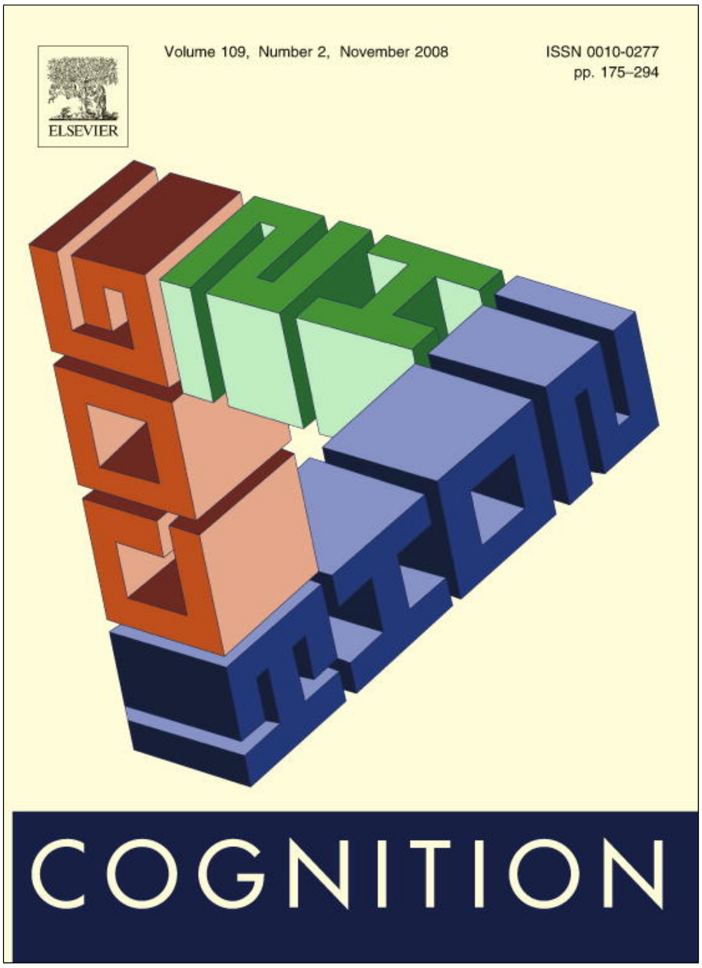
ELSEVIER
Volume 109, Number 2, November 2008
ISSN 0010-0277
pp. 175–294
COGNITION

Cognition 109 (2008) 175–192

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Inductive reasoning about causally transmitted properties ☆

Patrick Shafto [a,*], Charles Kemp [b], Elizabeth Baraff Bonawitz [c], John D. Coley [d], Joshua B. Tenenbaum [c]

[a] Department of Psychological and Brain Sciences, University of Louisville, 317 Life Sciences, Louisville, KY, USA
[b] Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA, USA
[d] Department of Psychology, Northeastern University, 360 Huntington Ave, Boston, MA, USA

ARTICLE INFO

Article history:
Received 31 August 2007
Revised 16 May 2008
Accepted 12 July 2008

Keywords:
Property induction
Inductive reasoning

ABSTRACT

Different intuitive theories constrain and guide inferences in different contexts. Formalizing simple intuitive theories as probabilistic processes operating over structured representations, we present a new computational model of category-based induction about causally transmitted properties. A first experiment demonstrates undergraduates' context-sensitive use of taxonomic and food web knowledge to guide reasoning about causal transmission and shows good qualitative agreement between model predictions and human inferences. A second experiment demonstrates strong quantitative and qualitative fits to inferences about a more complex artificial food web. A third experiment investigates human reasoning about complex novel food webs where species have known taxonomic relations. Results demonstrate a double-dissociation between the predictions of our causal model and a related taxonomic model [Kemp, C., & Tenenbaum, J. B. (2003). Learning domain structures. In *Proceedings of the 25th annual conference of the cognitive science society*]: the causal model predicts human inferences about diseases but not genes, while the taxonomic model predicts human inferences about genes but not diseases. We contrast our framework with previous models of category-based induction and previous formal instantiations of intuitive theories, and outline challenges in developing a complete model of context-sensitive reasoning.

1. Introduction

Any familiar thing can be thought about in a multitude of ways. A cat is a creature that climbs trees, eats mice, has whiskers, belongs to the category of felines, and was revered by the ancient Egyptians. Knowledge of all of these kinds plays an important role in inductive inference. If we learn that cats suffer from a recently discovered disease, we might think that mice also have the disease – perhaps the cats picked-up the disease from something they ate. Yet if we learn that cats carry a recently discovered gene, lions and leopards seem more likely to carry the gene than mice. Flexible inferences like these are a hallmark of human reasoning, which is notable for the selective application of different kinds of knowledge to different kinds of problems.

