Finding the Cause: Examining the Role of Qualitative Causal Inference through Categorical Judgments

Michael Pacer (mpacer@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94720

Woo-kyoung Ahn (woo-kyoung.ahn@yale.edu)
Department of Psychology, Yale U., 2 Hillhouse Ave
New Haven, CT 06520

Abstract

Previous work showed that people's causal judgments are modeled better as estimates of the probability that a causal relationship exists (a qualitative inference) than as estimates of the strength of that relationship (a quantitative inference). Here, using a novel task, we present experimental evidence in support of the importance of qualitative causal inference. Our findings cannot be explained through the use of parameter estimation and related quantitative inference. These findings suggest the role of qualitative inference in causal reasoning has been understudied despite its unique role in cognition. Further, we suggest these findings open interesting questions about the role of qualitative inference in many domains.

Keywords: Causal Reasoning, Qualitative Inference, Categorization

Qualitative and Quantitative Causal Inference

Studies have distinguished between two types of causal inferences: qualitative inferences ("whether an event X *is or is not* a cause of Y") and quantitative inferences ("to what extent is X a cause of Y"; Griffiths & Tenenbaum, 2005; Waldmann & Martignon, 1998). Causal reasoning has been traditionally construed as making quantitative inferences of causal strengths (e.g., Cheng, 1997; Rescorla & Wagner, 1972), in which data on presence or absence of two events[1] are used to estimate parameters, which describe the degree of a relationship between the cause and effect.

Within this causal strength paradigm, qualitative inferences are made by assessing the parameter estimates relative to a threshold. If the estimate is above that threshold, then it is a cause; if it is below, then it is not a cause. Such an approach casts qualitative causal inference as a process dependent upon quantitative inference. In the current study, we demonstrate that qualitative causal inference may play an important role such that people will make judgments which conflict with the judgments warranted by quantitative inference.

Quantitative and Non-Quantitative Data in Causal Inference

While quantitative data can be used in making qualitative causal inferences of whether one event causes another, non-quantitative data[2] can also be used in making qualitative inferences. Previous studies (e.g., Griffiths & Tenenbaum, 2005; 2009; Kuhn 1997; Waldmann & Martignon, 1998) have proposed that qualitative inference utilizes information such as intuitive theories (e.g., an inductive bias such as "novel foods may have unusual causal features, but common foods do not"), temporal information (e.g., events exert their influence on the future, not the past), or explicit claims (e.g., "X causes Y and Z").

For instance, if a person's lips turn green after eating a novel fruit, she may conclude from this single observation that the fruit caused the discoloration. However, it is unlikely that she will infer that the symptoms caused her to eat the fruit or that a fruit with which she had a great deal of experience (e.g. an apple) caused it. Additionally, if no abnormal symptoms present after eating the novel fruit, she may not think the fruit has any unusual causal features. Low probability events (e.g. green lips) prompt people to seek out causal explanations (e.g., Weiner, 1985), and if there is a novel, preceding event, that event may be identified as the cause even if quantitative data do not support that inference (e.g., Hilton & Slugoski, 1986; Kahneman & Miller, 1986).

The Role of Qualitative and Quantitative Inference

What remain unknown are details of how qualitative and quantitative inferences interact with one another (Griffiths & Tenenbaum, 2009), or whether one is more fundamental to inference generally. Previous work on causal inferences could be thought of as assuming that quantitative causal inferences are more fundamental, with qualitative inferences generally treated as an afterthought to be computed using a threshold and prior quantitative inferences.

Griffiths and Tenenbaum (2005) took another approach. Rather than estimating a single value for the strength of the causal relationship, Griffiths & Tenenbaum's model asks "whether or not a causal relationship exists" (i.e. a qualitative inference). In order to answer this, their model uses a large number of possible values for the strength of the causal relationship. In this way, their model gives the probability of the existence of a causal relationship without

[1] We refer to the type of information captured in simple contingency tables as quantitative information.

