the second leap plays a central role in rational causal discovery, in particular, why an associative definition of invariance, omitting this leap, is irrational for causal discovery.

We classify models as *causal* or *associative* depending on whether or not they have a definition of independence on causal influences. Whereas the ontological commitment to the existence of causation under the causal view enables this view to define independence on causal influences (e.g., Cartwright, 1989; Cheng, 1997; Lu, Yuille, Liljeholm, Cheng & Holyoak, 2008; Pearl, 2000; Sheps, 1958; Sloman, 2005; Yuille & Lu, 2008), the lack of this *a priori* assumption confines the *associative* view to defining independence on observations only (e.g., the cross product ratio; Fienberg, 1980/2007; Wickens, 1989). These two views differ most clearly for causes and effects that are represented by binary variables with a "present" value and an "absent" value; our argument therefore uses this variable type. For this variable type, observable events consist of the values of candidate cause *c* and of effect *e*. We denote the "present" and "absent" values by "1" and "0" respectively.

The Associative View The associative view defines independence on observations of *c* and *e* (we use *c* and *e* as variables or values depending on context): if *c* occurs independently of *e*, then

$$P(c=1, e=1) = P(c=1) \bullet P(e=1) \qquad \text{(Eq. 1)},$$

where $P(c=1, e=1)$ is the probability of the joint occurrence of *c* and *e*. This view computes associations, and leaves causal inference to a separate and subsequent interpretation of the associative output, for example, according to scientists' principles of experimental design or as Hume's "habit of mind". To enable predictions, this view typically amends Eq. 1 with additional assumptions, often variations of *linearity* or *additivity*. This amendment implicitly extends the definition of independence; deviation from linearity is what signals the need for interaction terms.

We illustrate the linear combination of associative strengths with the *ΔP model* (Jenkins & Ward, 1965; Salmon, 1965),

$$\Delta P = P(e=1 \mid c=1) - P(e=1 \mid c=0) \qquad \text{(Eq. 2)},$$

where $P(e=1 \mid c=1)$ and $P(e=1 \mid c=0)$, respectively, denote the probability that *e* occurs given that *c* occurs and given that *c* does not occur. Eq. 1 is a special case of Eq. 2, the case in which $\Delta P=0$. To tease apart the influence of *c* from all other influences on *e*, our analysis partitions all direct causes of *e* into *c* and *a*, where *a* represents a composite of *alternative* causes of *e* in the context. When *c* is absent, the occurrence of *e* is explained by *a*. Let w_c represent the *weight* (i.e., strength) of the association between *c* and *e*, and w_a represent that between *a* and *e*. ΔP has been shown to be a maximum-likelihood estimator of w_c in the Bayesian framework (Griffiths & Tenenbaum, 2009; Tenenbaum & Griffiths, 2001).

When there is *no confounding* (i.e., *a* occurs just as often whether or not *c* occurs), ΔP estimates w_c. Thus, replacing ΔP with w_c and $P(e=1 \mid c=0)$ with w_a, Eq. 2 can be rewritten and rearranged to give the linear equation:

$$P(e=1 \mid c=1) = w_c + w_a \quad \text{(Eq. 3)}.$$

That is, when multiple causes are present, the occurrence of *e* according to this model is explained by a sum of the associative strengths of the causes. Bayesian structure-learning models can likewise adopt the linear definition (Lu et al., 2008; Yuille & Lu, 2008; Tenenbaum & Griffiths, 2001).

Similarly, *generalized linear models* (*GLMs* [Fienberg, 2007; McCullagh & Nelder, 1989]), some process models in psychology (e.g., Rescorla & Wagner, 1972), and prominent causal models in epidemiology (Rothman et al., 2008) also adopt the definition in Eq. 1 amended with variants of linearity. For example, logistic regression, likely the most commonly used model for evaluating causal hypotheses in medical research and widely used in business research as well, amends Eq. 1 with a logistic scale transformation to better justify the linearity. A feature common across the generalizations in GLMs is "the presence in all the models of a *linear predictor* based on a linear combination of explanatory or stimulus variables" (McCullagh & Nelder, 1989, p. xvi).

