Estimating Causal Power between Binary Cause and Continuous Outcome

Motoyuki Saito (motoyuki.saito28@gmail.com)
Department of Psychology, University of California, Los Angeles
Los Angeles, CA 90095, USA
Department of Psychological Science, Kwansei Gakuin University
Nishinomiya, Hyogo 662-8501, JAPAN

Patricia, W. Cheng (cheng@lifesci.ucla.edu)
Department of Psychology, University of California, Los Angeles
Los Angeles, CA 90095, USA

Abstract

Previous studies of causal learning heavily focused on binary outcomes; little is known about causal learning with continuous outcomes. The present paper proposes a qualitative extension of the causal power theory to the situation where a binary cause influences a continuous effect, and induces causal power under various ceiling situations with the continuous outcomes. To test the predictions, we systematically manipulated the type of outcome (continuous vs. percentage vs. binary) and the contingency information. The experiment shows that people estimate causal strength based on the linear-sum rule for continuous outcomes and the noisy-OR rule for binary outcomes. In the partial ceiling situation where causal power is partially inferred but not precisely estimated, the distribution of participants' judgments was bimodal with one mode at the minimum value and the other at the maximum value, suggesting some participants made conservative estimates while others made optimistic estimates. These results are generally consistent with the predictions of the causal power theory. Theoretical implications and future directions are discussed.

Keywords: causal reasoning; causal inference; causal power; continuous variable; integration rules.

Introduction

The ability to learn causal relations is essential for explaining past events, controlling the present environment, and predicting future outcomes. Decision making based on causal knowledge enables us to achieve desired outcomes and to avoid undesired consequences. When there are two causes of a desired outcome, we should consider which cause has a high causal strength for producing the outcome. To estimate causal strength, we need to consider not only the states of the effect in the presence of the cause but also that in the absence of the cause (Rescorla, 1968). When a teacher thinks about the effect of active encouragement on students' homework performance, for example, he or she has to check whether the student finishes the homework both in the presence and absence of encouragement. It has been recognized that both children and adults readily form representations of causal networks (see Holyoak & Cheng, 2011 for a review).

As causal relations are unobservable, they must be induced from observable events, and covariation among observable events serves as a fundamental cue to learn causal relations (Hume, 1739/2000). For binary variables, covariation is represented as patterns of presence and absence. A measure of contingency is described by ΔP (Jenkins & Ward, 1965):

$$\Delta P = P(E = 1|C = 1) - P(E = 1|C = 0) \tag{1}$$

where $P(E = 1|C = 1)$ is the probability of effect E given the presence of candidate cause C, and $P(E = 1|C = 0)$ is the probability of E given the absence of C. Values of ΔP range from −1 to +1. Positive ΔP values indicate a generative causal relation; negative ΔP values indicate a preventive causal relation.

Because ΔP is a measure of associative strength, it does not address issues in causation such as confounding and ceiling effects. For example, although it is impossible to judge the causal effect when the outcome always occurs regardless of the presence or absence of the cause (i.e., $P(E = 1|C = 1) = P(E = 1|C = 0) = 1$), the ΔP model indicates that there is no causal relation (i.e., $\Delta P = 0$). To model causal strength, Cheng (1997) proposed the power PC theory and derived generative causal power as an estimate:

$$w_c = \Delta P / [1 - P(E = 1|C = 0)] \tag{2}$$

Causal power w_c is a function not only of contingency but also of the base rates of the effect. When the effect is always present, generative causal power is undefined, therefore explaining the (generative) ceiling effect. Buehner, Cheng, and Clifford (2003) systematically manipulated covariation information and demonstrated that judgments were well described by the causal power. Causal power is interpreted in the framework of causal Bayes nets by Glymour (2001) and of causal Bayesian models by Tenenbaum & Griffiths (2001; Griffiths & Tenenbaum, 2005).