Psychologists have confirmed experimentally that inductive inferences vary depending on the property involved. When adults are told about genes or other internal anatomical properties, they tend to generalize to taxonomically related categories (Osherson, Smith, Wilkie, L'opez, & Shafir, 1990). When told about novel diseases, however, adults may generalize to categories related by a causal mechanism of disease transmission, such as a food web

☆ Preliminary versions of this work were presented at the International Conference on Thinking 2004 and the 27th annual meeting of the Cognitive Science Society. This work was supported by the Albert Fellowship to C.K., NSF Grant BCS-0236338 to J.D.C., and the Paul E. Newton Chair to J.B.T. The authors thank Vikash Mansinghka for helpful comments and discussion.

* Corresponding author.
E-mail address: p.shafto@louisville.edu (P. Shafto).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.07.006

(Shafto & Coley, 2003). Across development, children demonstrate increasingly distinct patterns of inference for properties such as drinking versus riding (Mandler & McDonough, 1996, 1998a, 1998b), anatomic versus transient properties (Gelman & Markman, 1986), and anatomy versus beliefs (Springer, 1996; Solomon, Johnson, Zaitchik, & Carey, 1996). Psychologists have also suggested, at least in principle, how complex inferences like these might work. Flexible inductive inferences are supported by *intuitive theories* (Murphy & Medin, 1985; Carey, 1985; Keil, 1989), or "causal relations that collectively generate or explain the phenomena in a domain" (Murphy, 1993). In any given domain, more than one theory may apply, and different patterns of inference will be observed depending on which theory is triggered.

Although a theory-based approach is attractive in principle, formalizing the approach is a difficult challenge. Recent work by Kemp and Tenenbaum (2003) has proposed a model for taxonomic theories. Here we describe and test a Bayesian theory-based model of induction about causally transmitted properties. This new model is a rational analysis of reasoning about causal transmission in the sense of Anderson (1990). The model consists of two parts: a generative theory that defines prior beliefs, and Bayesian inferential machinery that generalizes novel concepts by combining observed examples with prior beliefs.

We begin by discussing the problem of context-sensitive induction, and explain why theories and causal knowledge are important to understanding context-sensitive induction. We then present our model of causal property induction and the Bayesian framework for theory-based inference. A first experiment investigates undergraduates' reasoning about species with familiar taxonomic and food web relations, demonstrating qualitative fits between model predictions and human inferences. A second experiment shows that the model predicts human inferences about the distribution of diseases over a more complex artificial food web. In a third experiment, we contrast the fits of causal and taxonomic models to human generalizations of diseases and genes over known species, showing that the causal model predicts inferences about diseases but not genes, and the taxonomic model predicts inferences about genes but not diseases. Finally, we discuss our contributions to understanding the relationship between prior knowledge and reasoning, and outline challenges in developing a full model of context-sensitive induction.

2. Context-sensitive induction

In category-based induction tasks (Rips, 1975), participants are given one or more examples of categories that have a novel property. For example, participants may be told, "Lions have gene XR-35", where the property is gene XR-35, and lions are one example of things that have the property. Participants are then asked to judge the probability that other categories have the property; for example, "How likely is it that tigers have gene XR-35, like lions?" Properties are chosen such that participants have no specific knowledge about which categories have the properties, and predictions must be generated based on prior knowledge about the kind of property and the categories in question. Many elements of the context may influence reasoning in these tasks. Several important sources of context are the property being generalized, the sampling of example categories, instructions, and general demand characteristics. In this paper, we are concerned with the effects of different kinds of properties on inductive generalization.

Research has confirmed that the properties used strongly influence the inductive inferences of both children and adults. For example, Gelman and Markman (1986) found that 4-year-old children generalize internal anatomical and behavioral properties ("has cold blood") but not idiosyncratic properties ("gets cold at night") between members of the same category. Working with adults, Heit and Rubinstein (1994) showed that inferences differ when reasoning about behavioral versus anatomical properties. For example, participants were more willing to generalize between taxonomically matched species such as bears and whales when reasoning about properties such as "has a liver with two chambers that act as one". However, when reasoning about a behavioral property such as "usually travels in a back-and-forth, or zig-zag, trajectory", participants were more willing to generalize between behaviorally matched species such as tuna and whales. More recent research has shown that fishermen generalize diseases, but not properties, over food web relations, with inferences being stronger from prey to predators than from predators to prey (Shafto & Coley, 2003). These experimental examples underscore the importance of properties in inductive reasoning.