Now, consider a situation in which representation in terms of observable events alone cannot capture the constancy of a causal relation across contexts. When effect *e* is binary, a factor's capacity to influence *e* may have no observable manifestations, even when there is no confounding. Suppose *c* is a cause of *e* that does not interact with any other cause of *e*. Yet, whenever *e* is already present (regardless of which other cause produced it), introducing *c* will yield no change in the state of *e*, indistinguishable from introducing a noncausal factor. For example, suppose someone is already dead (the binary outcome in question) from being hit by a car. Being hit by another car will show no change in the outcome (the person is still dead), despite the sameness of the forces underlying car accidents (the second car would have killed the person too). In such *occlusion* events, unobservable causal capacities lose their mapping onto observable changes. Given the lack of constancy in this mapping, postulating capacities becomes crucial for representing a stable causal world; observable changes, as used in associative models, or even actual causation in an episode, as used in epidemiological causal models (Rothman, Greenland & Lash, 2008), would be inadequate. Just as objects occluded in the 2-d visual input on our retinas are assumed to continue to exist in the world, so should occluded causal capacities.

The Causal View The causal view builds on Hume's insight – that causal knowledge is induced from noncausal data – but goes beyond it: Intervening between the observable input and the causal output is a causal explanation of the data. This explanation, under Kant's domain-general *a priori* causal framework, posits the existence of such things as causal relations: theoretical events that yield observed phenomena. We denote "causing" by "→" (e.g., "c→e" denotes "*c* causing *e*"). Once causal events are assumed to exist, the definition of their independence analytically follows:

if $c \rightarrow e$ is independent of $a \rightarrow e$, then

$$P(c \rightarrow e, a \rightarrow e) = P(c \rightarrow e) \bullet P(a \rightarrow e) \quad \text{(Eq. 4)}.$$

$P(c \rightarrow e)$ is the probability of *c* causing *e*; it corresponds to the theoretical probability that *e* would occur if *c* is present but no other (observed or unobserved, generative or preventive) cause of *e* were present. The probability is theoretical because it is impossible to know that a context has no unobserved causes. Note that P(c→e) is not a conditional probability involving two random variables, but instead the probability associated with a single random variable. Likewise, $P(a \rightarrow e)$ is the probability of *a* causing *e*, and $P(c \rightarrow e, a \rightarrow e)$ is the probability of one of the two causes, *c* or *a*, producing *e and* the other cause also producing *e* if *e* had not been already produced. ("No interaction" between the occurrences of *c* and *e*, as defined in Eq.1, is a special case of the independence of causal powers as defined in Eq. 4 when there is no confounding and ΔP=0.)

Notice that the definition in Eq. 4 centers on conjunctive causation in an "occlusion" event. The conjunctive causal event (e.g., a dead car-accident victim being killed by a second car) can never occur (rather than happen to not have occurred). Our "→" notation serves as a reminder that the causal events denoted are nonexistent and theoretical.

Although none of the events in Eq. 4 is observable, the intervening causal explanation of the data (e.g., when *e* occurred in the presence of *c*, it occurred *because c* caused it or *a* caused it) maps observable event frequencies (e.g., how often *e* occurred when *c* was present) onto their theoretical causal probabilities [e.g., $P(c \rightarrow e$ OR $a \rightarrow e)$]. Thus, $P(e=1|c=1)$ estimates $P(c \rightarrow e$ OR $a \rightarrow e)$. The latter in turn can be expressed in terms of the constituent events in Eq. 4:

$P(c \rightarrow e$ OR $a \rightarrow e) = P(c \rightarrow e) + P(a \rightarrow e) - P(c \rightarrow e, a \rightarrow e)$ (Eq. 5), where the final term equals the product, $P(c \rightarrow e) \bullet P(a \rightarrow e)$, if *c* and *a* produce *e* independently (Eq. 4).

Under this view, causal interpretation is integral to the computation of the numerical output (e.g., Cheng, 1997; Griffiths & Tenenbaum, 2009), rather than subsequent to it. Data analysis incorporates causal invariance.

