FEATURE INFERENCE AND THE CAUSAL STRUCTURE OF CATEGORIES: A MODIFICATION

JOOSHIM KIM

Northwestern University Department of Psychology
E-mail: jookim2017@u.northwestern.edu

Abstract - In extension to the causal feature inference experiment conducted by Rehder and Burnett (Rehder & Burnett, 2005), this paper further studies the psychological representation of causal knowledge based on Bayesian networks, especially in regards to the causal Markov condition. The results of this experiment indicate that people do perceive causality according to the causal Markov condition, in that they take into account only the causal factor and discard all other information when inferring unobserved features in a machine. Given that the study was a modified version of Rehder and Burnett, which previously observed people's violation of the causal Markov condition, the results suggest that at least one or more of the modifications made in this study had significantly influenced participants to represent causality according to the causal Markov condition. Exactly what modification made this change is unclear; nevertheless, the results provide direction for future research that will reveal additional insights regarding how people perceive, interpret, and interact with the world around them.

Keywords - Psychology, Psychology of thinking, Cognitive psychology, Human cognition, Social sciences, Human learning and education, Baysian network, Causal Markov condition, Perception of causality.

I. INTRODUCTION

Every day we encounter and interact with causality in many aspects of our lives. From something as small as turning off the light switch to something as big as getting into a car accident, we cannot imagine our lives without the force of causality playing important roles in everything we think, do, or say. It is not surprising, therefore, that many psychologists and philosophers find it fascinating to study how people cognitively represent causality. In fact, many theories and principles regarding people's reasoning from causality have already been brought up, such as the principles of Determinism, Priority, and Mechanism (Bullock, Gelman, & Baillargeon, 1982), the causal Markov condition (Pearl, 2000), and the inference of best explanations when causal events occur (Harman, 1965).

Among these many theories, this paper specifically focuses on the Bayes nets, or more precisely the causal Markov condition, which is the property that Bayes nets theory depends on significantly. This condition essentially claims that "conditioning on the states of the immediate causes…renders that variable statistically independent of all other variables in the net, except for those it causes (its "descendants")" (Rips, 2011). In other words, this principle explains that it is only the immediate causal factor that has the power to change the outcome of its descendants, and that other descendant variables stay independent of oneanother. There have been studies that explored whether people indeed represent causality according to this particular principle. In these studies, people were asked to infer unobserved features based on the causal information provided. Unfortunately, however, it has been found that people use their general knowledge about underlying mechanisms instead of just looking at the given causal knowledge when inferring their predictions (Park & Sloman, 2013). In another study conducted by Rehder and Burnett, the experimenters "observed a pervasive violation of…the causal Markov condition—because the presence of characteristic features invariably led participants to infer yet another characteristic feature" (p.1). Rehder and Burnett concluded that there might have been other reasons like the typicality problem ("people take an exemplar as "well functioning" category member when it has most or all of category's characteristic features" (p.1)) that played into people's cognition when deciding on their inferred answers (Rehder & Burnett, 2005).

Upon learning of this experiment, it seemed only natural to explore whether there would be any other way to prove that people indeed perceive causality according to the causal Markov condition. In order find an answer to this question, an experiment similar to what Rehder and Burnett used in their 2005 study was designed and conducted on Northwestern University's undergraduate students. The experiment was conducted in similar fashion to the Rehder and Burnett's experiment in that its purpose was to find out how "causal knowledge is used to make inferences about unobservable or unobserved features of [a] novel [object]" (Rehder & Burnett, 2005, p.2). However, the experiment was modified slightly in that the kind of novel object used and the procedures participants were asked to follow throughout the study were adjusted. These modifications were incorporated in order to find out which factors might be able to influence participants' cognition regarding causality and perhaps enable them to think in ways that are more aligned with the causal Markov condition.