Previous models of property induction have had difficulty explaining sensitivity to context. Consider first the similarity-coverage model (Osherson et al., 1990), the best known model of category-based inductive reasoning. It predicts inferences about novel properties based on similarities between pairs of categories and a hierarchy of taxonomic relations among categories. The model makes accurate predictions about human generalizations in default contexts, when people are reasoning about generic biological properties that seem to refer to anatomy or physiology. However, accounting for inferences about anatomical and behavioral properties such as those in Heit and Rubinstein (1994) would require extending the model to allow context-sensitive notions of similarity. Even if this amendment is allowed, similarity-based approaches cannot naturally account for the causal asymmetries demonstrated in Shafto and Coley (2003) because ratings of similarity between predators and prey do not show strong asymmetries (see also Medin, Coley, Storms, & Hayes, 2003). To be fair, the similarity-coverage model was not designed with multiple contexts in mind; nevertheless, any comprehensive model of category-based induction will have to deal with the general phenomenon of context-sensitive reasoning, and reasoning about causally transmitted properties in particular.

Sloman (1993) proposed a more flexible feature-based approach to modeling property induction. Instead of appealing to stable notions of similarity or taxonomy, Sloman posits that each category is represented by a large, potentially context-sensitive, set of features. The strength of an inference from one or more example categories to some target category is based on a measure of overlap between the features of the target category and those of the example categories. Depending on how the features are selected or generated, Sloman's model could accommodate various effects of context-sensitive induction. Yet the model does not really explain these effects, because it does not attempt to account for how the features used to represent categories are derived, or how they vary with context.

One possible source for flexible feature generation or weighting could be abstract prior knowledge – knowledge about the kinds of features likely to be relevant in different contexts of reasoning. Sloman (1993, 1994) suggests that such knowledge could be the basis for feature weights in his approach, but he does not pursue a formal account of knowledge-based feature weighting. In a sense, this is our goal here. We adopt a Bayesian framework rather than a feature-based framework, but the approaches are analogous: Bayesian inference operates over a space of implicit hypotheses for how properties apply to categories rather than a space of implicit features for categories; the prior probabilities of hypotheses are analogous to feature weights. We will show how abstract domain theories can be used to generate appropriate priors for either a default context of taxonomic reasoning or an important alternative inductive context – reasoning about causally transmitted properties. We now turn to describing the role of intuitive theories – and theories of causal transmission in particular – in guiding inductive reasoning. We will return to the relationship between our theory-based Bayesian modeling approach and the similarity-coverage and feature-based models in the general discussion.

3. Intuitive theories of causal transmission

Two important roles of intuitive theories are to specify causal relations between features and causal relations between entities as well as implications of those relations. Several studies have shown that causal relations among features influence both categorization (e.g. Ahn, 1998; Rehder, 2003; Rehder & Hastie, 2001) and inductive reasoning (e.g. Rehder, 2006; Rehder & Burnett, 2005). Here we consider inductive reasoning about causal relations between entities; specifically, causal transmission of properties over these relations.

Reasoning about causal transmission between entities is fundamental across a broad range of domains, including the domains of folkbiology, folkpsychology, and folkphysics (Wellman & Gelman, 1992). In the domain of biology, reasoning about diseases depends on knowledge about potential mechanisms of causal transmission, such as feeding relations, physical/sexual contact, or spatial proximity. A growing body of evidence suggests that people from many cultures and age groups use knowledge about causal and ecological relations to reason about novel diseases. Indigenous Mayans (López, Atran, Coley, Medin, & Smith, 1997) and American tree experts (Proffitt, Coley, & Medin, 2000) use knowledge about ecological relations and potential mechanisms of transmission among local species to reason about diseases. Commercial fishermen use knowledge about food web relations to reason about novel diseases (Shafto & Coley, 2003). Similarly, rural and urban children use knowledge about causal and ecological relations to guide inferences about diseases (Coley, Vitkin, Seaton, & Yopchick, 2005). Developmental studies in the domain of folk psychology have shown that young children know that beliefs are transmitted socially (e.g. Solomon et al., 1996; Springer, 1996). Evidence from the domain of folk physics shows that children differentiate cases when forces could be transmitted between entities and cases when they cannot (e.g. Shultz, 1982). Reasoning about causal transmission is thus a highly general problem faced in a wide variety of domains.