[2] By non-quantitative we simply mean any type of data that cannot be captured in a simple contingency table.

making assumptions about its strength. Peoples' judgments of causal strength actually better fit this type of qualitative model than previous quantitative ones (but see, Lu, Yuille, Liljeholm, Cheng & Holyoak, 2008). Like Griffiths and Tenenbaum, we argue that qualitative causal inference is important, perhaps fundamental to causal cognition, but our hypothesis includes no assumptions associated with any particular model.

We developed a paradigm which allows us to determine whether qualitative or quantitative inference is more predominantly used. The paradigm utilizes categorization judgments to infer participants' causal inferences, based on studies showing that background knowledge about causal features of category members determine categorization (e.g., Ahn, 1999; Ahn, Kim, Lassaline, & Dennis, 2000; Rehder & Hastie, 2001). We developed a novel problem in which different categorical judgments would be made depending on whether people engage in quantitative or qualitative causal inference. Participants' judgments can be used to determine which inference people used, as explained below.

In our experiments, participants read about two groups of ten people. One group ate mushrooms from Bag A, and nine people became sick, and one did not. The other group ate mushrooms from Bag B, and one person became sick and nine people did not.

Then participants read of a third group of ten people (Bag X group, henceforth), who after eating mushrooms either from Bag A or Bag B, five people became sick and five did not. Participants were asked from which of the two bags (Bag A or B) they thought the Bag X group likely ate.

Predictions of a Quantitative Approach

Formally, we treat this by comparing the probability of two hypotheses conditional on a set of data. We can use Bayes' rule (i.e. posterior $\propto$ likelihood*prior, or $P(H|D) \propto P(D|H)*P(H)$) to calculate the posterior odds of the two hypotheses given the known data. For our particular problem the equation (1) looks as follows[3]:

$$\frac{P(\textbf{Bag A}|D_{A,B,X})}{P(\textbf{Bag B}|D_{A,B,X})} = \frac{P(D_X|D_{A,B},\textbf{Bag A})P(\textbf{Bag A})}{P(D_X|D_{A,B},\textbf{Bag B})P(\textbf{Bag B})} \qquad (1)$$

Then to continue with our calculations we need to find the likelihoods for both hypotheses (**Bag A** and **Bag B**). We assume there is a unique, unknown value for the probability of an event (e) given some cause (c_i) ($p(e|c_i)$), which can be estimated ($\hat{p}(e|c_i)$).[4] In this scenario, $p(e|c_i)$ is the probability of getting sick given that the person ate a mushroom from Bag i. We estimate $\hat{p}(e|c_i)$ using maximum likelihood estimation (MLE) . Under these assumptions, the probability of someone getting sick given that they ate from Bags A and B are, respectively, $\hat{p}(e|c_A) = .9$ and $\hat{p}(e|c_B) = .1$.

Then if we treat the incidence of each person getting sick as an independent random variable, conditional on the person haven eaten from Bag i, our likelihood function is distributed as a Bin(10, $\hat{p}(e|c_i)$). Under these assumptions, the odds-likelihood equation is (2):

$$\frac{P(D_X|D_{A,B},\textbf{Bag A})}{P(D_X|D_{A,B},\textbf{Bag B})} = \frac{\binom{10}{5}0.9^5(1-.1)^5}{\binom{10}{5}0.1^5(1-.9)^5} = 1 \qquad (2)$$

That is, the likelihoods are equal to one another.

Because of this equality, we can surmise that, under this model, if we are to predict any preference for one hypothesis over another (i.e. *not* saying that Bag X is equally likely to be either Bag A or B), then this preference must derive from the prior. Since an equal number of people ate from each bag originally, it is reasonable to say that the priors are equal (i.e. P(**Bag A**) = P(**Bag B**)). There seems little reason to assume a priori that people are more likely to eat from either bag[5]. Using this model and its assumptions, the posterior prediction is that the probabilities that the people who ate from Bag X actually ate from Bag A or B are equal (i.e. $P(\textbf{Bag A}|D_{A,B,X}) = P(\textbf{Bag B}|D_{A,B,X})$).