Lu, Rojas, Beckers, and Yuille (2016) proposed a Bayesian theory of sequential causal learning. Their theory assumes that people select a different integration rule according to the type of outcome variable. On one hand, the noisy-OR rule is appropriate for a binary outcome and is consistent with Equation (2) in the causal power theory (Cheng, 1997). The rule assumes that two causes influence an outcome independently. It states that the effectiveness of two causes, both present, is the sum of the causal power of each minus their product (i.e., $P(E = 1|A = 1, B = 1) = w_A + w_B - w_A \times w_B$). On the other hand, the linear-sum rule is appropriate for a continuous outcome and is widely used in associative learning models such as the R-W model (Rescorla & Wagner, 1972). This rule simply calculates the sum of the influence of each cause (i.e., $P(E = 1|A = 1, B = 1) = w_A + w_B$). Lu et al. (2016) presented a sequential Bayesian model that explains previous findings on outcome-additivity in variations of the blocking paradigm.

Several empirical studies have provided supporting evidence for the use of the linear-sum rule for a continuous outcome. Rashid and Buehner (2013) systematically manipulated the quantity of continuous outcomes and tested which integration rules people use. The results appear inconsistent in that participants use the linear-sum rule for a generative cause and noisy-OR rule for a preventive cause. They suggested that the use of the linear-sum rule might be due to the absence of an upper limit for the quantity of the continuous outcome in their cover story. Prevention has a natural lower limit, the outcome quantity equal to 0, sharing that property with binary outcomes. Saito (2015) manipulated means and standard deviations in causal learning with continuous outcomes and found that judgments are largely explained by difference in the means, but not by difference in the standard deviations. White (2015) examined causal judgments of interventions in temporal sequences of a continuous outcome variable in single individuals and reported that most of the results were explained by the difference between the mean outcome value for the pre-intervention time periods and that for the post-intervention time periods. These results suggest that people use the linear-sum rule for continuous outcomes. However, these studies do not reveal whether people use a different integration rule depending on the type of outcome variable since they did not compare judgments for continuous outcomes with those for the binary outcomes. In addition, it remains unknown how people estimate causal strength under various ceiling situation with the continuous outcomes. Since integration rules are core parts of the models of causal learning, it is important to investigate how people choose an integration rule.

In this paper, we extend the causal power theory qualitatively to address continuous outcomes and derive predictions under various ceiling effects. For our purposes, we treat cardinal outcomes as a special case of continuous outcomes. We also report a study investigating whether people choose the appropriate integration rules according to the type of outcome variables and whether their judgments correspond to causal-power predictions.

Estimating causal power with continuous outcomes

The reasoner's goal is to induce the unobservable causal power of a candidate cause from observable events (Cheng, 1997). Consider a situation where a continuous effect E may be produced by a binary background cause B and/or a binary candidate cause C. Assume that:

(1) B and C influence E independently,
(2) B could increase E but not reduce it,
(3) The causal powers of B and C are independent of the frequency of occurrences of B and C, and
(4) E does not change unless it is influenced.

These assumptions are similar to those with binary cause and effect (cf. Cheng, 1997; Pearl, 1998).

The joint influence of background cause B and candidate cause C on the continuous outcome E is given by the linear-sum rule (cf. Lu et al., 2016). According to this integration rule, the influences of multiple causes are integrated by simple addition. Since the outcome can take on different values, expected value and conditional expected value are used to describe its state. The expected value of the continuous outcome is calculated as follows:

$$E[e] = P(b) \cdot w_b + P(c) \cdot w_c \tag{3}$$

In this equation, $P(b)$ and $P(c)$ denote the probabilities of occurrences of the background cause and candidate cause. Variables w_b and w_c are causal powers of the background cause and candidate cause. Although two different integration rules are used for a binary outcome (i.e., noisy-OR rule for generative cause; noisy-AND-NOT rule for preventive cause), there is no distinction between generative and preventive causes in case of a continuous outcome.

