Decomposing Human Causal Learning: Bottom-up Associative Learning and Top-down Schema Reasoning

Mark Edmonds[1,4] markedmonds@ucla.edu
Siyuan Qi[1,4] syqi@cs.ucla.edu
Yixin Zhu[2,4] yixin.zhu@ucla.edu
James Kubricht[3] kubricht@ucla.edu
Song-Chun Zhu[1,2,4] sczhu@stat.ucla.edu
Hongjing Lu[2,3] hongjing@ucla.edu

[1] Department of Computer Science, UCLA [2] Department of Statistics, UCLA [3] Department of Psychology, UCLA
[4] International Center for AI and Robot Autonomy (CARA)

Abstract

Transfer learning is fundamental for intelligence; agents expected to operate in novel and unfamiliar environments must be able to transfer previously learned knowledge to new domains or problems. However, knowledge transfer manifests at different levels of representation. The underlying computational mechanisms in support of different types of transfer learning remain unclear. In this paper, we approach the transfer learning challenge by decomposing the underlying computational mechanisms involved in bottom-up associative learning and top-down causal schema induction. We adopt a Bayesian framework to model causal theory induction and use the inferred causal theory to transfer *abstract* knowledge between similar environments. Specifically, we train a simulated agent to discover and transfer useful relational and abstract knowledge by interactively exploring the problem space and extracting relations from observed low-level attributes. A set of hierarchical causal schema is constructed to determine task structure. Our agent combines causal theories and associative learning to select a sequence of actions most likely to accomplish the task. To evaluate the proposed framework, we compare performances of the simulated agent with human performance in the OpenLock environment, a virtual "escape room" with a complex hierarchy that requires agents to reason about causal structures governing the system. While the simulated agent requires more attempts than human participants, the qualitative trends of transfer in the learning situations are similar between humans and our trained agent. These findings suggest human causal learning in complex, unfamiliar situations may rely on the synergy between bottom-up associative learning and top-down schema reasoning.

Introduction

The human capacity for inferring causal relations in unfamiliar environments is a hallmark of human intelligence (Mackie, 1974) that is often taken for granted in daily life. An illustrative example is that of the escape room—a prevalent social activity where groups of people inside of a locked room work together to complete sub-goals (puzzles) to achieve the goal—escape from the room. In order to succeed, teams must: (i) identify goal-relevant entities in the environment among distractors, (ii) develop a causal model for individual sub-goals, and (iii) interact with scene components to refine entity- and goal-based hypotheses. In this paper, we propose that inference in scenarios like the one above depends on two critical learning components. First, attributes relevant to candidate causal hypotheses are learned by interacting with entities in the scene, and second, causal hypotheses are refined based on newly encoded attribute-based knowledge.

It is worth noting that the above approach is generally inconsistent with early studies on causal learning in psychological research (Holyoak & Cheng, 2011). Early studies primarily focused on animal learning and conditioning experimental paradigms, framing causal understanding as learned stimulus-response relationships attained primarily through observation (*e.g.*, Shanks and Dickinson (1988)). Given associative weights on cue-effect links, the Rescorla-Wagner model was often utilized to explain how humans (and non-humans) construct expectations based on the co-occurrence of perceptual stimuli (Rescorla & Wagner, 1972). However, the knowledge that people have about causal mechanisms in the distal world has been shown to extend beyond the covariation between observed (perceptual) variables. For instance, adults interact with dynamic physical scenarios in ways that maximize information relevant to their causal hypotheses (Bramley, Gerstenberg, Tenenbaum, & Gureckis, 2018), and even infants test their beliefs about the physical characteristics of objects through exploration and experimentation (Stahl & Feigenson, 2015).

Contrary to the associative account, researchers have demonstrated that human learning and reasoning in novel (causal) environments rely heavily on the discovery of abstract causal structure (Waldmann & Holyoak, 1992) and strength (Cheng, 1997) rather than purely associative (statistical) dependencies. More recently, the integration of causal graphical models and Bayesian statistical inference (*i.e.*, Bayes nets) has provided a general representational framework for how this structure and strength is learned and transferred to novel situations (Griffiths & Tenenbaum, 2005, 2009; Tenenbaum, Griffiths, & Kemp, 2006; Bramley, Lagnado, & Speekenbrink, 2015; Bramley, Dayan, Griffiths, & Lagnado, 2017; Edmonds et al., 2018; Holyoak & Cheng, 2011). Under this framework, causal knowledge plays an essential role in constructing a flexible model of the world in which environmental states represent some status in the world, and connections between states imply the strength of a causal relationship.