Reasoning over causal relations requires concrete knowledge about the causal relations between entities, and more abstract knowledge about how properties are transmitted between entities (cf. Tenenbaum, Griffiths, & Kemp, 2006). We know that rabbits eat carrots, and that this is a potential means by which they may contract an illness. However, knowledge about the relationship between rabbits and carrots is tied to these examples: it does not tell us what other things are likely to eat carrots. Knowledge that eating is a route of disease transmission is abstract: it tells us that for any new pair of entities, the eater may contract an illness from feeding on infected or contaminated eatee. This distinction between concrete and abstract knowledge may explain why experts but not undergraduates showed context-sensitive reasoning about diseases in Shafto and Coley (2003); undergraduates may not have had the concrete knowledge required to make inferences based on causal transmission. Even if they knew that diseases may be transmitted from prey to predators, they would not have been able to use it. In this paper, we approach modeling knowledge about causal transmission using this multi-level approach, and we will investigate the validity of this assumption in our experiments.

4. Theory-based property induction

Bayesian models of category-based induction have been proposed before, but most suffer from an important limitation: the prior distribution plays a critical role in prediction, but previous models have not provided a formal account of the origins of this prior (Heit, 1998; Sanjana & Tenenbaum, 2003; Tenenbaum & Griffiths, 2001, but see Kemp & Tenenbaum, 2003). Theories offer a potential solution, provided they can be formally instantiated. Here we present a framework that combines Bayesian inference with theories that are formalized as probabilistic graphical models (Pearl, 2000; Spirtes, Glymour, & Schienes, 1993). The Bayesian inference engine implements rational statistical inference, and remains the same regardless of the inductive context. We model theories using probabilistic processes operating over graphical representations of the relationships between categories. Different probabilistic graphical models generate different prior distributions over hypotheses, and these priors lead to different patterns of inductive inference when combined with the Bayesian inference engine.

4.1. The Bayesian inference engine

In property induction, we observe one or more example categories D that have a novel property, and wish to compute the probability that another category (or set of categories) y also has the property. To compute this probability, we consider a hypothesis space H of all possible extensions of the property (see Fig. 1). Each hypothesis h specifies a particular subset of entities that have the property. The probability that y has the property, given D, can be computed by averaging the predictions of these hypotheses, weighted by their posterior probabilities

$$p(y|D) = \sum_{h \in H} p(y|h)p(h|D).$$

Note that $p(y|h)$ equals one if $y \in h$ and zero otherwise. We can expand $p(h|D)$ using Bayes' rule,

$$p(y|D) = \sum_{h} p(y|h)p(h|D) = \sum_{h} \frac{p(y|h)p(D|h)p(h)}{p(D)}. \tag{1}$$

The likelihood $p(D|h)$ is the probability of observing the data given that a particular hypothesis is true. We assume $p(D|h)$ is 1 if the data are consistent with the hypothesis and 0 otherwise; however, alternative sampling assumptions may be implemented within this framework (Tenenbaum, 1999; Tenenbaum & Xu, 2000). Because the numerator in Eq. (1) is zero for any h that is not consistent with y or not consistent with D, we can rewrite it as

$$p(y|D) = \sum_{h:y \in h, D \subset h} \frac{p(h)}{p(D)}.$$

The denominator can also be expanded by summing over all hypotheses

$$p(D) = \sum_{h} p(D|h)p(h) = \sum_{h:D \subset h} p(h).$$

Thus,

$$p(y|D) = \frac{\sum_{h:y \in h, D \subset h} p(h)}{\sum_{h:D \subset h} p(h)}. \tag{2}$$

The generalization probability $p(y|D)$ is therefore equal to the proportion of hypotheses consistent with D that also include y, where each hypothesis is weighted by its prior probability $p(h)$. If the conclusion y is included in most of the hypotheses with high prior probability that also contain the examples D, then the probability of generalization will be high.

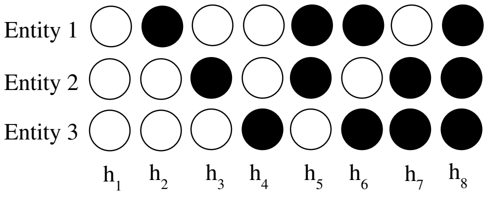

Fig. 1. All possible extensions of a novel property in a domain with three entities. Each candidate extension is a hypothesis, and black circles indicate that a given entity has the novel property.

4.2. Theory-based priors

The prior probabilities $p(h)$ in Eq. (2) represent *a priori* beliefs about the probabilities of different hypotheses. We suggest that reasoning about novel properties is guided by intuitive theories, which specify $p(h)$. We instantiate these intuitive theories using a combination of structured representations and probabilistic processes operating over the representations. In the next section, we formalize a simple theory of causal transmission, and explain how it results in a prior distribution over hypotheses. We will also introduce a theory of taxonomic inheritance (Kemp & Tenenbaum, 2003) and explain how it results in a prior distribution over hypotheses. Throughout, we will contrast predictions of the two models to highlight the importance of different kinds of knowledge in supporting inferences, and the ability of the Bayesian framework to support qualitatively different knowledge structures.

