Modeling Causal Inference from Emotional Displays

Dennis W.H. Teo (diswhdt@nus.edu.sg)
Department of Information Systems and Analytics, National University of Singapore

Zheng Yong Ang (zhengyong@comp.nus.edu.sg)
Department of Psychology and Department of Computer Science, National University of Singapore

Desmond C. Ong (dco@comp.nus.edu.sg)
Department of Information Systems and Analytics, National University of Singapore

Abstract

Can people learn causal relationships about the world from someone's emotions? We present a computational model integrating observational causal learning with emotional information, which uses emotional displays to disambiguate the beliefs, desires, and knowledge of other agents, in turn allowing causal inferences about the world. We compared our model predictions to human causal judgements on two observational learning tasks involving multiple possible causes or multiple possible outcomes. Across three studies ($N = 129, 127, 125$), emotional displays (compared to actions alone) led people to interpret agents' beliefs differently, which in some contexts resulted in different causal inferences. Our model closely reflected these patterns of belief and causal inference and revealed new insights on how people learn causal relationships from others' emotions.

Keywords: Causal learning; Emotional inference; Observational learning; Bayesian modeling

Introduction

You see your friend hit a switch on a machine and start frowning. You have no clue what the machine does, but you can easily infer that your friend expected something to happen but failed in their attempt. By extension, you learned something about how this machine works (or does not work).

Emotions are a powerful source of information for people to learn about the world (Ong, Zaki, & Goodman, 2016; Saxe & Houlihan, 2017; Wu, Schulz, Frank, & Gweon, 2021). For example, people—even very young children—can infer someone's beliefs and desires from their emotional expressions, as well as infer what events elicited those emotions (Wu, Baker, Tenenbaum, & Schulz, 2018; Wu & Schulz, 2018). These examples of emotion reasoning mainly involve inference about people's latent mental states. In this paper, we explore how emotions can also provide information about *causal relationships* in the environment. For instance, how do emotional displays help people learn the causal functions of an unfamiliar device? We propose that emotions contribute to *observational causal learning* through a rich intuitive theory of reasoning about others' minds.

Observational causal learning is the process of learning causal relationships by observing others (Meltzoff, Waismeyer, & Gopnik, 2012), typically by inferring from others' actions. It is distinct from learning via statistical covariation (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Cheng, 1997; Kushnir & Gopnik, 2005) or verbal testimony (e.g., Harris, Koenig, Corriveau, & Jaswal, 2018; Bonawitz & Shafto, 2016). Instead, observers have to reason about people's actions as motivated by their beliefs and desires (Goodman, Baker, & Tenenbaum, 2009; Teo & Ong, 2021), which explains how people can learn about a new technology simply by watching an experienced user working with it.

But inferring causality from actions alone might be misleading, such as in situations involving failed actions that are causally irrelevant. People need additional information such as verbal cues (e.g., "Whoops!") to accurately understand failed (Gweon & Schulz, 2011; Bridgers, Altman, & Gweon, 2017) or accidental actions (Gardiner, Greif, & Bjorklund, 2011; Gardiner, 2014). In this study, we consider the informational utility of emotional displays such as frustration that often accompany people's failures, which we propose that observers use to disambiguate such situations. Emotion understanding goes beyond recognizing expressions, and relies on rich intuitive theories of emotion that allow people to reason about how others would react in complex situations (Ong, Zaki, & Goodman, 2015, 2019; Saxe & Houlihan, 2017; Wu et al., 2021). Observers consider contextual factors that are motivationally-relevant for other agents—a third-person version of cognitive appraisal. For example, people understand that goal attainment often gives rise to happiness whereas failing to reach an expected goal often causes frustration. These intuitive theories allow people to make sense of others' behaviors and intentions from their emotional displays (Wu, Baker, et al., 2018) and even about their degree of knowledge (Wu, Haque, & Schulz, 2018; Wu, Schulz, & Saxe, 2018).