We propose that creative discovery in novel domains relies on both causal structure *and* associations. Knowledge of causal structure enables agents to simulate how interventions will influence the environmental state, and without associations to guide exploration, the number of causal hypotheses to consider becomes intractable. For problem domains where the number of possible interventions is particularly high, the need for associative "guidance" can drastically improve decision-making. To solve this problem, we propose a computational model that integrates two learning mecha-

nisms: (i) a bottom-up process that determines which object attributes are causally relevant, and (ii) a top-down process that learns which abstract causal structures accomplish a task. The outcomes of actions are used to update the causal hypothesis space, and simulated agents learn a dynamics model capable of solving a challenging task.

We implement the proposed model in a virtual "escape room" environment where agents (human and artificial) are trapped in a room containing a single locked door and a set of conspicuous levers. The door of this room will unlock after the agent has interacted with the levers in a specific sequence. An agent placed in such a room may begin to randomly push or pull on the levers and revise their theory about the door's locking mechanism based on observed changes. Once an agent discovers a single solution, they are placed back into the same room and tasked with finding the next solution. The agent "escapes" from a room after finding *all* of the solutions which can be used to unlock the door.

After escaping from a room, agents are placed in a similar room but with newly positioned levers. Although the levers are in different positions, the new room is governed by the same abstract rules as the last (unknown to the agent). Thus, the agent's task is to identify the role of each lever in a new room. If the agent makes use of some knowledge from previous trials, we expect to observe fewer attempts in solving the problem. Because these rules are abstract descriptions of the latent state of the escape room, we refer to the underlying theory as a causal schema (*i.e.*, a conceptual organization of events identified as cause and effect; Heider, 1958). Once learned, this schema enables agents to transfer between different arrangements of levers in the room. The present work models the causal learning process from a hierarchical Bayesian perspective and makes three major contributions:

1. Utilizes a bottom-up associative learning paradigm to determine which attributes of the scene contribute to causal relations.
2. Utilizes a top-down causal schema model of the generalized operation of the environment to quickly adapt to similar but new scenarios.
3. Leverages causal hypotheses to learn a world model capable of transferring knowledge between seemingly dissimilar but structurally and causally equivalent environments.

The remainder of the paper is structured as follows. First, the OpenLock environment and experimental procedure are described, followed by an analysis of human performance from Edmonds et al. (2018). Next, components of the proposed model are described and corresponding results are provided. Finally, the paper concludes with a discussion of results and directions for future work.

Experiment: OpenLock Task

Participants

A total of 160 undergraduate students (114 female; mean age= 21.6) from the University of California, Los Angeles (UCLA) Department of Psychology subject pool and were compensated with course credit for their participation.

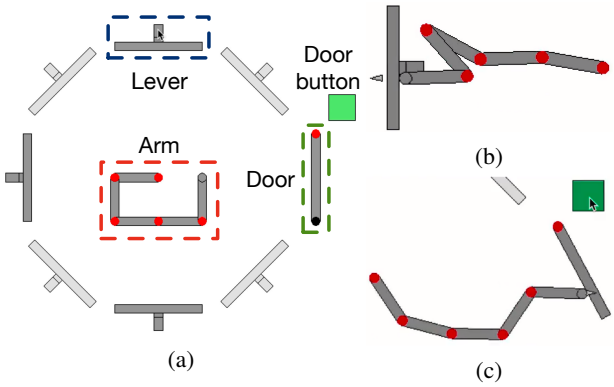


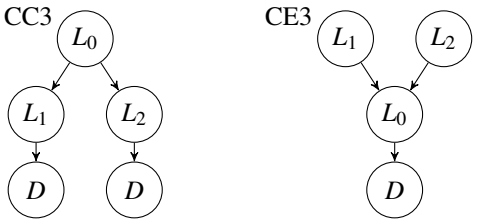
Figure 1: (a) Initial configuration of the room containing three active levers. All levers begin pulled toward the robot arm located at the center of the display. The arm interacts with levers by pushing/pulling them outward/inward. Only push actions are needed to unlock the door in each room (unknown to agents). White levers never move; this information is not explicitly stated. Once the door is unlocked, the green button can be clicked to command the arm to push the door open. The black circle located opposite the door's red hinge represents the door lock indicator (present if locked, absent if unlocked). (b) Pushing on a lever. (c) Opening the door by clicking the green button.