Predictions of a Qualitative Approach: the Importance of Qualitative Inference Hypothesis

The model above is one with strong ties to the tradition of quantitative causal inference, in that it uses MLE to estimate parameters used in its predictions and that it effectively assumes the existence of a causal relationship between at least one of the bags and getting sick[6]. The alternative hypothesis we wish to propose emphasizes the *importance of qualitative inference* (IQI).

Generally put, our hypothesis is that, qualitative causal inference (i.e. inference whether or not a cause exists) can play a unique role in judgment not able to be described by quantitative inference. In this case, we predict that participants will prefer to say those from group X are more likely to have eaten from Bag A than B.

In group A, nine of ten people became sick; this seems to be a fairly low probability event. This potentially suggests that there is some explanation, likely a causal one, for this anomaly. The common cause principle (Reichenbach, 1956) suggests that people would prefer in this case to appeal to a common cause instead of multiple causes varying over nine people. The common property across the nine people in this scenario is eating mushroom. Thus, people would infer that the mushroom had the causal feature of causing sickness. As a result, mushrooms from Bag A are categorized as being causes of sickness; this is a qualitative inference.

On the other hand, of those who ate mushrooms from bag

[3] Where **Bag A** and **Bag B** are the events where the Bag X group ate from Bag A and Bag B, respectively, D_X is the data from the Bag X group (i.e. that 5 people got sick and 5 people did not get sick), and $D_{A,B}$ is the data from the Bag A and B groups,

[4] Other models of quantitative causal inference (e.g. ΔP (Cheng & Novick, 1992) and Power–PC (Cheng, 1997)) also estimate $\hat{p}(e|c_i)$ with MLE.

[5] Even non-equal priors would predict no differences between conditions if the quantitative data are held constant.

[6] This assumption comes from allowing that $\hat{p}(e|c_A) \neq \hat{p}(e|c_B)$.

B, only one of the ten became sick. This is more probable to have occurred just incidentally in the world than nine of ten becoming sick. People do not need to search for a causal explanation including a *common* cause, since only one person became sick. As a result, people will be less likely to include "causes sickness" as a feature of mushrooms from Bag B; this too is a qualitative inference.

Now, consider the group who ate from the unknown Bag X. Five of these ten individuals became sick. This event, as in the case of Bag A, is rather unusual. It may even suggest that some underlying common cause for all five people's sicknesses. People may infer that the mushrooms from Bag X had the same feature of causing people to fall ill. Since only Bag A had been inferred to contain mushrooms with this property, people would conclude that Bag X is more likely to be Bag A. This judgment contradicts judgments based solely on quantitative inference, and we argue that it arises because of the inference that Bag A and Bag X share a qualitative causal feature.

Note, however, that according to IQI, if Bag B were to also share a causal feature with Bag X, then both Bags A and B would be judged to be equally likely to be Bag X. This can be accomplished by creating a new causal feature, which Bags B and X share, but which A does not. For example, we alter the scenario so that the people who do *not* get sick instead become high. That is, instead of 1 of the people who ate from Bag B becoming sick and 9 of them not becoming sick, 1 becomes sick and 9 become high. This change extends to those who did not get sick when eating from Bags X and A – of those who ate from Bag X, 5 became sick and 5 became high, and of those who ate from Bag A, 9 became sick and 1 became high. In such a scenario, the IQI hypothesis would predict the same thing as the quantitative model (assuming equal priors): no preference for either bag. Because these scenarios effectively discriminate our hypotheses, we use them in Experiment 1.

Experiment 1

Method

One hundred twenty Yale University undergraduates participated in exchange for candy. Participants received one of the two stimuli, as illustrated in Figure 1. In each stimulus, participants read about two bags of mushrooms, which contained one type of mushrooms. One group of 10 people ate mushrooms from each bag, and participants were told the events that occurred to the members of each group. They were then told about a new group of 10 individuals who ate from only one of the two bags (Bag X). They were given information about these individuals (e.g. "5 got sick and 5 did not get sick"). Last, they were asked to estimate either the likelihood that Bag X was Bag A, or the likelihood that Bag X was Bag B. Varying the question in this way controls for any potential biases inherent in the scale. Participants made a vertical mark on a line denoting the likelihood of the hypotheses, from 0 to 100%.