II. EXPERIMENT

In this experiment, participants learned about a novel machine (instead of a novel animal or objects used in Rehder and Burnett's study) and were asked to make inferences about one of its unobserved features. In the experimental condition (Common Cause condition), participants learned about how the machine works – (a) that the machine can be turned on into one of two different modes, (b) that the machine can perform three different programs regardless of its mode, (c) the likelihood of each program running every time the machine turns on (thus providing the base rate for the programs' operations), and (d) the causal relations between the modes and the programs. Specifically, participants learned that one of the two modes served as a common cause for the operation of the three programs. In the control condition, participants learned about the machine with the exception of (d), the causal relationship between the modes and the programs. In this way, it was possible to tease apart and compare the effect of causal knowledge between the Common Cause and the control groups.

The hypothesis was that if people indeed represent causality according to the causal Markov condition of the Bayesian networks principles, the participants in the Common Cause condition would only focus on the causal factor (in this case the mode that causes the programs to operate) and infer the answers to the unobserved feature consistently regardless of whether the other programs are operating or not. For the control group, however, participants would infer their answers only according to the base rate information given during the learning phase since the causal relationship had not been provided to them at all.
This experiment also put time constraints on the participants during the inference phase (both Common Cause and control groups) in order to restrain the participants from ruminating too many possibilities before they answer. Since the experiment aimed to find out how causality is cognitively processed in the most basic way, it was necessary to induce the most intuitive answers possible from the participants.

2.1. Method

2.1.1. Materials

A novel object called Machine Z was used for both the Common Cause and control conditions. This machine could turn on into two different independent modes called X Mode and Y Mode; each time the machine is turned on, it randomly turns on into either of the two modes. The machine also has the ability to perform three independent programs, called A, B, and C regardless of its mode (i.e. A, B, and C can equally and independently be performed when the machine is on X Mode as well as when it is on Y Mode). Out of a hundred times the machine is turned on (and once again regardless of its mode), it has been found that each program (A, B, and C) has a 75% chance of operating and a 25% chance of not operating. Most importantly, it was clarified that the machine's operation did not depend on the operation of the programs (i.e., the machine could still be "on" without any of the programs A, B, or C operating). This clarification was important because a lack of this information could have easily confused some participants into thinking that at least one of the programs must be operating for the machine to be functioning in the first place as that is generally how machines work.

After these general descriptions of how Machine Z works have been provided to the participants to learn at their own pace, participants in the Common Cause condition were given extra information regarding the causal relationship between X Mode and the three programs. The participants read the statement "X Mode causes A, B, C to operate" and were given a diagram visually explaining this relationship between X Mode and the three programs (see fig. 1, top). The way the diagram was presented was different from the diagram shown to participants in the original experiment by Rehder and Burnett (see fig. 1, bottom). By parsing the causal relations into three independent relationships, it was possible toemphasize the independence that each of the programs has within X Mode.

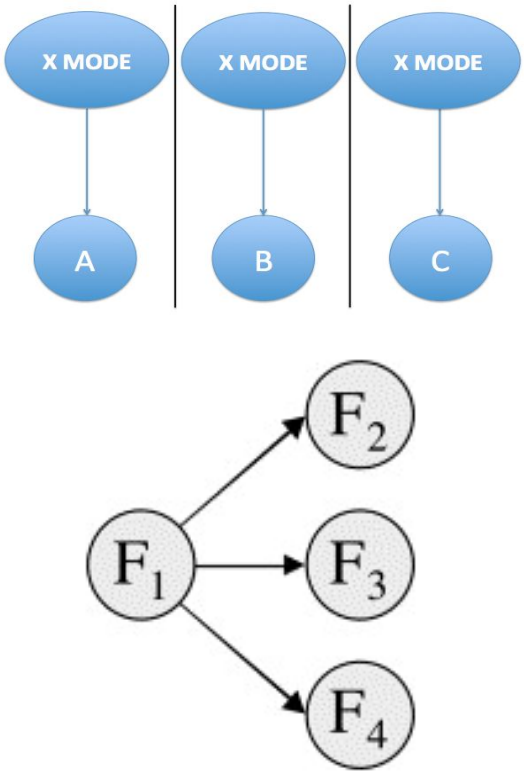

Fig. 1. (a & b) (Top) Visual diagram of Common Cause structure; (Bottom) Common Cause network (Rehder & Burnett, 2005).