Materials and Procedure

In this section, we outline the OpenLock task, initially presented in Edmonds et al., 2018. In the task, agents are required to "escape" from a virtual room by unlocking and opening a door. The door unlocks after manipulating the levers in a particular sequence (see Figure 1). Each room consists of seven levers surrounding a robotic arm that can *push* or *pull* on each lever. While a subset of the levers is always involved in the locking mechanism (*i.e.*, active levers; colored grey), other levers are not causally relevant (*i.e.*, inactive levers; colored white). Agents observe the color of the levers and are expected to *learn* that grey levers–but not white levers–are always part of solutions in each room. Importantly, agents are tasked with finding *all* possible solutions for opening the door within a room. Participants are explicitly told that their goal is to open the door and are informed of how many solutions they have remaining in this room.[1]

The mechanics underlying the environment obey one of two causal schemas: Common Cause (CC) and Common Effect (CE) (see Figure 2). Requiring agents to find all solutions within a specific room ensures that agents abstract CC or CE schema structures. While a single solution corresponds to a single causal chain, a schema relies on nodes that are shared between multiple chains. Agents operate under a movement-limit constraint, where only three movements can be used to either (i) *push* or *pull* on levers (active or inactive), or (ii) *push* open the door. This constraint was placed on the agent to confine the search depth of possible solutions. After three movements, the episode terminates and the environment re-

[1]The video instructions presented to participants can be viewed at `https://vimeo.com/265302423`



(a) (b)

Figure 2: Common Cause (CC) and Common Effect (CE) structures used in the OpenLock task, in which L_i indicates a lever in the scene, and D indicates the effect of opening the door. In (a) CC3 and (b) CE3 condition, both include three causal cues but with different causal structures.

sets, regardless of the outcome. Agents also operate under a limited number of episodes (30) in a particular room, regardless of whether all solutions are found. We denote three movements as an *attempt* and each room as a *trial*. After completing a trial, agents move to a new trial (*i.e.*, room) with the same underlying causal schema but a different lever arrangement. This setup ensures that agents do not overfit their understanding of the environment to a single trial; *i.e.*, if agents are forming a useful abstraction, the knowledge they acquired in previous trials should aid in their ability to find all solutions in new trials. Note that in a 3-lever room, an optimal agent should produce both solutions within 3 attempts. One attempt may be used to identify the role of the observed levers in the abstract structure, and the remaining attempts are used for each solution.

Human Results

The analyses reported herein expand on previous behavioral findings by examining the number of attempts needed to find *each* solution rather than accumulating *all* solutions (see Human Data, Edmonds et al., 2018). The purpose of this exploration was to tease apart the separate learning components involved in the OpenLock task. Participants who failed to find all solutions in the allotted maximum number of attempts in *any* trial were removed from the analysis (24 participants removed from each condition). Eighty human participants were assigned to each condition (CC and CE).

We first examined whether the number of attempts needed to find each solution varied across trials. The behavioral data from each experimental condition is depicted in Figure 4. For participants who trained under a Common Cause (CC) schema, attempts needed to find the *first* solution decreased significantly following both the first trial ($t(55) = 6.80; p < .001$) and second trial ($t(55) = 2.52; p = .02$). First solution attempts also showed a marginal decrease following the fifth trial ($t(55) = 1.99; p = .051$). For the *second* solution, the number of attempts needed decreased significantly following the first trial only ($t(55) = 4.40; p < .001$). A similar trend was observed for participants assigned to the Common Effect (CE) condition—attempts needed to find the *first* solution decreased following the first trial ($t(55) = 5.30; p < .001$) and third trial ($t(55) = 2.19; p = .03$), and attempts needed to find the *second* solution decreased following the first trial only ($t(55) = 2.36; p = .02$).

The human results demonstrate that regardless of which causal schema participants trained with, significant learning appeared to occur in the early trials for both the *first* and *second* solution. However, the learning rate for the *first* solution was much faster, and the learning rate for the *second* solution was relatively less pronounced. In the next sections, we describe our computational approach and report whether it can account for human performance.

Model Details

We begin by describing our agent's process for combining top-down (abstract) causal knowledge with bottom-up (associative) attribute knowledge. The agent decides which action to perform by (i) computing the posterior probability of each candidate causal chain and (ii) making a selection using the computed posterior and a model-based planner.