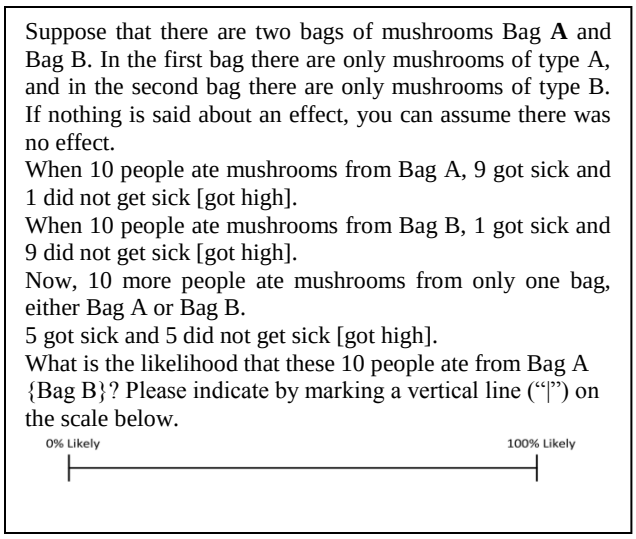

Figure 1: Scenario used in Experiment 1, 1-effect, Bag A condition, with the differences to the 2-effect condition in brackets ("[]") and the Bag B condition in curly brackets ("{}").

Participants' marks on the response line were converted into numbers ranging from 0 to 100 by measuring the distance of the mark from the left anchor of the line.

There were two conditions, *1-effect* and *2-effect*. In the *1-effect* condition, one type of effect was either present or absent (i.e., people got sick or did not get sick). In the *2-effect* condition, stimuli were the same except that "did not get sick" was replaced with "got high," resulting in two types of effects – getting sick and getting high. To sum, the study was a 2 (Question: Bag A or Bag B) × 2 (Condition: 1-effect vs. 2-effect) between-subjects design.

Results

Figure 2 summarizes the results. A 2 (Condition: *1-effect*, *2-effect*) × 2 (Question: *Bag A* or *Bag B*) ANOVA found a significant interaction, $F(1,116) = 4.42$, $MSE = 1591.40$, $p < .05$, but there were no significant main effects of either Condition, $F(1,116) = 2.45$, $MSE = 880.20$, $p > .1$, or Question, $F(1,116) = 1.80$, $MSE = 648.68$, $p > .1$. Planned comparisons showed that in the *1-effect* condition, mean likelihood ratings for *Bag A* (53.3%, $SD = 11.67$) were greater than those for *Bag B* (41.3%, $SD = 21.06$), $t(58) = 2.72$, $p <. 01$, $d = .70$. In the *2-effect* condition, however, there was no significant difference, $t(58) = .49$, $p > .6$ (*Bag A*: $M = 51.4\%$, $SD = 20.03$; *Bag B*: $M = 54.0\%$, $SD = 21.42$).

These results contradict the quantitative model and support the IQI hypothesis in two ways. First, the quantitative model predicted that in the 1-effect condition, participants should give equal likelihood ratings between Bag A and Bag B. Yet, as predicted by the IQI hypothesis, they gave higher likelihood ratings to Bag A. According to IQI hypothesis, this result was obtained because Bag A shared a causal feature with Bag X. Second, the quantitative model predicted no difference between the 1-effect and 2-

effect conditions, as the quantitative data are identical. Yet, people gave higher ratings to Bag A only in the 1-effect condition and not in the 2-effect condition. The IQI predicted no difference in preference for the two bags in the 2-effect condition because both Bags A and B have causal features that Bag X has.