The difference between the two views and its implications for rational scientific causal inference has not received attention. Like frequentist statistics for the experimental sciences, causal Bayes nets adopt the separation of statistics and causal inference. The "generic" parameterization most commonly adopted in causal Bayes nets uses neither the associative nor the causal definition, and the "noisy OR" parameterization in Eq. 5 is used for efficiency rather than rationality. In a similar vein, Bayesian causal models allow both the associative and causal definitions (Griffiths & Tenenbaum, 2009; Lu et al., 2008; Yuille & Lu, 2008).

The Rationality of the Two Views Is it rational to define causal invariance on unobservable, imaginary events, as the causal view does? Ceteris paribus, it is objectionable to use unobservable events. What is at stake, however, is logical consistency. What it means for the nature of a cause in our physical world to remain invariant across contexts is non-arbitrary. There is only one way for a causal mechanism in

a coherent world to operate the same way, without change. For binary causes and effects that are "present" or "absent," Eqs. 4 and 5 specify the only logically consistent definition of causal invariance (e.g., so that *c* causes *e* with indeed the same probability in one context as in another). In other words, systematic deviation from independence as specified in these equations indicates causal interaction. (Note that for other variable types and combinations of variable types, the singular meaning of causal invariance in the world is captured by other mathematical functions.)

We first explain the correlated influences inherent in associative amendments by illustrating how the linear model in Eq. 3 deviates from causal invariance. The additivity in Eq. 3 holds only if the capacities of *c* and of *a* to cause *e* are mutually exclusive [i.e., P(*c*→*e*, *a*→*e*) = 0; there are no occlusion events]. But, to define independence as mutual exclusivity (i.e., to define "no correlation" as a negative correlation) is self-contradictory.

To see the self-contradiction without the abstraction of causal inference, consider a simple concrete example involving two events regarding a deck of playing cards: drawing a diamond and drawing a face card. (Assume that the deck has diamonds and face cards, among other cards.) Defining independence between the two events as mutual exclusivity of the events would entail asserting that the chance of drawing a face card is the *same* for diamonds as for other suits *if* and *only if* face cards and diamonds are mutually exclusive: when there are no face cards that are diamonds. The chance of drawing a face card then would be 0 for diamonds but not for other suits. The mutual-exclusivity definition therefore implies a logical contradiction: "the chance of drawing a face card is the *same* across suits only if it is *not* the same across suits."

Our analysis so far may seem irrelevant to current frequentist statistics: nonlinear GLMs, which avoid a logical shortcoming of linear models for analyzing data with binary outcome variables, have long replaced linear models for that purpose (Fienberg, 1980/2007; Wickens, 1989). But, GLMs in fact do not sidestep the contradiction in other associative models. First, GLMs concur with the ΔP model in adopting the mutual-exclusivity definition for special cases involving data that have the feature of symmetry. We illustrate this agreement presently with a logistic-regression analysis of fictitious data in a story in an experiment designed for preschool children. Second, GLMs more generally carry the broader contradiction of defining independence as interaction. Because P(*e* =1|*c* =1) estimates P(*c*→*e* OR *a*→*e*), Eqs. 3 and 5 can be directly compared. They differ by the final (negative) term in Eq. 5 being omitted in Eq. 3. A scale transformation that would avert the contradiction would therefore need to result in subtracting the product, $w_c \bullet w_a$, from the right-hand-side of Eq. 3. But this is neither the intent nor the result of the transformations in GLMs. The logistic function, for example, is symmetric (see s-shaped curve in Figure 1), as is characteristic of associative models. In contrast, for every value of w_a, subtracting $w_c \bullet w_a$ from the sum, $w_c + w_a$, yields an asymmetric concave function of w_c (as w_c increases, an increasing amount is subtracted from the linear sum).

Without the *a priori* postulate that causal relations exist, associative models cannot coherently define independence on the missing relations, hence cannot justify the application of causal knowledge. They cannot, even when ideal experiments are concerned, because the error is logical.

An Illustration of the Associative and Causal Views Arriving at Opposite Conclusions We return to the mutual-exclusivity definition of causal invariance in associative statistics. In a story presented to preschoolers in our experiment, two brothers -- a farmer and a zookeeper – try to figure out what prevents red dots from appearing on the faces of animals at their farm and at zoo. The candidate preventive causes of red dots are two treats: a grain and a type of leaves. At the farm, the brothers gave the grain to all 10 animals there: 9 of them had red dots before eating the grain, and 6 did so afterwards. At the zoo, the brothers gave both treats -- grain and leaves -- to all 10 animals there: 4 of them had red dots before eating the two treats, and only one had red dots afterwards. The question is: which treat is one's best bet for removing red dots from the faces of farm and zoo animals?