We propose a Bayesian generative model that describes how people perform inference from emotional displays to infer the beliefs, desires, and knowledge of other agents, which in turn affords causal inferences about the world. This model extends past work on observational causal learning (Goodman et al., 2009; Teo & Ong, 2021) by allowing the model to learn from emotional displays together with actions and outcomes. The model also extends work on emotion inference (Ong et al., 2019) by showing how emotional displays can be used to infer causal relationships in the environment through the latent mental states of other agents. Additionally, the model gives insight on how conflicting information is handled between its information sources, such as between direct observation (of outcomes) and indirect testimonial evidence (e.g., from actions and emotional displays).

We compared our model predictions to people's inferences

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

on two observational learning tasks. In Studies 1a and 1b, we investigated a "multiple-causes" scenario where an agent performs two different actions sequentially, with the desired outcome occurring only after both actions were completed. Here, it is ambiguous whether both actions were intentional (and the outcome required both actions) or the agent's first action was intended but failed. In Study 2, we considered a "multiple-outcomes" scenario where an agent's action could lead to multiple possible outcomes. This scenario provides conflicting information from direct observation and testimonial evidence (e.g., via inference from actions and displays through beliefs). Across both scenarios, we manipulated the presence of emotional displays and hypothesized that such displays help resolve ambiguity over the agent's beliefs, allowing different inferences about causality.

Computational Model

We propose a computational model (Fig. 1) that describes how emotional information aids causal inference, integrating aspects of earlier models of observational causal learning (Goodman et al., 2009; Teo & Ong, 2021) and affective cognition (Ong et al., 2019; Wu, Baker, et al., 2018). We assume an agent interacting with the world via observable actions (A), causing outcomes (O). The model infers the latent causal structures of the world (W) by making two key assumptions about agent behavior. First, the agent acts rationally such that their actions are chosen to maximize their desires (D) given their beliefs (B) about the world (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). Second, the agent experiences emotions (E; inferred through observable emotion displays Y) after subjectively appraising the outcome of their actions, given their beliefs and desires (Wu, Baker, et al., 2018). As the agent may be knowledgeable about the world (or conversely, may have false beliefs about the world), the model allows for the agent's knowledge (K) to vary.

By observing O, A, and Y, the model performs Bayesian inference to jointly estimate the latent variables W, K, B, and D (marginalizing over emotions E). This is formalized by:

$$P(W,K,B,D|O,A,Y) \propto P(W)P(K)P(B|W,K) \\ P(D)P(A|B,D)P(O|W,A) \\ \sum_E P(E|B,D,O)P(Y|E) \qquad (1)$$

We implemented this model using PyMC3[1] for two types of learning scenarios. In a multiple-causes scenario (Study 1), the observer infers the likely cause of an outcome of interest, operationalized via an agent acting on multiple switches in order to turn on a light bulb. In a multiple-outcomes scenario (Study 2), the observer infers the likely causal effect of an event of interest, by observing an agent acting on a switch in a room with two different light bulbs.

[1] https://tinyurl.com/EmotionalCausalInference

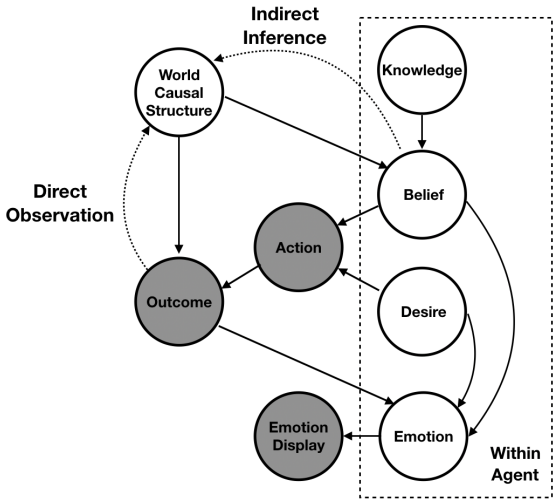

Figure 1: Graphical model. Nodes represent variables, shaded nodes are observable while clear nodes are latent, and edges between nodes represent causal influence.

Causal Structure and Belief, $P(W)$, $P(K)$, $P(B|W,K)$

In a multiple-causes scenario, W represents the likely cause of an outcome. We sampled W from a discrete space of candidate causes ("blue switch only"; "orange switch only"; "both switches") with a prior probability matching the estimated prior beliefs of a sample of participants (see Study 1 procedure for details). In a multiple-outcomes scenario (Study 2), W represents the likely causal effects of an event, and we modeled two possible causal effects (pink or orange bulb lighting up) from two independent $Beta(\alpha = .1, \beta = .1)$ distributions. These hyper-parameters reflect a prior that people favor deterministic relations (close to 0 or 1 probabilities).