2.1.2. Participants

Ten Northwestern University undergraduate students (6 female, 4 male) participated in the experiment; no tangible incentive was given for their participation.

2.1.3. Design

Participants were randomly assigned in equal numbers (5/5) into either the experimental condition (Common Cause; CC) or control condition (Control; C).

2.1.4. Procedure

Prior to starting the experiment, participants were given a slip of paper to choose from, which determined what condition the participant was to be assigned to (Common Cause or control) for randomization purposes. The slips the participants chose would also be used by the participant to write down his or her answers in the later feature inference phase.

The experiment had two phases: learning phase and inference phase. Both phases were conducted through a computer (PowerPoint slides), and the experimenter verbally answered any questions the participants had throughout the study.

Before proceeding to the learning phase, participants were given written instructions informing them that they will not be able to go back to a previous slide once they went on to the next slide. These instructions were created to make sure that participants read the descriptions carefully, given the complexity of the material. Once they started the learning phase, participants read descriptions of Machine Z at their own pace (no time constraint was enforced). At this point, participants in the Common Cause group were provided with the information regarding the causal relationships between the machine's modes and programs, both in written and visual form (diagram shown previously). When they were done with the reading and learning of the novel machine, they were tested on the learned materials through multiple-choice questions. Since the questions were on a PowerPoint slideshow, all participants gave answers to the questions verbally to the experimenter. Prior to beginning the test, participants were also given instructions that they would have to answer all questions correctly in order to proceed to the next phase (the inference phase). They were additionally told that they would have to go back to the first question (retake the test) if they answered incorrectly to any of the questions. When participants answered a question incorrectly, they were allowed to go back and reference the initial information screens before retaking the test. For the Common Cause condition, there were 9 questions in total—6 questions on the general functioning of the machine and the other 3 on the causal relationships between X Mode and the three programs (ex. What causes program A to operate?). For the control condition, there were only 6 general questions, which were the same 6 questions used in the Common Cause condition to determine comprehension regarding the general functioning of the machine.

Prior to beginning the feature inference phase, participants were once again presented with instructions, which notified them of the 10-second time constraint on each inference question. The experimenter kept track of the time while participants were working on each question and alerted the participants when the time remaining was less than 5 seconds. Throughout the inference phase, participants were given information of 1) what mode Machine Z was in and 2) whether two of the three programs were operating or not. The mode was color-coded in red when it was in X Mode and in blue when it was in Y Mode. After assessing the given information, they were required to infer the percentage of whether the unobserved program was operating or not. For example, participants were given a situation where the machine was in X Mode (in red) and where program A was operating and B was not operating. Then they were asked to infer, in one-to-three-digit numbers (0-100), the likelihood of program C operating as a percentage. Participants wrote down their answers on the slip they chose in the randomization process earlier in the experiment. After answering each question, there was a brief confirmation question asking the participants what mode the machine was in. Participants were asked to verbally convey to the experimenter the answer to this confirmation question. They were immediately informedupon answering whether they had answered correctly or incorrectly. When answered incorrectly, experiment asked the participant to write down the correct answer on the paper so that the participants implicitly understood the importance of the mode in answering the inference questions. Inferences were made in all 24 possible combinations in which one of the programs' operations is unobserved and each of the two other programs are either in operation or not within either the X or Y mode.

2.2. Results

Fig. 2 shows the results of the experiment. The Y-axis represents the participants' mean answer of how likely they think the unobserved program is going to be operating. The X-axis represents the number of observed programs operating, from none of them operating (0) to both of them operating (2). The red bar represents the answer given when in X Mode, and the blue bar in Y Mode. The top graph shows the results for the Common Cause condition, and the bottom one for the control condition.