Regardless of how "sameness" is defined, the rationale underlying the choice is: Assuming the grain operates "the same way" across contexts (i.e., farm and zoo), if the influence of the *intervention* (grain at farm vs. both treats at zoo) remains invariant across contexts, one's best guess would be that leaves had no effect – grain alone would already explain the outcome. But, if the influence of the intervention varies across contexts, one would attribute the difference to leaves.

According to the causal view, the grain operating with the same causal mechanism across contexts implies that *for every animal* (all 20), grain has the same causal power to remove red dots. We denote the two interventions by "*farm_iv*" and "*zoo_iv*" respectively and "red dots on the face" by "red" in the calculations below. The *causal power* of candidate cause *c* to prevent effect *e*, p_c, is estimated according to (Cheng, 1997):

$$p_c = \frac{P(e=1 \mid c=0) - P(e=1 \mid c=1)}{P(e=1 \mid c=0)} \qquad \text{(Eq. 6)}$$

Thus,

$$p_{farm\_iv} = p_{grain} = \frac{9/10 - 6/10}{9/10} = 1/3 \qquad \text{(Eq. 7)}$$

Likewise,

$$p_{zoo\_iv} = \frac{4/10 - 1/10}{4/10} = 3/4 \qquad \text{(Eq. 8)}$$

But, according to causal invariance (Eqs. 4 and 5),

$$p_{zoo\_iv} = p_{grain} + p_{leaves} - p_{grain} \cdot p_{leaves} \qquad \text{(Eq. 9)}$$

It follows that

$$3/4 = 1/3 + p_{leaves} - 1/3 \cdot p_{leaves} \qquad \text{(Eq. 10)}$$

Therefore, $p_{leaves} = 5/8$. Because 5/8 is greater than 1/3 (i.e., the leaves treat is a stronger cure than grain), the causal view prescribes choosing leaves.

Associative models, whether Bayesian or frequentist, all reach the opposite conclusion, prescribing grain instead. The mutual-exclusivity definition implies that the set of animals with "no red dots" *due to* grain, 3 out of 10 animals, has no overlap with the set *due to* the contextual cause at each place: grain should therefore heal 3 animals both at the farm and at the zoo. Because 3 animals indeed had their red dots "go away" at each place, leaves must have no effect. The ΔP model therefore prescribes grain.

Logistic regression is a GLM used for predicting the probability of the occurrence of a dichotomous outcome (e.g., red_dots vs no red_dots) by fitting data to a logistic function of a linear combination of input variables (e.g., grain, leaves, background causes at the farm and at the zoo). For the farm-and-zoo scenario (see Figure 1), because the pattern of events is symmetrical around the probability of .5, the same *reduction* in P(red_dots) occurs at the farm and at the zoo (see vertical dashed lines) at symmetrical segments of the logistic curve. Therefore, the grain (see heavy horizontal dashed lines) -- which explains the reduction in the probability of animals with red dots at the farm – explains the entire reduction at the zoo as well. That is, logistic regression detects no influence at all from leaves, either by itself or in an interaction, concurring with the ΔP model. Increasing sample size does not change this conclusion.

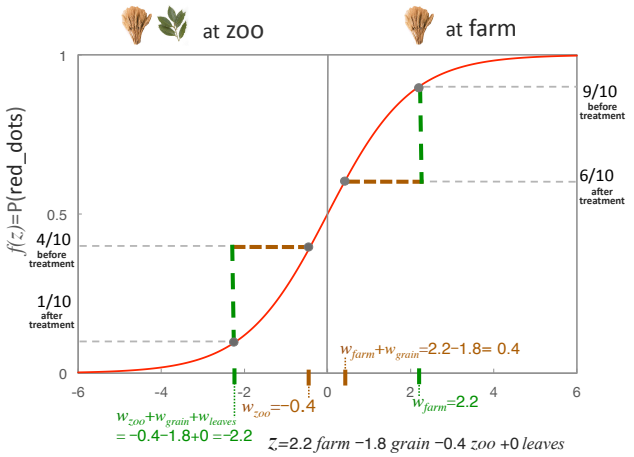

Figure 1. A schematic explanation of the probability of the outcome according to logistic regression: the probability of having red dots at the farm and at the zoo, before and after the respective interventions in the scenario, as a logistic function of the weighted sum of the four predictor variables.