4.2.1. A generative model of causally transmitted properties

Consider the case of disease transmission by feeding over food web relations (Shafto & Coley, 2003). We generate a prior distribution using a theory with two components at different levels of abstraction (Tenenbaum et al., 2006). At the concrete level, the theory states the predator–prey relations that hold over the domain. The set of relations can be represented as a food web (for examples see Fig. 2a and c). Note that different food webs may apply to different sets of animals. At the more abstract level are general principles that describe how diseases are spread over any food web. The theory assumes that each species has a probability of contracting the disease from a cause external to the food web, and once infected an animal can pass the disease to predators. These possibilities depend on the functional form of the causal relationship and two parameters, a background rate and a transmission rate. The noisy-or causal relationship captures the idea that a single exposure to a disease is probabilistically sufficient to transmit the disease. The background rate is the probability that an animal gets the property from a cause

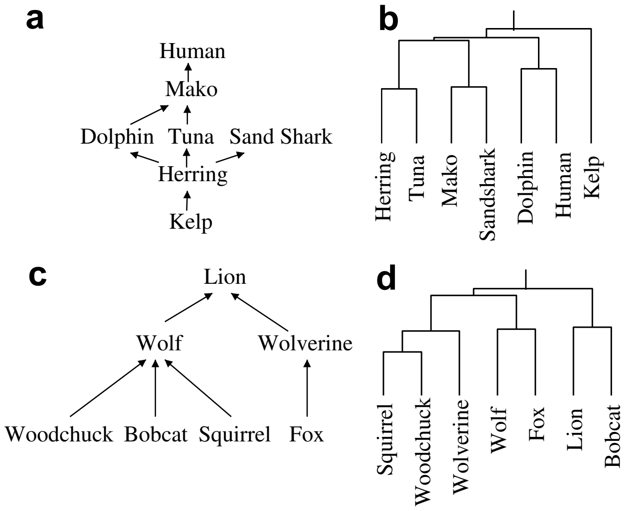

Fig. 2. Food web and taxonomic relations for two scenarios. (a) Food web for the island scenario. (b) Taxonomy for the island scenario. (c) Food web for the mammals scenario. (d) Taxonomy for the mammals scenario.

external to the web. The transmission rate is the probability that the property is transmitted from a species along an arrow to a causally related species. Assuming that the background rate affects each species independently and the transmission rate affects each arrow independently, we obtain a prior distribution over all extensions of the property. For simplicity we also assume for that the background rate and the transition rate are uniform across nodes and links; however, generalizations of this model could allow background rates and transmission rates to vary.

This basic model of causal transmission is similar to models used by scientists to understand transmission of diseases among people. In May and Lloyd's (2001) model of epidemics, nodes represent individuals, arrows between individuals indicate the kind of contact relevant to the disease (e.g. sexual contact in the case of HIV), and exogenous causes represent contacts with people not included in the network as well as other routes of causal transmission (e.g. sharing a needle with an infected person). The models used by scientists often include more detailed information than is included in our model, such as attributes of the nodes (e.g. gender), knowledge about the virulence of the disease, and frequencies of contacts between entities (e.g. Getoor, Rhee, Koller, & Small, 2004).

The prior probabilities, $p(h)$, represent the degree of belief in different possible extensions of the property, where each extension specifies for each animal whether it has the property or not. These priors can be computed by repeatedly simulating the arrival and transmission of disease in this model. A single simulation chooses a set of animals that acquire the disease from exogenous causes, and a set of causal links that are active (Fig. 3a). These choices imply that a certain set of animals will catch the disease, and that hypothesis is the output of the simulation (Fig. 3b). If we imagine repeating the simulation infinitely many times, the prior probability of any hypothesis is equal to the proportion of times it is chosen as output. In practice, approximate solutions can be obtained by repeating this simulation many times, and the quality of the approximation depends on the size of the network and the number of samples taken. For small problems the prior probabilities can be enumerated and calculated exactly, and we used this method for all of our calculations in this paper. Reflecting on the simulations should establish that the prior captures two basic intuitions. First, species that are linked in the web are more likely to share the property than species that are not directly linked. Second, property overlap is asymmetric: properties found in prey are more likely to be present in predators than vice versa.

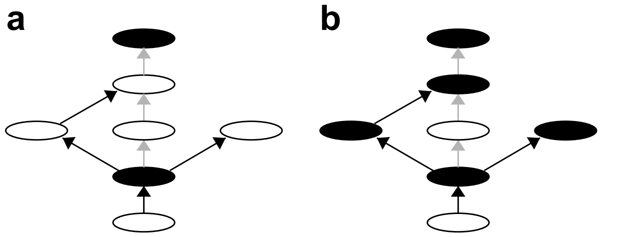

Fig. 3. One sample from the probabilistic model used to define prior beliefs. Black ovals indicate entities that have the disease. Black arrows indicate active routes of transmission. (a) Initial step showing active links and species that acquired the disease from exogenous causes. (b) Total set of species with disease via exogenous causes and causal transmission.

