

# HUMAN DYNAMIC CONTROL UNDER CHANGING GOALS

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

One practical reason that intelligent agents might learn to represent causal structure is that it enables flexible adaptation to a changing environment. For example, a causal model can enable rapid generalization of behavior in light of changing circumstances or goals. In this project we examine human goal flexibility when interacting with dynamic environments. Contrary to our predictions, information about changing goals affected neither participant ability to infer causal structure nor participant success in controlling the dynamic environment. These findings were corroborated by participants being better fit by models describing them as utilizing minimally complex, reactive control policies. The results show how despite being incredibly adaptive, people are in fact computationally frugal, minimizing the complexity of their representations and decision policies even in situations that might warrant richer ones.

## 1 INTRODUCTION

Considerable effort in machine learning has focused on how to infer causal structure between variables. The value of this approach is often framed as enabling human interpretable systems. However, the way that humans learn about and reason with causality remains poorly understood. While it has been shown that in some sufficiently simple scenarios people are able to infer abstract causal structure (Sloman, 2005), what is not known is the extent to which people leverage causal structure in more complex, time-varying, continuous settings. For example, operating room doctors receive continuous readings of blood pressure, heart rate, and respiration and must use their understanding of how these factors influence each other to take actions to keep patients healthy. In this project we study the conditions under which people spontaneously infer causal structure in continuous dynamic systems. Our hope is that studying the cognitive mechanisms that people employ in these situations can give insight into building more adaptive and intelligent machines.

The central issue in this report concerns how complete people’s representations of the environment are. For example, some recent work has found that people represent only the minimal amount of structure needed to achieve their goals (Davis et al., 2018). In other words, instead of developing complete models of causal systems people seem to learn the minimal amount necessary to achieve good performance in a given task. However, one argument is that people will adopt more comprehensive learning strategies when they must use their causal knowledge in multiple ways. To evaluate this claim, we attempted to induce spontaneous causal learning by giving people expectations about either static or changing goals, under the theory that having a richer representation of the environment is of higher utility when you must act in service of multiple goals rather than one.

## 2 METHOD

**Participants.** 100 participants (66 male; age  $M = 35.5$ ,  $SD = 9.7$ ) were recruited from Amazon Mechanical Turk. For a task of approximately 30 minutes they were paid \$3, with additional opportunity for bonus ( $M = 0.57$ ,  $SD = 0.52$ ). Participants were randomly assigned to one of two conditions, with 50 in each.

**Behavioral task.** During the dynamic control task (Fig. 1A), participants interacted with a system composed of three slider widgets arranged in a triangle. Participants initiated a 20 second trial by

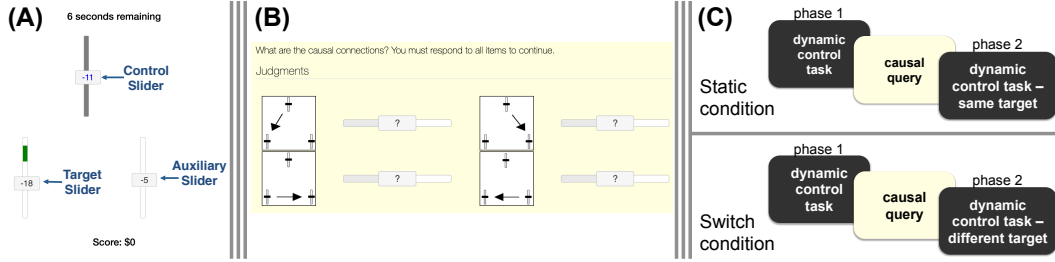


Figure 1: Experiment design. (A) Dynamic control task. Control slider remains the same across the experiment, Target Slider and Auxiliary Slider randomly chosen per trial. Green reward region also randomly chosen. (B) Causal query. Participants categorized the four potential causal links into “regular”, “none”, or “inverse” causal links. (C) Experiment flow. In the Switch condition, the Auxiliary and Target Slider were switched in phase 2 of the dynamic control task.

pressing the “Start” button at the top of the page, whereupon the sliders began updating according to an Ornstein–Uhlenbeck process (described below) at 100 ms increments. Participants could intervene on one slider (the ‘control slider’, top) by pressing the ‘o’, ‘k’, and ‘m’ keyboard buttons to make it move up by 10 arbitrary units, hold steady, or move down by 10, respectively. Manipulating the control slider exerted causal influence on the two bottom sliders, and participants acted to try to influence the (randomly selected) target slider to remain within the green reward region. Doing so required continuous adjustment and reaction to the dynamics of the system.