K represents the probability that the agent is knowledgeable, and is sampled from a $Beta(5,2)$ distribution. The chosen hyper-parameters reflect that observers tend to attribute high knowledge to the agent (mean reliability of .7). When the agent is knowledgeable, B reflects W. Otherwise, a random belief is sampled.

Agent's Desire and Actions, $P(D)$, $P(A|B,D)$

In the multiple-causes scenario, we sampled D from a $Binomial(p = .5)$ distribution (1 indicates desire to turn on the bulb). In the multiple-outcomes scenario, D was sampled from a Categorical distribution containing the likely desires (e.g., "turn on orange bulb"). Actions were sampled under the rational agent assumption: When the agent believes they know how to fulfill their desires, they act based on their belief. Otherwise, the agent is unlikely to act. Our model observed the object interactions of the agent ("push blue and orange switches" in Study 1 and "push blue switch" in Study 2).

Outcomes, $P(O|A,W)$

O represents the bulb(s) turning on and depends on A and W. For example, if pushing a blue switch is necessary to turn on

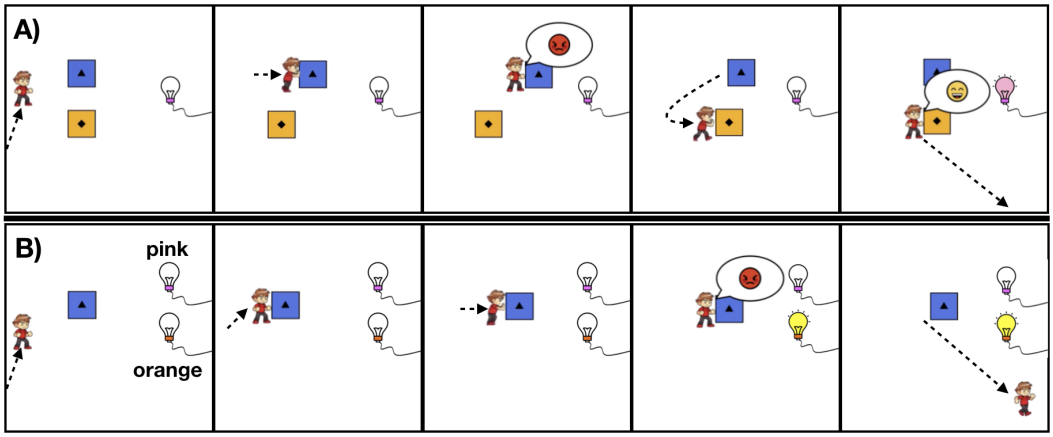

Figure 2: Experimental Materials for Studies 1a and 1b (top panels) and Study 2 (bottom panels). The panels show the action sequence of the agent with emotion displays. Note that no emotional displays were shown in the no-emotion conditions. The top **A)** panels match the 5 Measurement Points in Study 1b where participants provided causal ratings.

a bulb (W) and the blue switch was pushed (A), then the bulb will turn on ($O = 1$). Otherwise, the bulb remains off ($O = 0$). For Study 1, we allowed for the possibility of observing an outcome after some temporal delay from the causal event since the scenario involves an action sequence with multiple potential causes (similar to Teo & Ong, 2021).[2]

Emotions and Displays, $P(E|O,B,D)$ and $P(Y|E)$

We sampled E from a Categorical distribution ("happy", "frustrated", "neutral") with relative probabilities based on the following logic: By default, the model expects the agent to feel neutral. If the outcome matches the agent's desires (goal attainment), then the agent is more likely to experience (and express) happiness. Conversely, if the agent's plan did not go as expected and their desires are left unfulfilled (goal non-attainment), then the agent is more likely to feel frustrated. Y (emotional display) was set as equivalent to E.