Causal Theory Induction: To explain trends in human performance, we follow a Bayesian account of how hierarchical causal theories can be induced from data (Griffiths & Tenenbaum, 2005, 2009; Tenenbaum et al., 2006). The key insight in this framework is that hierarchy enables abstraction, and theories provide general background knowledge about a task or environment at the highest level. Theories consist of principles; for example, an analysis of evolutionary traits between species can be represented with a taxonomic tree and mutation processes (example from Tenenbaum et al. (2006)). Principles lead to structure; for example, a tree describing how primates evolved and split into species over time. Finally, structure leads to data; such as shared genes among primates.

The goal of this work is to model a human decision-making process where agents are required to learn *transferable* knowledge between different yet similar environments. We approach the problem from the perspective of *active* causal theory learning, where we expect an agent endowed with no information to learn the underlying abstract mechanics and commonalities between environments through interaction. This approach naturally places the focus of the learning task on how the agent decides the best action to take next and how to effectively integrate the results into the agent's model of the world.

In this work, we adhere to two general principles of learning: (i) *causal relations induce state changes in the environment, and non-causal relations do not* (referred to as our bottom-up β theory), and (ii) *causal structures that have previously been useful may be useful in the future* (referred to as our top-down γ theory). Specifically, the environment provides a set of attributes, such as position and color, and our agent learns which attributes are associated with levers that induce state changes in the environment. Our agent also learns a distribution over abstract causal structures (*i.e.*, schemas) that provide generalized notions of task structure.

We define a causal chain hypothesis space, Ω_C, over possible causal chains, $c \in \Omega_C$. Figure 3b shows the structure of the causal chain. Each chain is defined by a tuple of subchains $c = (c_0, \ldots, c_k)$, where each subchain is defined as a tuple $c_i = (a_i, s_i, cr_i^a, cr_i^s)$. Each a_i represents an action node that the agent can intervene on (execute), and the space of actions, Ω_A, consists of pushing and pulling on every lever and

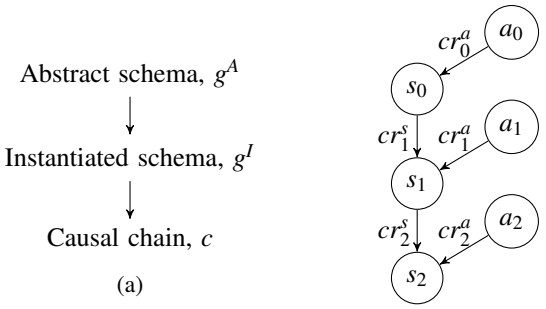


Figure 3: (a) An illustration of hierarchical structure of the model. A bottom-up associative learning theory, β, and a top-down causal theory, γ, serve as priors for the rest of the model. The model makes decisions at the causal chain resolution. (b) Atomic causal chain. The chain is composed by a set of sub-chains, c_i, where each c_i is defined by: (i) a_i, an action node that can be intervened upon by the agent, (ii) s_i, a state node capturing the time-invariant *attributes* and time-varying *fluents* of the object, (iii) cr_i^a, the causal relation between a_i and s_i, and (iv) cr_i^s, the causal relation between s_i and s_{i-1}.

pushing on the door. Each s_i represents a state node. The state node is defined as a tuple, $s_i = (\phi_i, f_i)$, where ϕ_i is a vector of time-invariant *attributes* and f_i is a vector of time-varying *fluents*. The state node is influenced by taking action a_i according to the causal relation cr_i^a and may be affected by a previous state node through the causal relation cr_i^s. For instance, in Figure 1a and Figure 3b, the action *push* for the leftmost lever may transition the lever from the fluent *pulled* to *pushed* through cr_0^a, which in turn transitions the uppermost lever from *locked* to *unlocked* according to cr_1^s.

The space of attributes is denoted as Ω_ϕ, consisting of position and color. The space of fluents, Ω_F, consists of binary values for lever status (*pushed* or *pulled*) and lever lock status (*locked* or *unlocked*). The space of states is defined as $\Omega_S = \Omega_\phi \times \Omega_F$. The space of causal relations is defined as $\Omega_{CR} = \Omega_F \times \Omega_F$, capturing the possibly binary transitions between previous fluent values and the next fluent values.

We assume agents can directly intervene on (*i.e.*, control) *actions*, but cannot directly intervene on *fluents*. This distinction adds significantly more complexity to the causal chain hypothesis space but means that we do not assume the effects of actions, nor do we assume an agent can directly intervene on the value of a particular fluent. We assume that an agent can execute any action within the action space (through an intervention on the action node in the causal chain), but how that action affects the state of the world must be learned (*i.e.*, the effects of the actions are learned).