After phase 1 of the task, participants reported the causal structure of the system they were interacting with (Fig. 1B) by rating the causal strength of different pairwise links between the slides. Upon responding to all potential causal links, participants initiated phase 2, which was either identical to phase 1 (Static condition) or the Target and Auxiliary Sliders were switched (Switch condition). See Fig. 1C. Participants were tested on 10 different causal graphs (patterns of causal influences between the sliders) that were roughly balanced across factors such as the number of inverted and regular links and the number of links between each variable. These graphs were presented in random order.

**Critical manipulation.** Prior to the start of the experiment, participants completed an interactive instruction that showed them demonstrations of the types of causal links (regular, inverse, and none, see below) and instructed them that they would be rewarded \$.01 for every 100ms the target variable was within the target region. In addition, those assigned to the switch condition were instructed that the target slider would change between phases. This is the critical manipulation in the study: do expectations about the system changing in the future alter the type of causal learning that participants spontaneously engage in while controlling the dynamic systems.

**Causal Systems.** The stimuli in our task were generated using a new approach for simulating continuous causal systems. Briefly, Ornstein–Uhlenbeck (OU) networks represent causality with autoregressive processes that move towards a basin point as a function of time (see Davis et al. (in press) for a full explication of the generative process). When one variable is causally influenced by another variable (as defined by the causal structure of the OU network), this is modelled by making the effect’s basin point nonstationary, following some function of the state of its cause(s). We here restrict these functions to the effect either asymptoting to a value equal to the cause’s value (“regular” connections) or to the inverse of the cause’s value (“inverted” connections). If a variable has more than one cause, we model it as attracted to the sum of the basin points defined by each of its causes. To accommodate interventions we use the “Do( )” operator (Pearl, 2009), whereby an intervened on variable takes the assigned value with a probability of 1, ignoring all other endogenous or exogenous factors.

### 3 RESULTS

We first analyze performance on the dynamic control tasks. On average, participants gathered 59.7 (SD=32.0) points out of a theoretical maximum of 200. A 2x2 ANOVA was performed to assess the

impacts of phase and condition on the amount of reward participants gathered, finding no evidence of main effects for either condition ( $F(1, 196) = .35, p = .56$ ) or phase ( $F(1, 196) = 1.95, p = .16$ ). In addition, there was no evidence of an interaction between phase and condition,  $F(1, 196) = .04, p = .84$ .

On the explicit causal queries that followed phase 1 of the dynamic control task, participants were above chance (.33) at identifying causal links ( $M=.43, SD=.10$ ),  $t(99) = 10.3, p < .001$ . However, contrary to our prediction, we did not find evidence of a difference between the static ( $M = .44, SD = 0.10$ ) and switch ( $M = .42, SD = 0.09$ ) conditions,  $t(98) = .72, p = .48$ .

## 4 MODELING

The descriptive results suggest that our manipulation of expected goal flexibility did not elicit different structure learning strategies. We now test this result by comparing people to models that make different assumptions about the extent to which people represent causal structure in our task.

**Proportional-Integral-Derivative controller.** The Proportional-Integral-Derivative (PID) controller acts on a function of its error between desired and observed outcomes. It has been successful in accounting for human control behavior (Davis et al., 2018; Ritz et al., 2018), although it has not been tested in environments with changing goals. The PID controller computes a history of distances from its state to a desired setpoint, and performs some simple operations on this buffer. In particular, it computes a weighted sum of (P): its current distance to setpoint, (I): its history of distances to the setpoint over some buffer, and (D): the derivative of its movement towards the setpoint, computed as the current error minus the error on the previous timepoint. The PID recommends an action with the same sign as the sum of its errors (e.g. a large positive sum would recommend the action ‘o’ in our task). One additional complication of our task is that the PID must learn the relationship between actions and outcomes. We model this as a simple timelagged correlation between control and target variables, of the form  $\rho(\text{Control}_{t-1}, \text{Target}_t)$ . The sign of this learned correlation modifies the sign of the error term, such that a negative correlation negates the computed sum of errors.

**Model-based controllers.** We also compare participant actions and causal judgments to two model-based controllers—the Causal Model Based Controller (CMBC) and the Local Computations (LC) controller. The CMBC agent, at each timepoint, inverts the generative OU model to optimally estimate the probability of there being causal connections between sliders (see Davis et al., in press). The LC model is more limited, evaluating the individual contribution of each variable to each other variable’s value, ignoring the possible contributions of other links. In other respects the LC model is identical to the normative model. The causal models use their online estimate of the causal structure of the environment to act. In particular, at each timepoint they elaborate a full decision tree of possible actions (‘o’, ‘k’, ‘m’, or nothing) over the next three timepoints, and assign an expected value to each action that reflects the maximum expected reward over all following decisions.