Study 1a and Study 1b: Multiple causes

In Study 1a, we investigated a multiple-causes scenario where an agent attempted two distinct actions to turn on a light bulb. In the no-emotion condition, an agent pushed a blue switch (with no observable effect), and then pushed an orange switch, which is followed by the bulb lighting up. This is ambiguous, and consistent with several hypotheses: for example, the agent thinks that both switches are necessary to turn the bulb on; alternatively, the agent initially thought that the blue switch was sufficient, but realized it failed to reach the desired outcome, and so went on to push the orange switch.

In the emotion condition, we manipulated emotional displays by having the agent express frustration after pushing the blue switch and express happiness after pushing the orange switch. We predicted that the emotion displays help to disambiguate the agent's beliefs, by affording the inference that the agent only intended to push the blue switch and believed it was sufficient to turn on the bulb (although this resulted in failure). As a more ambitious test of our model, we ran a follow-up study, Study 1b with the same stimuli, where we additionally measured causal inferences at multiple key moments of the scenario rather than merely at the end.

Without emotional displays, we predicted that observers are more likely to infer that both switches were causally important because presumably the agent was rational and pushed both switches intentionally (Goodman et al., 2009; Teo & Ong, 2021). Our model predicted that observers in the emotion condition would instead infer that pushing the blue switch led to failure and discount its causal relevance relative to other possibilities like the orange switch.

Participants. For Study 1a, we recruited 158 participants from Prolific and excluded 29 participants who failed our attention checks. The final sample included 129 participants (41.9% females; $M_{age} = 28.8$; $SD_{age} = 9.2$). For Study 1b, we recruited 155 Prolific participants, excluding 28, for a sample of 127 (44.9% females; $M_{age} = 26.1$; $SD_{age} = 8.7$).

Materials. We created two videos involving an agent, two switches (blue and orange), and a light bulb (see Figure 2). The agent starts by pushing the blue switch, with no visible change in the light bulb—in the emotion condition, the agent additionally displays a frustrated expression. Next, the agent pushes the orange switch. This time, the bulb lights up in response—in the emotion condition, the agent displays a happy expression. Finally, the agent exits the scene.

Procedure. In Study 1a, participants were randomly assigned to the emotion (N=59) or no-emotion (N=70) condition, and were shown the corresponding video. Following the video, participants were presented with several candidate causes of the bulb lighting up ("blue switch only"; "orange switch only"; "both switches") and asked to rate their like-

[2] An earlier action (e.g., push blue switch) has a .2 probability of causing the bulb to turn on even though some time has passed between these events.

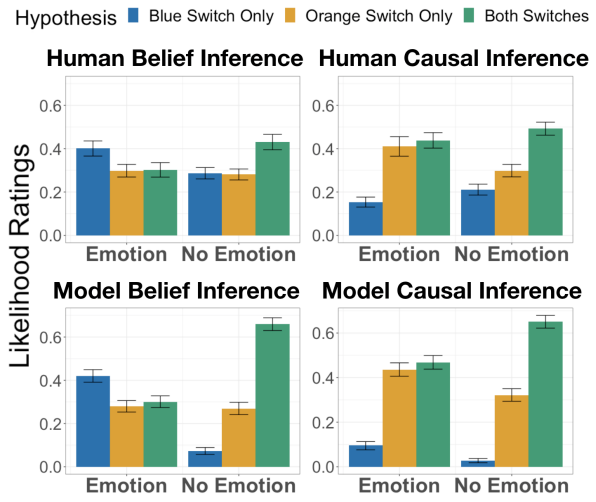

Figure 3: Study 1a results: inferences in a multiple-causes scenario. Mean causal ratings and agent-belief ratings across conditions for 3 hypotheses (top row). Corresponding model predictions are presented below. Error bars represent 95% confidence intervals.

lihoods on a Likert scale from 1 (Extremely Unlikely) to 9 (Extremely Likely). They were also asked about the agent's beliefs (how likely the agent expected the respective switches to cause the bulb to light up), desire (how likely the agent wanted to switch on the light bulb), and knowledge (how likely the agent knew how to turn on the light bulb). The causal and belief ratings were normalized to sum to 1 within each participant. Participants then responded to attention checks and demographic questions.