When the cause is present (i.e., $P(c) = 1$), the conditional expected value given the presence of the cause is

$$E[e|c] = P(b|c) \cdot w_b + w_c \tag{4}$$

Similarly, the conditional expected value given the absence of the cause (i.e., $P(c) = 0$) is

$$E[e|\neg c] = P(b|\neg c) \cdot w_b \tag{5}$$

Subtracting Equation 5 from Equation 4 yields the difference in conditional expected values (i.e., $\Delta E = E[e|c] - E[e|\neg c]$). The difference in conditional expected values is:

$$\Delta E = P(b|c) \cdot w_b + w_c - P(b|\neg c) \cdot w_b \tag{6}$$

If we assume that the background cause and the candidate cause occur independently, two conditional probabilities equal to one another (i.e., $P(b|c) = P(b|\neg c) = P(b)$). Therefore, the causal power of the candidate cause w_c is represented as follows:

$$w_c = \Delta E - \{P(b|c) - P(b|\neg c)\}\ w_b = \Delta E \tag{7}$$

Within the range of outcome values greater than the minimum and less than the maximum, predicted values of the causal power are simply the differences in conditional expected values.

Predictions of the value of w_c vary depending on the value of the continuous outcome. To illustrate these predictions, consider a situation where a teacher investigates the effect of active encouragement on students' homework performance and gives 100 homework problems to each student. For example, a student had finished 25 out of 100 previous homework problems assigned; after the encouragement, the student finished 75 out of 100 new homework problems assigned. The causal power w_c is the difference in performance before and after the encouragement (i.e., $\Delta E = E[e|c] - E[e|\neg c] = 75 - 25 = 50$).

However, this is not the case where one of the outcome values reaches the upper limit. We hypothesize that depending on the reasoner's assumption about the counterfactual value of the outcome if there were no upper limit, w_c has a range of possible values. We replace $E[e|c]$ in Equation (4) with the assumed counterfactual value $E'[e|c]$. Suppose a student had finished 50 out of 100 previous homework problems assigned and then finished 100 out of 100 new homework problems assigned. It is inferred that the causal power is equal to or larger than 50, but not precisely determined. Thus, the prediction of the causal power theory is an interval. Whereas some cautious reasoners might estimate the minimum value in the interval (50 in this case, resulting from $E'[e|c] = 100$), other reasoners might estimate a higher value in the interval (e.g., 100, resulting from $E'[e|c] = 150$). When both outcomes are at the maximum value (e.g., a student finished 100 out of 100 homework problems regardless of the encouragement), the interval spans the entire range from 0 on and causal power becomes undefined. We call the former the partial ceiling situation and the latter the full ceiling situation. The difference between the partial and full ceiling situations is a unique feature in causal learning with continuous outcomes that have maximum values. The predictions of the causal power theory are shown in Table 1.

The purpose of the present study is to investigate whether people use proper integration rules according to the type of outcome variable and whether people differentiate between the partial and full ceiling situations. In addition to the conditions with continuous outcomes and binary outcomes, we added a condition with percentage outcomes. This is because the upper limit for percentage outcomes has a clear maximum of 100, unlike that for continuous outcomes.

Table 1: Design and predictions of the experiment.

Continuous & Percentage				Binary			
$E[e\|c]$	$E[e\|\neg c]$	ΔE	causal power	$P(e\|c)$	$P(e\|\neg c)$	ΔP	causal power
100(0)	0(0)	100	≥100*	1.00(0)	0.00(0)	1.00	1.00
100(0)	25(4.3)	75	≥75	1.00(0)	0.25(0.4)	0.75	1.00
75(4.3)	0(0)	75	75	0.75(0.4)	0.00(0)	0.75	0.75
100(0)	50(5)	50	≥50	1.00(0)	0.50(0.5)	0.50	1.00
75(4.3)	25(4.3)	50	50	0.75(0.4)	0.25(0.4)	0.50	0.67
50(5)	0(0)	50	50	0.50(0.5)	0.00(0)	0.50	0.50
100(0)	75(4.3)	25	≥25	1.00(0)	0.75(0.4)	0.25	1.00
75(4.3)	50(5)	25	25	0.75(0.4)	0.50(0.5)	0.25	0.50
50(5)	25(4.3)	25	25	0.50(0.5)	0.25(0.4)	0.25	0.33
25(4.3)	0(0)	25	25	0.25(0.4)	0.00(0)	0.25	0.25
100(0)	100(0)	0	NA	1.00(0)	1.00(0)	0.00	NA
75(4.3)	75(4.3)	0	0	0.75(0.4)	0.75(0.4)	0.00	0.00
50(5)	50(5)	0	0	0.50(0.5)	0.50(0.5)	0.00	0.00
25(4.3)	25(4.3)	0	0	0.25(0.4)	0.25(0.4)	0.00	0.00
0(0)	0(0)	0	0	0.00(0)	0.00(0)	0.00	0.00