Two qualitative predictions emerge from the model, and will be tested in the experiments that follow. One phenomenon, *causal asymmetry* (Shafto & Coley, 2003; Medin et al., 2003), predicts that generalizations from prey to predator will be stronger than generalizations from predator to prey. Intuitively, this is because causal transmission is directed: when a prey has the property, by transmission the predator may acquire it. However, when a predator has a property, it may be because it acquired the property from the prey, or from some other cause. A second phenomenon, *causal distance*, predicts that the strength of generalization will decrease with increasing distance in the web. Intuitively, this is because the mechanism of causal transmission is fallible, so the probability that one species will receive a property transmitted by a second species decreases with the distance between them. Appendix A shows more formally how these qualitative predictions follow from our model. In addition, the model makes fine-grained quantitative predictions which will also be tested.

4.2.2. A generative model for taxonomic properties

The taxonomic model is based on two key ideas: species fall at the leaves of a known taxonomic tree (see Fig. 2b and d), and the novel properties are generated by a mutation process over the tree. Imagine a property that arises at the root of the tree, and spreads out towards the leaves. The property starts out with some value (on or off), and at each point in the tree there is a small probability that the property will mutate, or switch its value.

The taxonomic model has a single parameter – a mutation rate – which intuitively corresponds to the number of mutations a feature is expected to undergo while traveling down the tree. As in the previous generative model, this stochastic process induces a prior distribution over all possible extensions of a novel property. This prior captures a key intuition about taxonomic properties: the closer two species lie in the tree, the more likely they are to share a property. We call this prediction *taxonomic distance*, and we will test it as well as the qualitative predictions of the causal model in the experiment that follows.

This taxonomic model is related to the model for causally transmitted properties. Both models are based on structured representations (food webs or trees), and incorporate stochastic processes over those representations. Both are theories which generate knowledge-specific prior distributions over hypotheses, and explain phenomena in a domain. Both priors are combined with evidence by the same Bayesian inferential framework. These models thus provide insight into how domain-specific knowledge can be combined with domain-general inference mechanisms to explain context-specific inferences.

Both the causal and the taxonomic model were built by thinking about how properties are distributed in the world: diseases are distributed in the world by direct transmission between entities, while genes are distributed according to a mutation process over the evolutionary tree. Because both models are simple models of how the world works, both correspond to simple versions of models used by scientists – models like the causal model are used by epidemiologists, and models like the taxonomic model are used by evolutionary biologists. The resemblance between our models and scientific models reflects the assumption that people are approximately rational with respect to simple formulations of problems in the world.

5. Experiment 1: Reasoning about real-world causal transmission

People have a wide variety of knowledge about plants and animals which they might use to guide their inferences, making it important to establish whether there exist contexts that elicit reasoning based on causal and taxonomic knowledge. We asked participants to make judgments about the distribution of two different kinds of properties: novel physiological properties and novel diseases. Previously, Shafto and Coley (2003) demonstrated that experts' reasoning about diseases but not genes showed causal asymmetry. The current experiment investigated whether undergraduates' reasoning about diseases and physiological properties of known species would show context-sensitive use of causal and taxonomic knowledge.

To test this question, we assembled sets of three-species food chains, each composed of one prey, a predator of that prey, and a predator of the predator (e.g. carrots, rabbits, and wolves). We used simple chains to insure that participants knew the requisite food web relations, and were therefore able to apply their knowledge of causal transmission, if they deemed it appropriate. Taxonomic questions were drawn from items that did not have food web relations.

5.1. Method

5.1.1. Participants

Fourteen people participated in this experiment in exchange for a small monetary reward. Participants included both undergraduates and members of the broader M.I.T. community.

5.1.2. Materials and design

We identified six sets of three-link predator–prey chains that were familiar to undergraduates through pre-testing. These chains were: grass-sheep-wolves, plankton-salmon-grizzly bears, acorns-squirrels-hawks, sunflower seeds-sparrows-house cats, grain-mice-owls, and carrots-rabbits-fox. The full set of questions used in the experiment included all pairs within each causal chain (six per chain, 36 total), and six specifically taxonomically related pairs (asked in both orders, 12 total), and an additional 12 questions that were neither taxonomically close nor causally related. All of the taxonomic and unrelated pairs were created using species from the causal chain stimuli.