In the control condition, participants' answers were consistent (75%) throughout the entire inference phase regardless of how many programs were operating or which mode the machine was in. In the Common Cause condition, however, there seemed to be a slight yet gradual increase in the participants'

answers regarding the number of programs operating. Yet, after running t-tests on the results, it has been found that X-0, X-1, and X-2 (Mode-number of programs running) were not significantly different from each other (all three t-tests [X-0:X-1, X-1:X-2, X-0:X-2] had a p-value equal to .37). Likewise, the differences among Y-0, Y-1, and Y-2 were not statistically significant. However, it was important that X-0, X-1, and X-2 were all significantly larger than their pairs Y-0, Y-1, and Y-2 (p=5.6) because it demonstrates the effect of the causal knowledge between X Mode and the three programs. Since it was X Mode and not Y Mode, which was causally connected to the programs, participants seemed to have inferred higher percentage numbers when the machine was in X Mode than when it was in Y Mode. (For specific data analysis: Common Cause Mean—X-0: 73.3, X-1: 78.3, X-2: 81.7, Y-0: 20, Y-1: 20, Y-2: 25, Standard Error—X-0: 2.6, X-1: 1.2, X-2: 1.3, Y-0: 2.2, Y-1: 1.4, Y-2: 3.9; Control Mean—X-0: 75, X-1: 75, X-2: 75, Y-0: 75, Y-1: 75, Y-2: 75, Standard Error—X-0: 0, X-1: 0, X-2: 0, Y-0: 0, Y-1: 0, Y-2: 0)

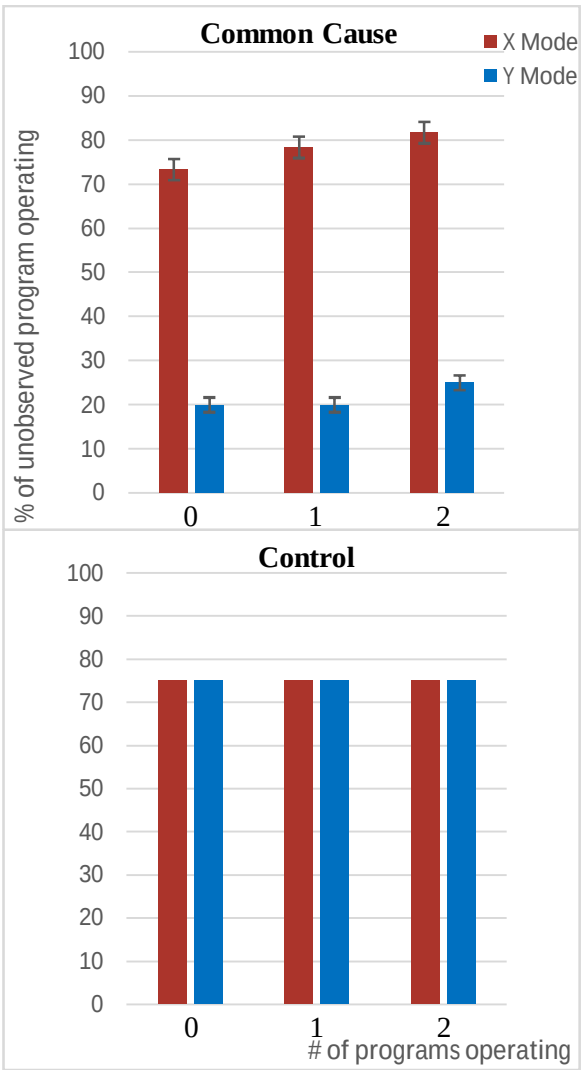

Fig. 2. (a & b) Feature inference results from Experiment 1; Top: Common Cause condition, bottom: control condition.

III. DISCUSSION

The results of the experiment seem to have clearly fulfilled the foremost purpose of this paper, which was to prove that people actually represent causality according to the causal Markov condition. In the Common Cause condition, participants determined their answers based on what mode the machine was in instead of looking at the operation of other programs. In other words, the inferences participants made were consistent throughout the whole inference phase according to the mode the machine was presented to be in, not according to how many of the programs were presented to be in operation or not. In fact, when the experimenter asked the participants what their deciding factors were when inferring their answers, most of them answered that they "only looked at the modes" and not at the other programs. This result is significantly aligned with the causal Markov condition in that none of the other programs (the "descendants") influenced participants' inferences, but only the causal factor (the mode) was taken into consideration when they were determining their answers. Therefore, it is possible to suggest that people indeed represent causality according to the Bayes nets theory, specifically the causal Markov condition. One of the reasons that could explain why the participants only looked at the modes and not the programs is that the design of the study put a lot of emphasis on the modes throughout the inference phase. As explained above, the modes in each of the inference questions were color-coded to either red or blue depending on the mode and there was also a brief confirmation question reminding the participants to focus on the modes. Additionally, for each question that described the machine's mode, it was described as the machine being in "X Mode, not Y Mode" and vice versa, hence highlighting the machine's current mode once again. All of these factors could have turned the participants' attention from the operation of other programs to what mode the machine was in, thus making them base their answers on only the modes of the machine when inferring their answers. Another reason could be that the inference questions were timed—because they were given only 10 seconds to decide on each of their answers, their strategy could have changed into having the least effort put into the task by only looking at the modes and not assessing what the other programs' operation or lack thereof might mean for each question.