Note. Numbers in parentheses are standard deviations. The causal power theory predicts "≥100" for the continuous group and "100" for the percentage group.

Method

Participants

A total of 136 participants were recruited from Amazon Mechanical Turk (http://www.mturk.com/). An additional 35 participants were tested but excluded for failing to pass the comprehension question (see below for details). All were native English speakers and residing in the US.

Experimental design

Participants were randomly assigned to one of three groups differing on the type of outcome (continuous, percentage, or binary). For all groups, the candidate cause was a binary variable (i.e., presence or absence of encouragement). Exclusion by the comprehension question resulted in unequal group sizes (56 participants in the continuous group, 43 in the percentage group, and 37 in the binary group). In addition to manipulating type of outcome, contingency information was manipulated within-subject (see Table 1). In the continuous and percentage groups, there were 15 contingency conditions resulting from the combination of five levels (100, 75, 50, 25, 0) of conditional expected values in the presence and absence of the cause. The difference between $E[E|C = 1]$ and $E[E|C = 0]$ for each condition yielded five levels of nonnegative values in the outcome magnitude ($\Delta E = E[E|C = 1] - E[E|C = 0] = 100$, 75, 50, 25, 0). Similarly, the binary group had five levels of nonnegative values in the difference (i.e., $\Delta P = P(E = 1|C = 1) - P(E = 1|C = 0) = 1.00$, .75, .50, .25, .00). Participants in each group completed the causal learning task for all contingency conditions. The order of the contingency conditions was randomized across participants.

Procedure

Instructions Participants were asked to read the instructions carefully and answer each question thoughtfully. The exact instructions in the continuous group were as follows (italicized sentences differed across groups):

> A math teacher wants to investigate the effect of active encouragement on students' homework performance. Students are given *100 math homework problems* of similar difficulty. The teacher randomly assigns some students to receive encouragement and assigns other students to receive no encouragement.
>
> Imagine that you are a teaching assistant for the class. You are responsible for checking whether or not a student receives encouragement and *how many out of the 100 homework problems the student finishes (0-100)*.
>
> You will see several sets of student records. Each set contains the records of students from a school ordered in a random sequence. Each record describes

> a student's homework performance before and after the experiment. After observing the records of sixteen students from a school, you will be asked to judge how much the encouragement increases performance at that school.

For the continuous group, the effect was a continuous variable (i.e., number of finished homework problems). The same instructions were used in the percentage group with one exception: the outcome observation was described as "what percentage of the homework problems the student finishes (0-100%)." In the binary group, both cause and effect were binary variables. Specifically, the instructions stated the outcome observation as "whether or not the student finishes the homework problems."

After reading the instructions, participants were asked to answer the comprehension question that checks the understanding of random assignment. The exact question was (italicized sentences differed across groups):

> Before you begin viewing the records of the students' homework performance, consider the following situation. Suppose we conduct a study, and find that: *the average number of the homework problems students in the experimental group (those who received encouragement) finish is 65. Likewise, the average number of the homework problems students in the control group (those who did not receive encouragement) finish is 65 as well*. Recall that the students are were randomly assigned to one or the other group. Can the homework performance in the experimental group be attributed to encouragement?

Participants were required to provide a "yes" or "no" answer and to justify their answer briefly. This question was intended to exclude participants who did not read the instructions properly and to encourage the assumption that the influence of background causes (i.e., causes other than encouragement) on homework performance was constant across the two groups (cf. Buehner et al., 2003). Similar questions were used in the percentage and binary groups with the corresponding modifications of the descriptions in terms of percentages. Participants received no feedback on their answers to this question.