5.1.3. Procedure

The experiment was conducted on computer using MATLAB. Participants provided judgments in both the disease and physiological property conditions. Order was counterbalanced across participants. For each property, participants rated the likelihood of 60 statements of the form, "Carrots carry the bacteria XD. How likely is it that rabbits also carry the bacteria?" In the physiological property condition, questions asked whether a species would "have the XD enzyme for reproduction". Color pictures of the animals appeared with the questions. Ratings were made on a sliding scale that varied continuously between 1, "very unlikely", and 7 "very likely". The letter combination identifying the property was different for each question. Order of questions was randomized within each condition.

5.2. Results and discussion

We analyzed people's judgments for the presence of the two effects predicted by the transmission model, causal asymmetry and causal distance, and one effect predicted by the taxonomic model, taxonomic distance. We expect that effects predicted by the transmission model will be observed for diseases, but not physiological properties, and effects predicted by the taxonomic model will be observed for physiological properties but not for diseases. Because the predictions are across different items, we collapsed the data across participants. To test for causal asymmetry we compared inferences up the food chain, from prey to predators, to inferences down the food chain, from predators to prey (see Fig. 4, left panel). There was a marked difference between inferences up the chain ($Mean_{\text{up}} = 5.09$) and down the chain ($Mean_{\text{down}} = 4.42$) for diseases ($t(22) = 3.17$, $p < 0.005$), but no difference for physiological properties ($Mean_{\text{up}} = 3.16$, $Mean_{\text{down}} = 3.24$, $t(22) = 0.17$, $p > 0.5$).

To test for causal distance we compared inferences one link up the food chain (e.g. carrot-rabbit) to inferences two links up the food chain (e.g. carrot-wolf). Participants rated one link inferences significantly more likely than two-link inferences for diseases ($Mean_1 = 5.09$, $Mean_2 = 2.71$, $t(16) = 7.79$, $p < 0.001$) as well as for physiological properties ($Mean_1 = 3.16$, $Mean_2 = 1.58$, $t(16) = 4.03$, $p < 0.005$). Because distance in the food web is correlated with distance in a taxonomy, it is possible that the causal distance effect observed with physiological properties may be artifactual. We conducted a follow up analysis to test this possibility, which focused on the subset of causal distance items that were of the same taxonomic distance (Fig. 4, middle panel). The results showed no change in the effect for diseases ($Mean_1 = 5.01$, $Mean_2 = 2.71$, $t(10) = 7.72$, $p < 0.001$), and a markedly decreased effect for physiological properties ($Mean_1 = 2.11$, $Mean_2 = 1.58$, $t(10) = 2.30$, $p = 0.04$).

To test for taxonomic distance, we compared inferences of distances 1 (e.g. between two mammals), 2 (e.g. between a mammal and a bird), and 3 (e.g. between a mammal and a plant). For physiological properties, participants rated taxonomic distance 1 inferences ($M_1 = 5.43$) more likely than distance 2 ($Mean_2 = 3.68$, $t(24) = 8.48$, $p < 0.001$), and distance 2 more likely than distance 3 ($Mean_3 = 1.79$, $t(42) = 12.39$, $p < 0.001$). For diseases, there was no difference between items of taxonomic distance 1 and 2 ($Mean_1 = 4.37$, $Mean_2 = 4.11$, $t(24) = 0.60$, $p > 0.50$)

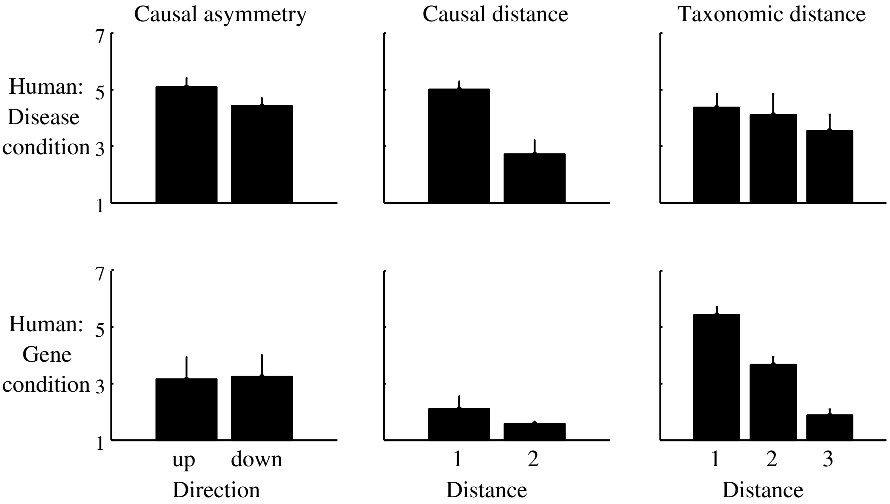

Fig. 4. Qualitative effects from Experiment 1. Error bars indicate two standard errors of the mean.