Study 1b employed similar stimuli and procedures (N's= 58, 69 in the emotion and no-emotion condition respectively). The main difference was that instead of obtaining one set of ratings after the video, the video was paused and participants provided causal ratings at five measurement points, as shown in Fig. 2. These were: 1) when the agent was in front of the switches; 2) while the agent was pushing the blue switch; 3) after pushing the blue switch in the no-emotion condition and after the agent displayed an emotion in the emotion condition 4) while the agent was pushing the orange switch and; 5) at the end of the scene. As the first measurement point was before anything happened in the scene, we used participants' mean causal ratings at this point as the model priors on the causal structure W (for both Study 1a and Study 1b).

Study 1 Results

When shown only the agent's actions in the no-emotion condition, participants were more likely to infer that the agent believed that pushing *both* switches would lead to their desired outcome, and were more likely to infer that *both* switches were causally necessary for the light bulb to turn on. By contrast, when the agent displayed frustration after pushing the blue switch, participants presumably took it as a failed action, leading them to infer that the blue switch is causally irrelevant and favor other possibilities such as the orange switch.

Comparing the two conditions, we found that participants in the emotion condition (compared to no-emotion), were less likely to infer the agent's belief that both switches were necessary to turn on the bulb (Mann-Whitney-Wilcoxin Test, $W = 1028.5, p < .001$), and more likely to infer the agent's belief that the blue switch was necessary ($W = 3055.5, p < .001$). These differences in belief inferences resulted in different causal inferences. Participants in the emotion (compared to no-emotion) condition were more likely to infer that the orange switch turned on the bulb ($W = 2924.0, p < .001$). We note that in both conditions, participants still gave the most weight to the "both switches" hypothesis.

Our computational model captured these qualitative patterns (see Figure 3) across the experimental conditions and the two variables (*B*elief inference and *W*orld causal inference), albeit with more polarized inferences for the no-emotion condition. Quantitatively, our model's predictions also correlated well with people's ratings across the several latent variables measured, including inferences of the *W*orld causal inference, and the agent's *B*elief, *D*esire, and *K*nowledge ($r(14) = .883$, $p < .001$).

In Study 1b, we evaluated our model's predictions of people's causal inferences at key moments of the scene (see Figure 4). The first measurement point served as our estimate for people's priors over the world's causal structure. Upon seeing the agent push the blue switch (second measurement point), both our model and participants inferred through the agent's actions that the blue switch is causally relevant. But upon observing no change in the light bulb (third measurement point), both model and humans decreased their confidence in the blue switch (although in the no-emotion condition, our model predicted a larger decrease compared to participants). At this point, observers also begin favoring different hypotheses across the experimental conditions. Observers in the emotion condition began to favor other hypotheses upon learning that the agent failed their action ("orange switch only" and "both switches"). By contrast, observers in the no-emotion condition leaned toward "both switches" as they assumed that the agent intended both actions. This trend continued to the last measurement point. Overall, our model co-varied with these dynamic causal inferences ($r(28) = .746$, $p < .001$), giving evidence that our model closely reflects people's underlying reasoning processes.

Study 2: Multiple outcomes

In Study 2, we considered a multiple-outcomes scenario. An agent pushes a switch next to two light bulbs, one pink and one orange, and only the orange bulb lights up. By manipulating the agent's emotional displays—frustration vs no reaction—we gave conflicting (vs consistent) testimonial evidence that the agent expected the pink (vs orange) bulb to light up.

Our model predicts that observers would make different belief inferences across conditions, but similar causal infer-