Learning phase The learning phase consisted of 16 trials that presented information about the cause and effect in a pre-post design. For the continuous group, participants were requested to observe whether a student receives encouragement (present or absent) and how many out of the 100 homework problems the student finishes (0-100) before and after encouragement. On each trial, homework performance before the encouragement for a student was described with the illustration and text (e.g., "A student (ID: 12345) at this school finished 25 out of 100 previous homework problems assigned"). Student ID was a five-digit random number and designed to show that each trial described a different student. The states of the encouragement were provided with the sentence (e.g., "The student received encouragement" or "The student did not receive encouragement"). The other two groups followed an identical procedure, except that the outcomes were expressed in percentage terms for the percentage group (e.g., "The student finished 75% of new homework problems") and as present or absent in the binary group (e.g., "The student finished the new homework problems"). The inter-stimulus interval was 1000-ms, and the button to proceed to next trial was presented 500-ms after the presentation of all the information. Each trial was separated by a 500-ms blank screen. Participants were required to learn causal strength of the encouragement on homework performance through trials.

There were 16 trials for each contingency condition in Table 1. Encouragement was present on 8 trials and was absent on 8 trials. For the continuous group, the outcomes were normally distributed with the variance set to be ten times that in the binary group (see standard deviations in parentheses in Table 1). The order of trials was randomized within-subject. To familiarize participants with the procedure, practice trials were presented prior to the learning phase.

Test phase After the 16 learning trials, participants were asked to estimate the causal strength of the candidate cause in a counterfactual question. In the continuous group, the question was "Suppose the next student (ID: 23456) at this school finished 0 out of 100 previous homework problems assigned. If the student now receives encouragement, how many out of 100 new homework problems will the student finish?" The responses were made on a rating scale ranging from 0 to 100. Our scale limits the maximum strength to 100 so that responses can be compared across groups. Similar questions were used in the percentage and binary groups with modifications of the descriptions corresponding to the outcome type (e.g., for the binary group, "If these 100 students now receive encouragement, how many of them will finish their new homework problems?"). In addition, participants were also asked to report confidence in their judgment on a scale ranging from 0 (not confident at all) to 100 (extremely confident). After their judgments, participants completed the next contingency condition. To encourage the independence of judgments in each condition, participants received the following instructions: "Recall that the schools have students from very different socioeconomic backgrounds, and encouragement may have different effects on the students from school to school. Please evaluate each school separately."

Results

Participants who failed to pass the comprehension question were excluded from our analysis below. This procedure reduced noise, but did not alter the general pattern of the results. Since the causal power theory makes different predictions for the non-ceiling and ceiling situations, separate analyses were conducted. Figure 1 shows the mean ratings of causal strength in non-ceiling situations. Overall, participants clearly differentiated between continuous and