Preschoolers in our experiment chose leaves, in accordance with the causal view. Recall that the causal view avoids the incoherence of the associative view by defining causal invariance on counterfactual causal events. Note that the causal explanations involve no prior domain-specific knowledge; the causal-invariance assumption is domain-general and the input consists of data alone. This view thereby achieves objectivity without sacrificing coherence.

If the world happens to be causal, then a leap of faith to assume unobservable causal capacities would be adaptive, by enabling a coherent definition of causal invariance in our representation. Coherence is essential because there are infinitely many possible representations of the world based on available observations, only some of which support generalization to new contexts. Reasoners use logical consistency and, more generally, parsimony of the represented explanations to prune the vast search space and efficiently converge on truth, if truth exists (Kelly, 2007). Causal discovery should therefore require general-purpose Sherlock Holmeses, who make use of coherence to infer how things work.

Discussion

In summary, noting a simple logical consequence of Kant's *a priori* assumption of causation for rational causal inference, we have shown that -- contrary to the unspoken consensus among scientists -- the causal invariance assumption critically affects causal discovery. To evaluate a causal relation involving a binary outcome variable that is "present" or "absent", only invariance defined on causal capacities is logically consistent and supports generalization to new contexts. Thus, associative statistics, for which invariance is only defined on observations, may arrive at a fallacious conclusion even when applied to data from a perfect experiment.

The potential for the associative and causal views to arrive at opposing recommendations has obvious implications. For example, a critical linear-regression analysis in the influential Seven Countries Study (Keys, 1980), a large longitudinal study on how diet affects coronary heart disease and other health outcomes, shows that controlling for saturated fat, consumption of sugar is unrelated to death (a binary outcome). Medical and public-health dietary advice in the U.S. based on this and other analyses in the study (Keys et al., 1984; Menotti et al., 2003), using linear models as was common practice, has created a food industry that produces low-fat but high-sugar foods (e.g., fat-free salad dressings with added sugar to compensate for taste). More generally, these associative analyses formed the foundation for three decades of dietary advice to adhere to a low-fat diet, without special attention to sugar intake (as distinct from caloric content). There is currently no causal analogue of logistic regression, which allows predictor variables that are continuous (e.g., consumption of sugar) as well as discrete. As we have shown for binary outcome variables, coherent causal generalization requires a causal framework, and applying causal instead of associative statistics to evaluate the influences of fat and sugar intake could potentially reverse estimates of the magnitude of their harm or change their assessed causal status. The researchers could have found that consumption of sugar causes coronary heart disease, diabetes, cancer and other diseases constituting the metabolic syndrome, as recent evidence indeed suggests. A more rational statistical approach could have profoundly altered the course of the obesity epidemic in the U.S. and worldwide.

Note that one interpretation of associative models that would remove the incoherence we noted is to posit a mediating continuous variable and to assume that the causes operate independently on this continuous variable rather than on the observed binary outcome variable. The linear definition of causal invariance holds for continuous outcome variables, thereby removing the incoherence. Regardless of the plausibility of the revised hypothesis with the mediating variable, note that it is *deviation from* the criterion of causal invariance that signals the need to revise the simple hypothesis (Carroll & Cheng, 2010), the tenacious goal being to achieve causal invariance.

Acknowledgments

The research reported in this article was supported by AFOSR FA 9550-08-1-0489. We thank Chris Carroll, Tom Wickens, and Ying Nian Wu for helpful discussion.

References

Carroll CD, Cheng PW (2010) The induction of hidden causes: Causal mediation and violations of independent causal influence. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds Ohlsson S, Catrambone R (Austin, Texas: Cognitive Science Society), pp. 913-918.