Decomposing states into time-invariant *attributes* and time-varying *fluents* aids in the computational complexity of learning and inference; our agent assumes attributes cannot be changed by actions or other states. In addition, because the attributes are time-invariant, attributes offer a grounding upon which the agent can learn knowledge, regardless of the executed action sequence or lever configuration. In contrast, the fluents are time-varying and include the latent state of the lever's internal locking mechanism; *i.e.*, *locked* or *unlocked*. The agent learns how to influence these latent states through observational cues about which attributes are associated with a particular fluent. Attributes are defined by low-level features of an object, *e.g.*, position, color, shape, orientation, *etc.*. These low-level attributes provide general background knowledge about how specific objects change under certain actions (for instance, which levers can be pushed or pulled).

A background theory encodes general knowledge that can be used to induce or evaluate a structural representation. We use two background theories—one for bottom-up features, denoted β, to learn beliefs about which attributes of objects indicate the object can be interacted with to produce a causal effect. This low-level knowledge about object attributes and their propensity to be involved in causal relationships provides information to transfer between similar but different environments governed by common underlying dynamics. The second background theory provides a top-down abstraction, denoted γ, that assumes tasks have similar causal structure across slightly different environments; *i.e.*, changes in the observable environment do not alter the underlying causal structure of a task.

Attribute Learning: Attributes provide time-invariant properties of an object. Categories of objects often share common attributes; *e.g.*, all cups share a common shape, all stop signs are red, *etc.*. However, objects in a category may vary in their physical form but share common functionality; for instance, light switches come in a number of shapes and sizes, but all examples share a common mechanism to transit between states.

We learn which attributes are relevant to our causal hypotheses via a Bayesian learning process, based on our assumption that causal relationships induce state changes. Therefore, an object changing states under an action indicates that the object's attributes may be related to a causal relationship. These attributes provide generalization clues for the agent, such as insights into which low-level attributes indicate that the corresponding object is part of a solution. This knowledge is invariant across trials and causal schemas.

The agent's belief in an attribute being causal is modelled with a multinomial distribution $\text{Mult}(\theta)$ parameterized by θ. The posterior distribution of θ given observed data $\mathbf{X}$ and the bottom-up theory β follows a Dirichlet distribution: $p(\theta|\mathbf{X};\beta) = \text{Dir}(\alpha')$, where α' is given by a maximum a posteriori (MAP).

Attributes are learned in two different time scales: a global timescale to learn attributes across all trials (between trials) and a local timescale to learn attributes specific to this trial (within trials). This separation allows the agent to adapt quickly to trial-specific knowledge while maintaining a global understanding across all trials. In each timescale, we perform this attribute learning in the following steps: (i) draw a sample (produce an observation by selecting an intervention and observing the result), (ii) accept the sample if the environment changed state in any way (*i.e.*, there was an effect from the intervention), and (iii) increase α of each attribute's Dirichlet distribution according to observed outcome.

A Dirichlet distribution, $\text{Dir}(\alpha^G)$, is used to model the posterior of the global attribute distribution. After finishing a

trial, the agent's global Dirichlet parameters, α^G, are updated to incorporate the observed data within a trial.

For each trial, we initialize the parameters of the local attribute Dirichlet distribution, $\text{Dir}(\alpha^L)$, with a scaled sample from the global Dirichlet, $\alpha^L = k\theta$, where $\theta \sim \text{Dir}(\alpha^G)$. This scaling factor k reduces the variance and enables fast adaptation of the agent's local attribute beliefs. In our experiments, we set k to initialize the local Dirichlet to have $\alpha^L \in [1, 10]$.

We introduce an additional variable, ρ to represent a casual event according to our background theory β; *i.e.*, that causal events induce state changes in the environment. We use a local prior over attributes as our bottom-up associative learning theory. We compute the likelihood that the attributes of a particular chain c are causally relevant given a background theory β as:

$$p(\rho|c;\beta) = \prod_{c_i \in c} p(\rho_i|c_i;\beta), \tag{1}$$

where $p(\rho_i|c_i;\beta)$ is computed as

$$p(\rho_i|c_i;\beta) \propto \prod_{\substack{\phi_{ij} \in s_i \\ s_i \in c_i}} p(\rho_i|\phi_{ij};\beta) \tag{2}$$

where ϕ_{ij} is the j-th attribute from the i-th subchain. The term $p(\rho_i|\phi_{ij};\beta)$ represents the probability that attribute ϕ_{ij} adheres to the background theory β. Here, β represents the probability that attribute ϕ_{ij} is associated with objects that induce state changes. Note that $p(\rho_i|\phi_{ij};\beta)$ is parameterized by a sample from the local attribute Dirichlet distribution. After finishing an attempt, we update the parameters α^L of the local distribution to incorporate the outcome of the attempt and resample θ.