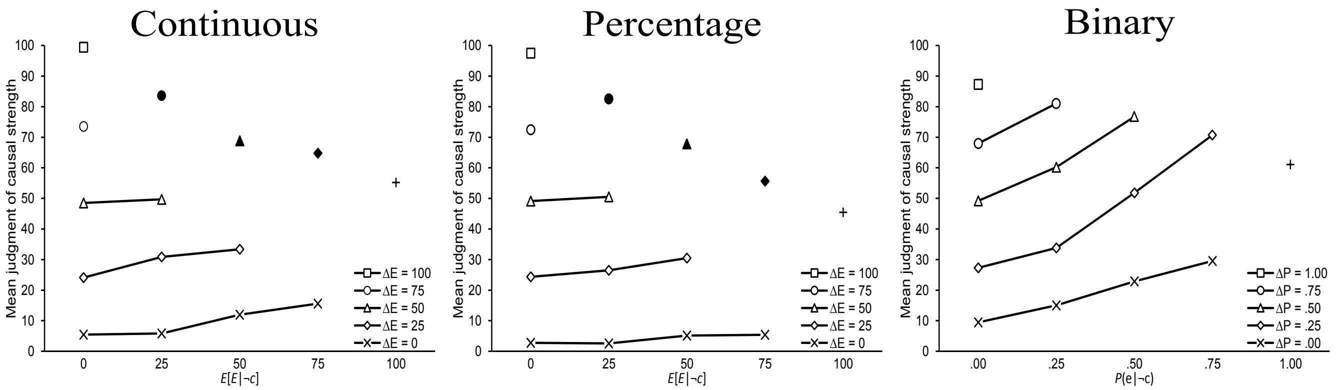

Figure 1: Mean ratings of causal strength in each contingency condition. Judgments with the same level of ΔE or ΔP in the non-ceiling situations are connected by lines. Judgments in the partial and full ceiling situations are represented by black symbols.

percentage outcomes on one hand and binary outcomes on the other. In the continuous group, judgments generally corresponded to the difference between the conditional expected values (i.e., $\Delta E = E[E|C = 1] - E[E|C = 0]$). Similar results were obtained in the percentage group, but the trend was much more evident. In contrast, judgments in the binary group were affected by both the difference between conditional probabilities ΔP and the base rates of the effect $P(E = 1|C = 0)$. These descriptive analyses were confirmed by statistical analyses.

A two-way mixed ANOVA with the type of outcome (continuous vs. percentage vs. binary) as between-subjects factor and the contingency condition (11 contingency conditions except for the ceiling situations) as within-subject factor resulted in a significant two-way interaction, $F(20, 1330) = 6.45$, $MSE = 210.3$, $p < .001$, $\eta^2_G = .069$. To explore the results in greater detail, we analyzed the effect of the type of outcome for each ΔE and ΔP condition. In the $\Delta E = 50$ and $\Delta P = .50$ conditions, a two-way mixed ANOVA revealed a significant interaction between the type of outcome and contingency condition, $F(2, 133) = 5.54$, $MSE = 116.1$, $p = .005$, $\eta^2_G = .036$. As expected, judgments varied as a function of the base rate of the effect in the binary group, $F(1, 36) = 10.95$, $MSE = 206.5$, $p = .002$, $\eta^2_G = .118$, but not in the continuous and percentage groups, Fs < 1. The interaction was also significant in the $\Delta E = 25$ and $\Delta P = .25$ conditions, $F(4, 266) = 7.76$, $MSE = 153.6$, $p < .001$, $\eta^2_G = .057$, and in the $\Delta E = 0$ and $\Delta P = .00$ conditions, $F(6, 399) = 2.33$, $MSE = 242.7$, $p = .032$, $\eta^2_G = .015$. Although the incremental pattern of the results in the binary group in the $\Delta P = .00$ condition was inconsistent with the predictions of the causal power theory, it may be explained by misperception of contingency for sequential trials due to working memory limitations. This outcome-density effect is consistent with causal-power predictions given the misperceptions (Cheng, 1997). It is worth noting that a similar but smaller trend was found in the continuous group, but not in the percentage group. This might be because the continuous group needs an assumption of equal upper limits to compare outcomes whereas the percentage group does not.

Figure 2 depicts distributions of individual judgments in the partial and full ceiling situations. In the partial ceiling situation (i.e., 100-25, 100-50, 100-75 conditions), a range of causal power is inferred (e.g., equal to or larger than 75 in the 100-25 condition). The distribution of participants' judgments appears bimodal with one mode at the minimum value of the interval and the other at the maximum value of the interval given our scale. These results indicate that some participants made conservative estimates while others made optimistic estimates. Dip tests confirmed the bimodality both in the continuous group ($D = 0.08$, $p = .013$ in the 100-25 condition, $D = 0.13$, $p < .001$ in the 100-50 condition, $D = 0.14$, $p < .001$ in the 100-75 condition) and percentage group ($D = 0.13$, $p < .001$ in the 100-25 condition, $D = 0.16$, $p < .001$ in the 100-50 condition, $D = 0.16$, $p < .001$ in the 100-75 condition). In contrast, the bimodality was not observed in the binary group, and the mode of the distribution corresponded to the point estimate of causal power (i.e., $w_c = 1$).