Furthermore, it is important to notice another surprising aspect of the results: the lack of the typicality effect for both the Common Cause condition as well as the control condition. Originally in Rehder and Burnett's experiment, they found the effect of typicality in the control condition (participants' inferred answers got higher as more characteristic features were present even when they learned that the base rates for all features were 75%).

In the present study, however, all participants answered in a highly consistent manner throughout the whole inference phase, abandoning any possibility of the typicality effect. The reason for this result could be that the object used in the experiment was a novel "machine" – participants would not have been able to apply any background category knowledge to how the machine functions and would have had to only rely on the previously provided descriptions about the machine when answering the inference questions. On this note, however, it had already been found that participants again violated the causal Markov condition even when a whole new object called "dax" was used in the same experiment (Rehder & Burnett, 2005). Knowing this, this explanation may not present the most convincing argument for such results. Another possible explanation is that the presentation of the descriptions regarding the machine highly emphasized the independence of the three programs. In Fig. 1 (top), the diagram shows three different relationships between X Mode and the three programs instead of having X Mode be one cause that branches out its effects on three different programs like the one used in the original study. Also, when participants learned the base rates for the programs' operation, they were given three different sentences explaining the base rates for each of the three programs (ex. Program A has 75% chance of operating and 25% chance of not operating; Program B has…). Such presentation of the programs' independence could have clarified the relationships among the machine's modes and the three programs, thus enabling the participants to think more in a causal Markov condition manner.

CONCLUSIONS

The purpose of this research paper, which was to prove that people represent causality using the causal Markov condition, was supported by the experiment. However, it is unclear which modification exactly caused the drastic difference of participants' answers from that of the Rehder and Burnett experiment. It can be assumed that one or more of the modifications adopted in the experiment prompted people to think in a causal Markovian way. It is possible that the use of a novel machine called Z, which did not quite fit into any of the pre-existing categories of objects, enabled participants to focus only on the causal knowledge provided. Or perhaps the usage of "modes" and "programs" may have clarified the causal relationship better than the original Rehder and Burnett study where a category member had four characteristic features. Also, the use of different wordings and visual information that emphasized the independence of each of the programs may have influenced the participants to look at the operations of programs separately from one another. A completely different perspective may be that it was all of the modifications working together that enabled the participants to comply with the causal Markov principles. In any way, however, it should be noted that the results of this experiment are not completely convincing given the small sample size and the limited scope of sample demographics. Further studies on the effects of each distinct modification adopted in this paper may hugely benefit the cognitive psychology field in better understanding how and when people utilize the Markovian way of perceiving causality.

Learning causality serves as a fundamental basis for humans to understand the world. Thus, further investigation regarding how we perceive causality will not only yield meaningful insights into how people function in society, but may also allow us to tackle many complex social problems in the future.

REFERENCES

[1] Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. The developmental psychology of time, 209-254.

[2] Harman, G. H. (1965). The inference to the best explanation. The Philosophical Review, 88-95.

[3] Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. Cognitive psychology, 67(4), 186-216.

[4] Pearl, J. (2000). Causal inference without counterfactuals: Comment. Journal of the American Statistical Association, 428-431.

[5] Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. Cognitive Psychology, 50(3), 264-314.

[6] Rips, L. J. (2011). Lines of thought: Central concepts in cognitive psychology. Oxford University Press, USA.

★★★