Recall our associative theory: causal relationships induce state changes in the environment; practically, $p(\rho_i|\phi_{ij};\beta)$ represents the probability that attribute ϕ_{ij} is associated with objects that produce state changes, under the assumption these attributes are independently associated with causal events. In our domain, an agent using this theory should learn that grey levers are involved in causal events and white levers are not. Additionally, the agent should initially believe that position is an important attribute for detecting causal relationships. However, as the agent observes multiple configurations of levers with different positions of grey levers, every position will be involved in causal events, and therefore this belief should approach the uniform distribution.

This bottom-up inference enables agents to leverage low-level associative information about causal relationships. We then transfer this belief between trials, thereby enabling our agent to leverage the knowledge acquired in one trial to transfer to the next trial. The agent updates its belief regarding which attributes it believes are causal after each attempt.

Abstract Schema Learning: Learning attributes that correspond to causal cues is critical for an agent expected to learn how an environment operates. However, many environments share common high-level abstract causal structures. For instance, switches come in all different shapes and sizes tailored to specific tasks—from a light switch to a circuit breaker to a railroad switch. Each of these domain-specific mechanisms share a common abstract functionality—changing the state of some object from one discrete state to another.

We propose a model to learn abstract structural models that can be used to instantiate domain-specific models to achieve a task in an environment. This abstract knowledge is assumed to be useful across domains, and agents may acquire a collection of useful abstract models of different functionality. Our model considers learning abstract knowledge as a form of model selection, where the agent hypothesizes a space of potential abstract structures and updates the beliefs in those abstract structures based on its experience in the environment.

More specifically, we consider an abstract causal schema, g^A, from a hypothesis space of abstract schemas, Ω_{G^A}, to be a structural description of some causal relationships (see Figure 2). The space Ω_{G^A} is enumerated in this work; *i.e.*, all possible structural combinations of $N = 2$ trajectories (*i.e.*, causal chains) with length $K = 3$ are considered (since there are two solutions and three actions per attempt). We introduce a prior over abstract schemas, $p(g^A;\gamma)$, that is a multinomial distribution parameterized using a sample from the abstract schema Dirichlet distribution, $\text{Dir}(\alpha^A)$. After completing a trial, the abstract schema that encodes the solutions found in this trial receives a parameter update in the Dirichlet distribution—*i.e.*, an increase to the solution abstract schema's α^A.

These abstract structures are not bound to any particular instantiation of attributes, states, or actions. Instead, they encode common structural properties under varying instantiations—knowledge that may be useful when an observational setting is changed. In our task, abstract schemas encode the abstract structures, some of which are useful for solving OpenLock (*i.e.*, CC or CE), and we should expect agents to have a biased prior towards these structures.

Next, we consider an instantiated schema, g^I, to be a composition of causal chains, $c \in \Omega_C$. Instantiated schemas share the same structure as abstract schemas, but contain specific assignments for each a_i, s_i, cr_i^a, and cr_i^s of each subchain in the schema. We compute the belief in an instantiated schema g^I according to the hierarchical structure in Figure 3a:

$$p(g^I|do(q);\gamma) = \sum_{g^A \in \Omega_{G^A}} p(g^I|g^A, do(q))p(g^A;\gamma), \tag{3}$$

where $do(q)$ represents an intervention where the agent performs q—the solutions found thus far, a set of action sequences $q = \{A_0, A_1, \ldots, A_n\}$, where A_i is an action sequence. The $do()$ operator is the intervention operation presented by Pearl (2009), which allows the agent to bias its top-down inference towards instantiated schemas that contain solutions already found. Next, we compute the top-down belief in a causal chain by summing over instantiated schemas that contain the chain:

$$p(c|do(q);\gamma) = \sum_{g^I \in \Omega_{G^I}} p(c|g^I, do(q))p(g^I|do(q);\gamma). \tag{4}$$

These terms enable top-down inference on which chain is most likely to adhere to instantiated schemas that reflect abstract causal structures that have been useful in the past.

Learning which abstract schemas were successful in previous trials can be leveraged when the agent faces a new room configuration with the same underlying abstract mechanism governing the lock.