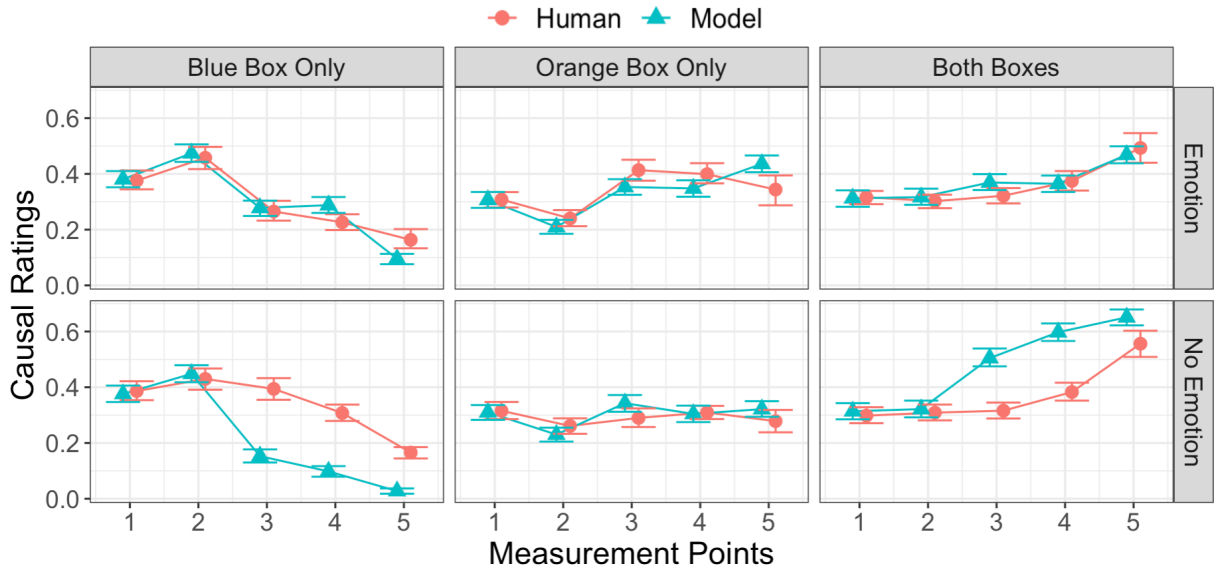

Figure 4: Study 1b results: mean causal ratings from human participants and model predictions across five time points for each of three possible causal hypotheses (columns), by condition (rows). Error bars represent 95% confidence intervals.

ences. Even though the agent signalled that they expected the other bulb to turn on in the emotion condition, observers are likely to infer that the agent's beliefs are false since they contradict direct (physical) observations which are often more reliable (Shafto, Eaves, Navarro, & Perfors, 2012).

Participants. We recruited 147 participants on Prolific, excluding 22 who failed our attention checks, resulting in a final sample of 125 (40.1% females; $M_{age} = 28.1$; $SD_{age} = 11.0$).

Materials. The videos (Figure 2) depict an agent, a switch, and two light bulbs—one pink and one orange. The agent pushes the switch just before the orange bulb lights up. In the emotion condition, the agent additionally displays a frustrated expression. The agent then exits the room.

Procedure. Participants were randomly assigned to the emotion (N=68) or no-emotion (N=57) conditions. After watching the video, they rated different causal hypotheses (causal relationship between the switch and pink bulb; switch and orange bulb), the agent's beliefs (how likely the agent expected the pink/orange bulb to light up), desire (how likely the agent wanted to switch on the pink/orange bulb), and knowledge (how likely the agent had knowledge of the function of the switch). The causal and belief ratings were not normalized (unlike Study 1) to allow the two hypotheses to vary independently of one another (scoring high on both reflects the belief that both bulbs are causally related to the switch whereas scoring low on both reflects the belief that neither bulbs are causally related to the switch).

Study 2 Results

When presented only with the agent's actions in the no-emotion condition, participants were more likely to infer that the agent believed that the switch caused the orange bulb to turn on—presumably, via the rational agent assumption, participants may have assumed the agent pushed the switch in order to turn on the orange bulb (and succeeded). By contrast, when the agent displayed a frustrated expression (in the emotion condition), participants were instead more likely to infer that the agent believed that the switch activated the (other) pink bulb—presumably they intended to turn on the pink bulb as they were dissatisfied with the outcome. In support of this, we found that both conditions significantly differed in their belief inferences. Participants in the no-emotion condition (compared to emotion) were more likely to infer that the agent expected the orange bulb to turn on ($W = 444.0, p < .001$) whereas participants in the emotion condition (compared to no-emotion) were more likely to infer that the agent expected the pink bulb to turn on ($W = 3309.0, p < .001$).

Despite these different inferences over the agent's beliefs, participants across both conditions inferred that the switch was causally related to the orange bulb but not the pink bulb. There was no statistically significant differences in participants' ratings of the causal relationship between the switch and the pink bulb ($W = 1842, p = .624$) or between the switch and the orange bulb ($W = 1881.5, p = .745$).