**Model evaluation.** We evaluate the models on two criteria. The first is on their ability to predict the participants’ explicit causal judgments. For this metric, the two model-based controllers have a single per participant softmax  $\tau$  parameter of their posterior distribution over causal graphs, fitted to maximize the log-likelihood of participant causal judgments. The PID model’s two fitted per participant parameters were (a) a significance threshold, where the PID infers a causal link if the correlation between control and target variable drops below a fitted significance level, and (b) a lapse rate, which models participants as responding according to PID’s predictions with some probability. The PID learns nothing about other variables, and so we assume random responding on all other potential links. For the dynamic control task, all models were fit with (a) a perseverance parameter reflecting the probability of repeating the action made on the previous timepoint, and (b) a softmax  $\tau$  parameter, which maximizes the log-likelihood of participant actions. For the model-based controllers these operations were performed on normalized expected values. For the PID these operations were performed on a one-hot vector of its favored action at that timepoint.

**Modeling results.** Fig. 2 shows the modeling results, clustered into causal learning performance (panels A and B) and dynamic control performance (panels C and D). The PID model provided the best account of participant judgments of causal structure, both in overall BIC as well as per participant fits. These results confirm that participants attended to the direct impact of their actions on the control variable but not the system’s causal structure. The PID was also the most accurate in

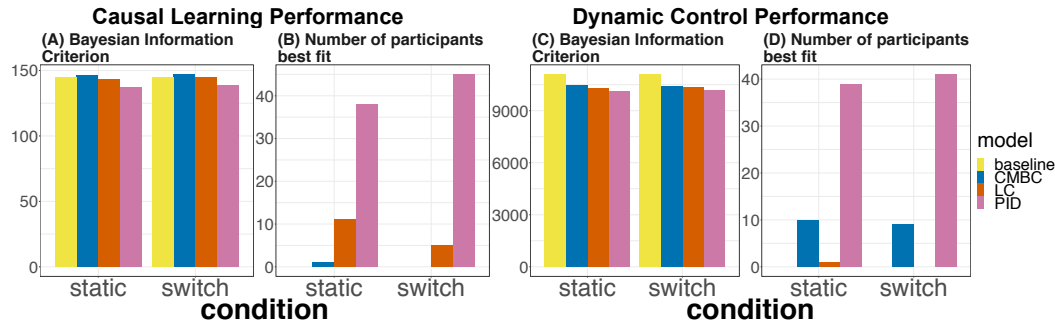


Figure 2: Modeling results. (A) Mean Bayesian Information Criterion value per participant on the Causal Query. (B) Number of participants best fit as a function of BIC. (C) Mean Bayesian Information Criterion value per participant on the Dynamic Control Task. (D) Number of participants best fit as a function of BIC.

predicting participant actions, according to both BIC and the number of participants best fit (Figs. 2C and 2D). As in the descriptive results, the modeling results reflect no differences across conditions.

## 5 DISCUSSION

Learning causal structure is important to acting effectively in the world. In this project we attempted to explore under which conditions people will invest their cognitive resources to do so, especially in the context of a dynamic control task. We specifically attempted to manipulate expectations of changing goals under the idea that learning robust representations is motivated by the need to generalize to new situations. Our results gave little evidence that this manipulation fundamentally altered the representational strategy of our human participants. Instead, people seemed to use a simpler, reactive control policy (a PID-type controller).

An important factor that was likely to have contributed to this result is the simplicity of the task. In general, we suspect that subjects found discovering the relationship between the control and target variables relatively straightforward. As a result, the potential performance advantage of learning the system’s causal structure during phase 1 so that it could be applied to performance during phase 2 was likely to be modest. Our conjecture is that, in this task, minimal representations were sufficiently rewarding to disincentivize more effortful causal-based strategies. These results suggest that people balance the efforts and rewards of developing causal knowledge. Building an understanding of the factors affecting people’s spontaneous causal learning may help machine learning models adaptively develop causal knowledge in response to situational factors such as potential for reward or computational limitations.

## REFERENCES

- Zachary J Davis, Neil R Bramley, Bob Rehder, and Todd M Gureckis. A causal model approach to dynamic control. In *Proceedings of the 40<sup>th</sup> Annual Conference of the Cognitive Science Society*, Madison, WI, 2018. Cognitive Science Society.
- Zachary J Davis, Neil R Bramley, and Bob Rehder. Causal structure learning in continuous systems. *Frontiers in Psychology*, in press.
- Judea Pearl. *Causality*. Cambridge university press, 2009.
- Harrison Ritz, Matthew R Nassar, Michael J Frank, and Amitai Shenhav. A control theoretic model of adaptive learning in dynamic environments. *Journal of Cognitive Neuroscience*, 30(10):1405–1421, 2018.
- Steven A Sloman. *Causal models: How people think about the world and its alternatives*. Oxford University Press, 2005.