Intervention Selection: We formulate our intervention selection as a combination of the top-down and bottom-up causal chain beliefs, and we consider our learning mechanisms, γ and β, to be independent. We compute the posterior of the chain based on our top-down belief and bottom-up likelihood, assuming a uniform prior $p(\rho)$:

$$p(c|\rho, do(q); \gamma, \beta) \propto p(c|do(q); \gamma) p(\rho|c; \beta). \quad (5)$$

Our agent maintains an explicit notion of the goal of the task—to open the door. Human participants were also told the precise goal of the task. Thus, we frame our intervention selection process as a form of model-based planning. Our agent seeks to infer the causal chain most likely to achieve the goal—opening the door—given the agent's current model of the environment. The agent's model of the environment comes from two forms of learning: bottom-up associative attribute learning and top-down abstract schema learning.

We define a target goal of our planner as a particular state of the environment, denoted s^*. Given a target goal our agent models its current state as a tuple of (n, q), where n represents the number of solutions remaining, and q the set of solutions already executed. The agent seeks to execute a causal chain c in the hopes of transitioning n to $n-1$. The agent replans after every attempt until it finds all solutions the room; *i.e.*, when $n = 0$. Thus, our final planning objective at time t is to pick the causal chain with the maximal posterior subject to the constraints that the chain contains the target goal state s^* (*i.e.*, the door being *pushed*) and is not in the agent's set of solutions executed q:

$$c_t^* = \arg\max_{c \in \Omega_C} p(c|\rho, do(q); \gamma, \beta) \quad \text{s.t. } s^* \in c \wedge c \notin q, \quad (6)$$

where $p(c|\rho, do(q); \gamma, \beta)$ is defined in Equation 5. This state definition matches information provided to human participants and places the focus of our planner on achieving task-level goals.

Among the chains that satisfy the constraints, we rely on our chain posterior to capture which chains are causally plausible. The posterior combines the top-down structural knowledge with the bottom-up attribute knowledge. This combination is powerful for two reasons: (i) bottom-up knowledge biases beliefs towards structures that contain attributes that have been present in causal events in the past, and (ii) top-down knowledge allows the agent to bias beliefs towards structures that have been useful in the past.

Model Results

We train our agent in the same fashion as humans; specifically, we allow the agent to complete 80 trials in CC and CE escape rooms (same number as human participants). The agent is also limited to 3 actions in an attempt and 30 attempts within a trial. Any agent that did not complete all trials was removed from the study (same as human participant data—no agents were removed from the CC condition; 7 agents were removed from the CE condition).

Figure 4 compares human and model performance. The model shows a similar trend as humans but with slightly worse performance in each trial[2]. For the agent assigned to the CC condition, the number of attempts needed to find the *first* solution decreased significantly following the first trial ($t(79) = 8.09; p < .001$) and second trial ($t(79) = 4.04; p < .001$). The CE agent required less attempts to find the *first* solution following the first trial only ($t(72) = 6.23; p < .001$). Decreases in first and second solution attempts were not significant between the remaining trials.

These results demonstrate that our model is roughly capable of capturing learning rates of human participants but does not capture all significant changes in the number of attempts needed: *e.g.*, in both the CC and CE conditions, the number of attempts needed by participants to find the *second* solution consistently decreased following the first trial. However, our model overall effectively captures general trends in human performance: the number of attempts needed to find *all* solutions matches well to humans and decreases near-monotonically, albeit at a lesser rate.

[2]Example solution executions for human participants and the model can be viewed at `https://vimeo.com/334518941`