Experiment 2

Experiment 2 replicates the findings of Experiment 1 using different manipulations. We developed a new condition to compare to the *1-effect* condition from Experiment 1. This new condition, called 2-cause condition, is exactly like the *1-effect* condition, except that we explicitly state that mushrooms within both bags contain chemicals known to cause sickness.

Because the amount or potency of the chemicals is not mentioned, the quantitative model would still predict no preference for either Bag A or B in the *2-cause* condition, just as in the *1-effect* condition. The IQI hypothesis predicts that in the *2-cause* condition participants should be more ambivalent about Bag X's identity, making participants more prone to think that the two bags are equally likely. Thus, the quantitative model predicts no difference between 1-effect and 2-cause conditions, whereas the IQI hypothesis predicts that preference for Bag A would be obtained only in the 1-effect condition.

Method

We recruited one hundred three participants through Craigslist. They participated in an online survey using Qualtrics®, with the option of entering a lottery for a $20 gift certificate to an online store.

There were two conditions, using either 1-effect or 2-cause scenario. The *1-effect* scenario was the same as in Experiment 1. The *2-cause* scenario was the same as the 1-effect scenario except that after the first two sentences, the following was inserted:

> *Mushrooms of type **A** contain **Adelosinol**, mushrooms of type **B** contain **Blemosine**. **Adelosinol** and **Blemosine** are known to cause people to become sick.*

Instead of asking participants to give a scalar response, we gave participants a 3-alternative-forced-choice question ("Please choose whichever bag you think that it is most likely that these 10 people ate from."), with the answer choices of "Bag A", "Bag B", or "Both bags are equally likely." That is, we explicitly provided an option for "both bags are equally likely" instead of inferring this choice from a scalar response. This way, we highlighted to participants that both bags could be equally likely, as predicted by the quantitative model. Yet, the IQI hypothesis predicted the preference for this option only in the 2-cause condition.

Results

Figure 3 depicts the proportion of each option chosen by condition. Although the stimuli differed only minimally, participants' responses were significantly different between the two conditions, χ^2 (2, N = 103) = 6.58, p < .05, ϕ = .25.

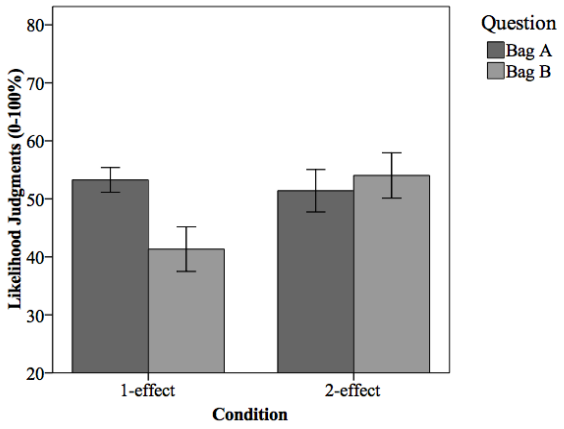

Figure 2: Mean Likelihood Judgments with SE bars in Experiment 1.

Given that the quantitative information was *exactly* the same across scenarios, this finding is unexplainable using our quantitative model.

More importantly, we can verify the IQI predictions using a χ^2-Goodness of Fit Test to compare the proportion of responses within conditions to an uninformed random response (i.e. each choice being chosen one-third of the time). The IQI predicts that the most frequent response in the *1-effect* condition would be Bag A, and this was the case ("Bag A" chosen 49%, χ^2 (2, N = 51) = 8.59, p < .05, ϕ = .41). The IQI predicts that in the *2-cause* condition, however, the most frequent response would be "both bags are equally likely"; this too was the case ("Both bags" chosen 56%, χ^2 (2, N = 52) = 12.04, p < .005, ϕ = .48). Thus both predictions of the IQI hypothesis are fulfilled.