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

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

Cheng PW (2000) Causality in the mind: Estimating contextual and conjunctive causal power. In *Explanation and Cognition*, eds Keil F, Wilson R. (MIT Press, Cambridge), pp. 227-253.

Cheng PW, Novick LR, Liljeholm M, Ford C (2007) Explaining four psychological asymmetries in causal reasoning: Implications of causal assumptions for coherence. In *Topics in contemporary philosophy, Volume 4: Causation and explanation*, ed O'Rourke M (MIT Press, Cambridge), pp. 1 – 32.

Fienberg SE (1980/2007) *The analysis of cross-classified categorical data* (2nd edition (MIT Press, Cambridge; Spring, New York).

Griffiths TL, Tenenbaum JB (2009) Theory-based causal induction. *Psychological Review*, *116* (*4*): 661-716.

Hume D (1739/1987) *A treatise of human nature* (2nd edition, Clarendon Press, Oxford).

Jaynes ET (2003) *Probability theory: The logic of science.* (Cambridge Univ Press, Cambridge).

Jenkins HM, Ward WC (1965) Judgment of contingency between responses and outcomes, *Psychological Monographs*: *General and Applied*, 79 (1, Whole No. 594).

Kant I (1781/1965) *Critique of pure reason*, trans Smith NK (Macmillan, London).

Kelly K (2007) Ockham's razor, empirical complexity, and truth-finding efficiency, *Theoretical Computer Science*, *383*: 270-289.

Keys A (ed, 1980) *Seven countries: a multivariate analysis of death and coronary heart disease* (Harvard, Cambridge).

Keys A, Menotti A, Aravanis C, Blackburn H, Djordevic BS, Buzina R, Dontas AS, Fidanza F, Karvonen MJ, Kimura N, Mohacek I, Nedeljkovic S, Puddu V, Punsar S, Taylor HL, Conti S, Kromhout D & Toshima H (1984) The Seven Countries Study: 2,289 Deaths in 15 Years, *Preventive Medicine*, *13*: 141-154.

Kitcher P (1995) Revisiting Kant's epistemology: Skepticism, apriority, and psychologism. *Noûs*, *29*: 285-315.

Lu H, Yuille A, Liljeholm M, Cheng PW, Holyoak, KJ (2008) Bayesian generic priors for causal learning, *Psychological Review*, *115*: 955-984.

Lustig RH (2012) The toxic truth about sugar. *Nature*, *482*: 27-29.

McCullagh P, Nelder JA (1989). *Generalized linear models* (2nd edition, CRC Press, Boca Raton).

Menotti A, Pudd, PE, Lanti M, Kromhout D, Blackburn H & Nissinen A (2003) Twenty-five-year coronary mortality trends in the seven countries study using the accelerated failure time model. *European Journal of Epidemiology, 18*: 113–122.

Pearl J (2000) *Causality: Models, reasoning, and inference* (Cambridge Univ Press, Cambridge).

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current theory and research*, eds Black AH, Prokasy WF (Appleton-Century Crofts, New York), pp. 64-99.

Rothman KJ, Greenland S, Lash TL (2008) *Modern epidemiology* (3rd edition, Lippincott, Williams & Wilkins, Philadelphia).

Salmon WC (1965) The status of prior probabililities in statistical explanation. *Philosophy of Science, 32*:137-146.

Sheps MC (1958) Shall we count the living or the dead? *The New England Journal of Medicine*, *259*: 1210-1214.

Sloman S (2005) *Causal models: How we think about the world and its alternatives.* (Oxford Univ Press, New York).

Spirtes P, Glymour C, Scheines R (1993/2000) *Causation, prediction and search* (2nd edition, MIT Press, Cambridge).

Tenenbaum JB, Griffiths TL (2001) Structure learning in human causal induction. In TK Leen, TG Dietterich, & V Tresp (eds), *Advances in neural processing systems*, *13*: 59–65 (MIT Press, Cambridge).

Wickens TD (1989) *Multiway contingency tables analysis for the social sciences* (Erlbaum Associates, Hillsdale).

Yuille A, Lu H (2008) The noisy-logical distribution and its application to causal inference. In *Advances in neural processing systems*, vol. 20 (MIT Press, Cambridge).