Our computational model reflected these qualitative patterns (see Figure 5). We found that our model's predictions were highly correlated with participants' mean ratings of the *W*orld causal inference, and agent's *B*elief, *D*esire, and *K*nowledge ($r(12) = .918$, $p < .001$).

We compared our model against lesioned models by comparing their predictions for Studies 1a and 2 to assess our model's contribution over previous models that do not include emotions and/or knowledgeability (Goodman et al., 2009; Teo & Ong, 2021; Ong et al., 2019). Upon lesioning emotions, the correlation between human inferences and model

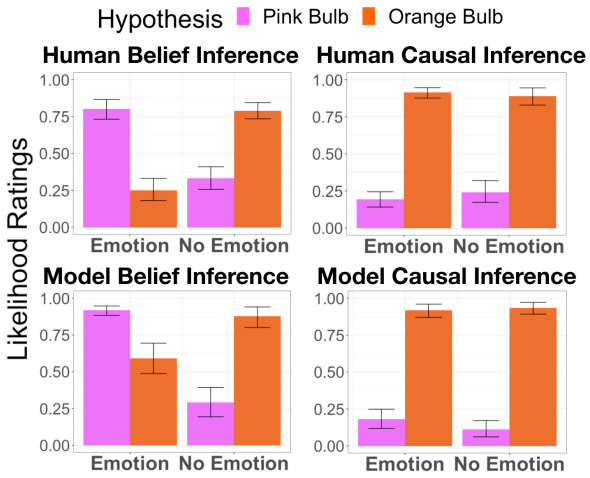

Figure 5: Study 2 results: inferences in a multiple-outcomes scenario after observing the orange bulb lighting up. Mean causal and agent-belief ratings across conditions for 2 causal hypotheses (top row). Corresponding model predictions are presented below. Error bars represent 95% confidence intervals.

predictions (on W and B) dropped from .91 to .67 (Fisher's Z test, $p = .045$). Upon lesioning knowledgeability (fixing it to 1 such that $W = B$), the correlation between human inferences and model predictions also dropped from .91 to .67 ($p = .043$). Without considering the emotions or knowledgeability of agents, observers would incorrectly assume that observed actions and outcomes are all intended by the agent, making it difficult to decipher any failed actions by the agent.

Causal Learning from Emotional Displays

Through emotional displays, people can draw insights about others' mental states, including their beliefs, desires (Ong et al., 2015; Wu, Baker, et al., 2018; Wu & Schulz, 2018), and knowledge (Wu, Haque, & Schulz, 2018; Wu, Schulz, & Saxe, 2018). People can also make causal inferences about the world merely by observing others' actions (Goodman et al., 2009; Meltzoff et al., 2012; Teo & Ong, 2021). In this paper, we presented a computational model showing how these processes can be integrated within a coherent Bayesian framework. Across three studies, we showed how both people and our model used emotional displays to disambiguate social situations, resulting in different inferences over agent beliefs and causal relationships: in particular, we also demonstrated that our model's inferences closely track people's causal inferences as the scene unfolds.

There were some notable differences between our model's predictions and human judgments, which provide interesting observations about people's reasoning. In the no-emotion condition in Study 1a, our model made more 'polarized' inferences compared to people in the no-emotion condition, attributing much less weight to the "blue switch only" hypothesis and much more weight to the "both switches" hypothesis. From Study 1b (Fig. 4), we see that these deviations began after the agent performed their first action (with no outcome; Measurement Point 2 $\rightarrow$ 3), where the model starts to dismiss the "blue switch only" hypothesis (the action at Point 2 that failed) in favor of the "both switches" hypothesis. This is sensible because the (lack of an) outcome shows clear evidence against the blue switch as a causal antecedent. But people are somehow less able to make that inference from the absence of an outcome. Contrast this with the same point in the emotion condition where the agent displays a frustrated expression at Measurement Point 3: people did dismiss the "blue switch only" hypothesis then. We know that (personal) failure is a strong motivator of causal search (Weiner, 2012) and perhaps, watching another person fail (unambiguously, given the emotion display) might give similar motivation for observers to reason about the absence of an outcome.