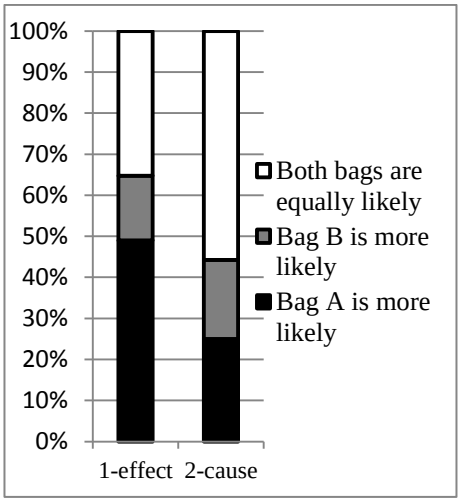

Figure 3: Proportion of times participants in Experiment 2 chose each option in response to the prompt "Please choose whichever bag you think that it is most likely that these 10 people ate from."

Discussion

Two experiments pitted the predictions of two hypotheses for predicting participants' categorization judgments against one another. One model uses MLE to make quantitative inferences based on quantitative information specified in the scenarios, just as many models of causal induction estimate parameter values of causal strengths based on quantitative information. However, the IQI hypothesis uses qualitative inferences about objects' causal features to guide judgment. Participants' categorization judgments better fit the predictions made by the hypothesis that emphasizes the role of qualitative inference than those made by the model that considered only quantitative inference.

Assuming that equal priors for people choosing to eat from Bag A and B[7], the quantitative model predicts no differences in participants' judgments across all experiments and conditions. Particularly, it predicts that participants should claim that Bag X is equally likely to be either Bag A or Bag B. This model is unable to explain any of the differences found between the conditions, or the general preference for Bag A (or against Bag B) found in both studies' *1-effect* conditions.

However, the results match the predictions made by our instantiation of the IQI hypothesis. Participants in the *1-effect* conditions expected that Bag X was more likely to be Bag A than Bag B. The IQI predicted this due to the inference of a causal feature shared by Bags X and A that Bag B lacked. On the other hand, in both the *2-effect* and *2-cause* conditions, participants judged that the bags were equally likely. The IQI predicted these results because participants presumably made qualitative inferences that Bag X shared a causal feature with both Bags A and B. By using a forced choice paradigm in Experiment 2's *2-cause* condition, we were able to establish not only that participants lacked a preference for Bag A or B, but that the option, "both bags are equally likely," was indeed participants' most preferred response, the strongest possible prediction made by our hypothesis. Every prediction made by the IQI hypothesis obtained.

Previous evidence supporting the idea that qualitative causal inferences are not merely a function of quantitative inference and a threshold was limited to causal judgment tasks, specifically to how well different models fit participants' responses in these tasks (Griffiths & Tenenbaum, 2005; Lu et al., 2008). The experiments presented here go one step further and investigate the role of qualitative causal inference in other cognitive processes, using its effects on a categorization judgment (i.e., should Bag X be categorized as Bag A or B?) to infer its presence. The current studies demonstrate not only that qualitative causal inference can result in judgments that contradict the judgments suggested by quantitative inference, but that qualitative causal inferences play a larger role in cognition than previously suspected. Still, there is much yet to be learned about the complete role of qualitative inference in reasoning.

[7] If priors were unequal, this would have predicted a systematic bias for Bag A or B that did not change across conditions.

Future Work

There are many open questions in relation to precise relationship between qualitative and quantitative causal inferences and how they affect other cognitive processes. What other types of judgments and inferences rely on qualitative inference independent of quantitative inference? What is the time course of these inferences? For example, are qualitative inferences judged first, only to estimate the strength of the relationship later? Does use of one type of inference differentially affect causal learning, especially with regards to active data search and planned intervention?

What *is* the underlying computation that is driving our effects? Bayesian models of qualitative causal inference exist (Griffiths & Tenenbaum, 2005; 2009), however, much work is still needed before we understand how to include qualitative inferences into hypotheses tests as in (1). For instance, the task used in the current study did not provide all information required to derive predictions of existing quantitative models of causal induction, namely, frequencies of all four cells in a contingency table. Extensions of our studies which provide this data will allow the testing of various models' predictions. Discovering the roles played by qualitative and quantitative inference is an important part of *all* inferential tasks, and as such advances on this problem may have wide implications for cognition generally (Griffiths & Tenenbaum, 2009).