In the full ceiling situation where the causal power cannot be estimated (i.e., 100-100 condition), the distributions of the continuous and percentage groups appear bimodal while that of the binary group appear trimodal. This might be because participants had no option to answer "I don't know" in our materials.

Discussion

The present study qualitatively extended the causal power theory to deal with the continuous outcomes and tested whether people differentiate between continuous and binary outcomes. The results showed that people estimate causal strength based on the linear-sum rule for continuous outcomes and the noisy-OR rule for binary outcomes. In the partial ceiling situation where the estimation of causal power has a range, the distribution of participants' ratings was bimodal with one mode at the minimum value and the other at the maximum value, suggesting some participants made conservative estimates while others made optimistic

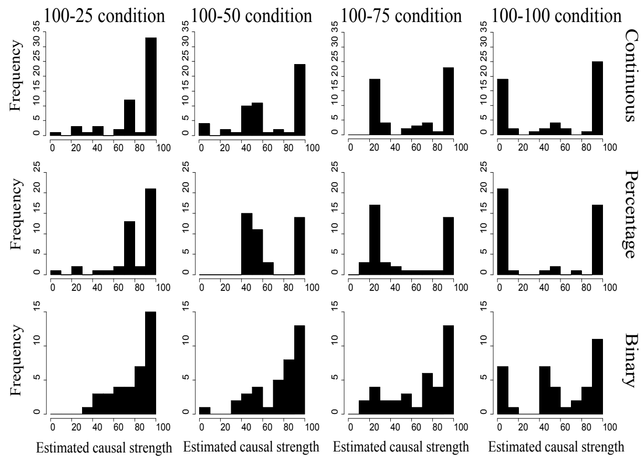

Figure 2: Distributions of individual judgments in the partial and full ceiling situations.

estimates. These results are generally consistent with the predictions of the causal power theory.

The present study has theoretical implications for understanding how people estimate causal power. Whereas covariation models (e.g., ΔP model, Jenkins & Ward, 1965) and associative models (e.g., R-W model, Rescorla & Wagner, 1972) adopt one integration rule, Bayesian models generally assume multiple integration rules (Griffiths & Tenenbaum, 2005, 2009; Lu et al., 2008, 2016). Our results demonstrate that people choose the proper integration rule according to the type of outcome, supporting the Bayesian models. Notably, this finding implies that people assume the invariance of causal power as a default, consistent with the proposal that causal invariance plays a key role in the construction of generalizable causal knowledge (Cheng & Lu, in press). The two integration rules respectively represent the invariance of causal power for the two outcome variable types. Another theoretically important aspect is the bimodal distributions in the judgments in the partial ceiling situations. Computational models generally predict averaged results. The observed bimodality suggests that models incorporating different conservatism values and/or priors may explain individual differences in the partial ceiling situations. Further investigations will shed more light on the question of how people estimate causal power.

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., & Lu, H. (in press). Causal invariance as an essential constraint for creating a causal representation of the world: Generalizing the invariance of causal power. In M. R. Waldmann (Ed). *The Oxford Handbook of Causal Reasoning*. Oxford, England: Oxford University Press.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661–716.

Hume, D. (1739/2000). *A treatise of human nature*. Oxford: Oxford University Press.

Holyoak, K, J., & Cheng, P, W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology, 62*, 135-163.

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

Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2016). A Bayesian theory of sequential causal learning and abstract transfer. *Cognitive Science, 40*, 404–439.

Lu, H., Yuille, A., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*, 955-984.

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

Rashid, A. A. A., & Buehner, M. J. (2013). Causal reasoning with continuous outcomes. In M. Knauff, M., Pauen, N., Sebanz, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 115-120). Austin TX: Cognitive Science Society.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*, 1-5.

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 research and theory* (pp.64-99). New York: Appleton-Century-Crofts.

Saito, M. (2015). How people estimate effect sizes: The role of means and standard deviations. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp2075-2079) . Austin TX: Cognitive Science Society.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 59–65). Cambridge, MA: MIT Press.

White, P. A. (2015). Causal judgements about temporal sequences of events in single individuals. *The Quarterly Journal of Experimental Psychology, 68*, 2149-2174.