Similar to past work on testimonial learning (Bridgers, Buchsbaum, Seiver, Griffiths, & Gopnik, 2016), we observed an asymmetry in people's reliance on direct observations compared to others' knowledge in Study 2: people did not rely on testimonial evidence when it contradicted direct observation. This stems from the uncertainty an observer has about another person's knowledge compared to the certainty of physical evidence (Shafto et al., 2012). This account implies that increasing confidence in the expertise of testimonial evidence, or lowering the certainty of physical evidence, may lead to people trusting others' knowledge more than physical evidence. Bridgers et al. (2016) demonstrated the latter in a context where event relationships are probabilistic, and in which people's inferences relied on testimony even when these reports contradicted what was observed. Future work should examine how such findings translate to an observational learning context, where there is uncertainty not just in the agent's reliability, but also in their mental states (emotions, beliefs, desires). It might also be worth investigating certain emotions (e.g., confidence or anxiety) that may additionally convey information about an agent's expertise.

Our model's treatment of emotions is still simple. So far, we have only considered goal-related appraisals leading to frustration and happiness: other emotions like surprise or excitement could also provide information about expectations and facilitate causal learning. We have also not considered other contextual factors: In our scenarios, frustration in context suggested that the agent had false beliefs about the causal structure. However, in situations where people can fail due to other factors, frustration could signal that the agent is knowledgeable (e.g., frustration from failing to answer a test question due to lack of time, or failing to push the correct switch due to lack of strength). Much more work can be done to develop models of emotion understanding across contexts and to integrate such information into causal learning.

In summary, our work shows how emotion displays—via a rich intuitive theory of appraisals—provide information beyond mental states to causal relationships in the world, broadening our understanding of observational causal learning.

Acknowledgments

This research was funded by a Singapore Ministry of Education Academic Research Fund Tier 1 grant to DCO.

References

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10.

Bonawitz, E., & Shafto, P. (2016). Computational models of development, social influences. *Current Opinion in Behavioral Sciences*, *7*, 95–100.

Bridgers, S., Altman, S., & Gweon, H. (2017). How can I help? 24-48-month-olds provide help specific to the cause of others' failed actions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *39*, 162–167.

Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology*, *52*(1), 9.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367.

Gardiner, A. K. (2014). Beyond irrelevant actions: Understanding the role of intentionality in children's imitation of relevant actions. *Journal of Experimental Child Psychology*, *119*, 54–72.

Gardiner, A. K., Greif, M. L., & Bjorklund, D. F. (2011). Guided by intention: Preschoolers' imitation reflects inferences of causation. *Journal of Cognition and Development*, *12*(3), 355–373.

Goodman, N. D., Baker, C. L., & Tenenbaum, J. B. (2009). Cause and intent: Social reasoning in causal learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*, 2759–2764.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620.

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*(6037), 1524–1524.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*(9), 678–683.

Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology*, *48*(5), 1215.

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141–162.

Ong, D. C., Zaki, J., & Goodman, N. D. (2016). Emotions in lay explanations of behavior. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *38*, 360–365.

Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational models of emotion inference in Theory of Mind: A review and roadmap. *Topics in Cognitive Science*, *11*(2), 338–357.

Saxe, R., & Houlihan, S. R. (2017). Formalizing emotion concepts within a Bayesian model of Theory of Mind. *Current Opinion in Psychology*, *17*, 15–21.

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*(3), 436–447.

Teo, D. W., & Ong, D. C. (2021). Learning from Agentic Actions: Modelling Causal Inference from Intention. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 2478–2484.

Weiner, B. (2012). *An attributional theory of motivation and emotion*. Springer Science & Business Media.

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive Science*, *42*(3), 850–884.

Wu, Y., Haque, J., & Schulz, L. (2018). Children can use others' emotional expressions to infer their knowledge and predict their behaviors in classic false belief tasks. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *40*, 1193–1198.

Wu, Y., Schulz, L., & Saxe, R. (2018). Toddlers connect emotional responses to epistemic states. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *40*, 2711–2716.

Wu, Y., & Schulz, L. E. (2018). Inferring beliefs and desires from emotional reactions to anticipated and observed events. *Child Development*, *89*(2), 649–662.

Wu, Y., Schulz, L. E., Frank, M. C., & Gweon, H. (2021). Emotion as information in early social learning. *Current Directions in Psychological Science*, *30*(6), 468–475.