Inferential tasks are notoriously difficult and even young children are skilled inferential learners. Perhaps the use of qualitative inference bestows unique advantages when one is learning. And these advantages might change over development as one gains more experience and data. For example, learners (e.g. children) use qualitative inference to guide their exploratory behavior by focusing first on gathering information about objects with unknown causal structure (Gweon & Schulz, 2008), only to estimate precise parameters later.

There are theoretical similarities between the qualitative inference of hidden causal features as we discuss them here and essentialism, which some argue plays an integral role in categorization (Gelman, 2003). For instance, a raccoon transformed to look like a skunk which then shares many surface features with the category "skunk" is still classified as a raccoon presumably because of people's qualitative inference to the hidden essence (Keil, 1989). One might consider these surface features "quantitative data". Then, this task similarly pits inference which considers only quantitative data (something related to quantitative inference as we discuss it) against qualitative inference. Under this perspective, these essentialist judgments could be thought to be in line with our hypothesis that qualitative inference sometimes overrides quantitative inference in guiding judgment.

Essentialism has been implicated as a process relevant to many different domains of reasoning (Gelman, 2003). Thus

if the sorts of qualitative inferences we discuss here *are* related to essentialism, qualitative causal inference may play a role in many domains. Yet, what role it plays in each domain may vary. Thus further investigation into how the role of qualitative inference changes across domains may prove very fruitful.

Conclusions

Our experiments use categorical judgments to explore the roles of qualitative and quantitative causal inference. We test the predictions of one computational model using MLE to estimate parameters in the spirit of traditional models of quantitative inference against the predictions of a hypothesis that suggests people infer causal features in objects and will use these qualitative inferences in later categorization judgments. This hypothesis suggests that qualitative inference plays an important, unique role in cognition. The results support the notion that qualitative inference can result judgments that conflict with the judgments suggested by quantitative inference.

Because of the generality of the problem of inference, and the potential role of qualitative inference in other inferential tasks, there is need for further study of the role of qualitative inference in many varieties of cognitive processes.

Acknowledgments

Many thanks to Benjamin Rottman and Joshua Tenenbaum for commentary on the project, and Elizabeth Seiver, Chris Holdgraf, and Caren Walker for comments on a draft of this paper. This research was supported in part by a National Institute of Health Grant MH57737 given to the second author.

References

Ahn, W. (1999). Effect of Causal Structure on Category Construction. *Memory & Cognition, 27*, 1008-1023

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality, *Cognitive Psychology, 41*, 361-416.

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

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, **99**, 365-382.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gopnik, A. & Schulz, L. (eds.) (2007) *Causal learning: Psychology, philosophy, and computation*. New York: Oxford University Press

Griffiths, T.L., & Tenenbaum J. (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.

Gweon, H., & Schulz, L. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers' exploratory play. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 570-574.

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

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, **93**, 136-153.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kuhn, D. (1997). Is good thinking scientific thinking? In D. R. Olson & N. Torrance (Eds.), *Modes of thought: Explorations in culture and cognition* (pp. 261-281). New York: Cambridge University Press.

Liljeholm, M. & Cheng, P.W. (2009). The influence of virtual sample size on confidence and causal strength judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **35**(1:1) 157–172.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, **115**, 955-982.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt & Company.

Rehder, B. & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. Journal of Experimental Psychology: General, 130, 323-360.

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, NY: Appleton-Century-Crofts.

Reichenbach, H. (1956). *The Direction of Time*. University of California Press, Berkeley.

Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Deny (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 1102-1107). Mahwah, NJ: Erlbaum.

Weiner, B. (1985). 'Spontaneous' causal thinking. *Psychological Bulletin*, **97**, 74-84.