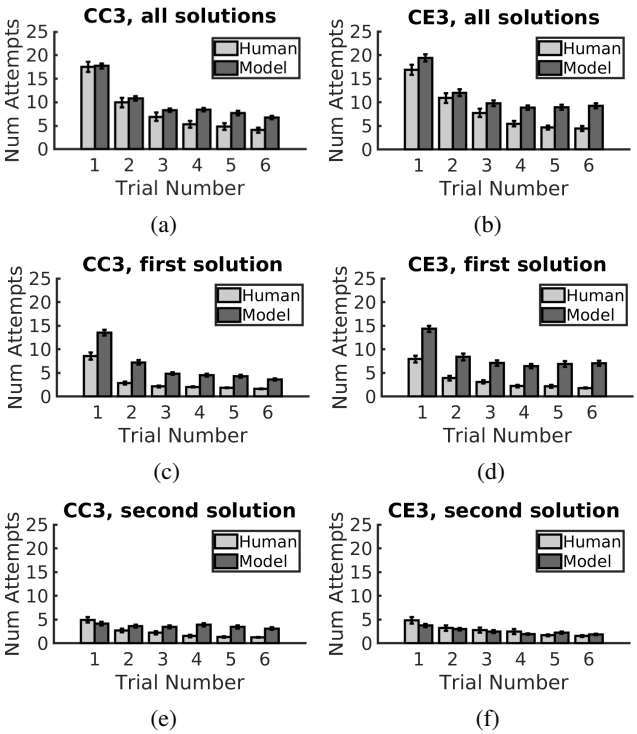


Figure 4: Comparison of human and model results for the common-cause CC3 condition and the common-effect CE3 condition. (a) and (b) compare the total number of attempts to find all solutions; (c) and (d) compare the number of attempts to find the *first* solution; (e) and (f) compare the number of attempts to find the *second* solution.

Conclusion

In this work, we showcase a hierarchical model based on associative learning and schema reasoning. Our model integrates two learning mechanisms: (i) a bottom-up theory that learns which attributes have causal associations in the environment, and (ii) a top-down theory that learns useful abstract structures in the environment. Our agent chooses an intervention based on the posterior of causal chains and updates its model using the observed outcome of the intervention. Model results show that our hybrid agent is able to capture general trends observed in human participants and captures some of the statistical significance observed in human performance. These results suggest that human causal learning may consist of a mechanism that combines bottom-up associative learning with top-down reasoning about causal structure.

The underlying computational framework presented here is broadly applicable outside of the OpenLock environment; it can be applied to any reinforcement learning environment where: (i) underlying dynamics are constrained by some causal structure; (ii) interactive elements have observable features which signal causal relevance; and (iii) physical locations of key elements change over time. In the future, we hope to expand our model to account for more extreme observational changes. For example, what if levers could suddenly be rotated instead of pushed/pulled? What if new colors were introduced which provided further cues about causal relevance? And what if the environment began operating in a probabilistic fashion where levers may fail to actuate properly? Future behavioral and computational work should examine how these processes integrate in more complex scenarios that provide a closer approximation to the real world.

Discussion

What other theories may be useful for learning causal relationships? The background theories presented here—namely that causal relationships induce state changes and abstract causal knowledge can be reused—provide reasonable background theories. However, other background theories may also be appealing. For instance, Pearl (2009) defines a stricter definition of causal relations based on whether or not a causal relation is *identifiable* in a directed acyclic graph.

How can hypothesis space enumeration be avoided? The spaces of Ω_{g^A} and Ω_{g^I} are enumerated in this work. Hypothesis space enumeration can quickly become intractable as problems increase in size. While this work used a fixed, fully enumerated hypothesis space, future work will include examining how sampling-based approaches to iterative generate causal hypotheses (*e.g.*, see Bramley et al. (2017)).

What are the other possibilities of bottom-up associative criteria? Our method treats low-level attributes as the criteria for our bottom-up associative learning. However, other possibilities are equally valid. For instance, a modeler could pair attributes with specific actions and learn distributions of causal effects over this pairing. This decision ultimately comes down to the resolution of the problem being considered and what is appropriate to correctly model the problem.

How is this work connected to reinforcement learning (RL)? The model-based planner is closely related to model-based RL. Our problem setting could be cast in terms of a 0-1 reward function—the agent receives a reward of 1 if the door is opened, and 0 otherwise. However, model-based RL typically assumes a world model is provided, but our agent iteratively updates its conception of world dynamics through associative learning and schema reasoning.

Acknowledgement

The authors thank Prof. Tao Gao, Prof. Ying Nian Wu, Feng Gao, and Chi Zhang for their helpful discussions. The work reported herein was supported by DARPA XAI N66001-17-2-4029, ONR MURI N00014-16-1-2007, NSF BSC-1655300, and an NSF Graduate Research Fellowship.

References

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neuraths ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, *105*, 9–38.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 708.

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

Edmonds, M., Kubricht, F., James, Summers, C., Zhu, Y., Rothrock, B., Zhu, S.-C., & Lu, H. (2018). Human causal transfer: Challenges for deep reinforcement learning. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334-384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661-716.

Heider, F. (1958). *The psychology of interpersonal relations*. Psychology Press.

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

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64–99.

Shanks, D. R., & Dickinson, A. (1988). Associative accounts of causality judgment. *Psychology of learning and motivation*, *21*, 229–261.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*(2), 222–236.