but there was a just-significant difference between items of distance 2 and 3 ($Mean_3 = 4.21$, $t(42) = 3.30$, $p = 0.04$). Controlling for causal distance across taxonomic distances by not including items with causal distance 2 resulted in no change in the effect for the physiological property ($Mean_3 = 1.88$, $t(30) = 8.48$, $p < 0.001$) but eliminated the effect for diseases ($Mean_3 = 3.55$, $t(30) = 1.14$, $p > 0.25$; v Fig. 4, right panel).

The results are consistent with the predictions of the models. We unexpectedly observed a prediction of the causal model, causal distance, for physiological properties. This effect was clearly smaller than for diseases; however, we will return to this issue in the third experiment and discuss it more fully there. These results suggest that undergraduates chose, from the wealth of information that they know about animals, taxonomic and food web knowledge to guide inferences in these two contexts. This result provides evidence that models of both kinds of knowledge are necessary to account for people's reasoning. In the second experiment we turn our attention to more systematically investigating predictions of the causal model using a more complex food web scenario.

6. Experiment 2: Testing the causal model

As we have described them, intuitive theories include two components: concrete knowledge about how a given set of entities can be organized into a structured representation, and more abstract knowledge about how properties are distributed over one of these representations. Both components are needed for inductive inferences about novel properties, but any given person may only have one of them. A visitor to a foreign country may know quite well that a predator can catch a disease by eating an infected prey animal, even if she does not know which animals eat each other in the local ecosystem. On the other hand, a child may often have observed an animal of species X eating an animal of species Y, but may not have made a connection between feeding relationships and the transmission of disease.

Although both components of intuitive theories are important, the average city-dwellers cannot be expected to have common knowledge of vast food webs. In this experiment, we taught our participants the novel food webs over blank (unnamed) animals (see Fig. 5). These novel situations take advantage of the abstract nature of knowledge about causal transmission, allowing use of a more richly structured food web to test the detailed quantitative predictions of the causal model. Quantitative fits of model predictions to human judgments provide a more stringent test of the model than qualitative effects – not only does the model have to predict individual qualitative effects, but also their relative importance. In the experiment, after learning the food web, participants were given a series of inductive problems. In each case, participants were told about a single animal that had a disease and then were asked how likely another animal was to have the disease. We will assess qualitative effects as well as the agreement between the model prediction and people's judgments for each question.

6.1. Method

6.1.1. Participants

Twenty people participated in this experiment in exchange for either course credit or a small monetary reward. Participants included both undergraduates from

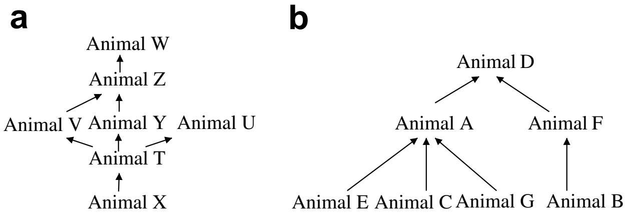

Fig. 5. Graphical depiction of blank food webs. Arrows point in the direction of transmission, from prey to predators (up the web). (a) Food web for the island scenario. (b) Food web for the mammals scenario.

Northeastern University and M.I.T. and members of the broader M.I.T. community.

6.1.2. Materials and design

Participants were tested on two different food webs (see Fig. 5, we refer to these as the (a) island scenario and (b) mammals scenario). The animals featured in these scenarios were blank (unnamed), but the scenarios were based on hypothetical scenarios with real animals. Each participant saw both scenarios, and the order of scenarios was counterbalanced across participants. Information about the food webs was conveyed to participants using two sets of seven cards. On each card was printed the name of one animal (e.g. "Animal A") and the immediate predators and prey of that animal. For example, the card corresponding to animal A read, "Animal A is eaten by animal D" and "Animal A eats animals C, E, and G". When the animal did not have any predators or prey, the card read "Animal D is not eaten by any other animals" or "Animal C does not eat any other animals". The statements appeared side by side on the cards. A test was also created to familiarize the participants with the food web information. The test was true/false and contained statements like, "Animal A is eaten by animal B" and "Animal D eats animal F, which eats animal B". At no point in the experiment did participants see a graphic representation of the food web; rather, they had to infer the structure from information on the cards.

6.1.3. Procedure

The experiment was conducted on computer using the Psyscope program (Cohen, MacWhittney, Flatt, & Provost, 1993). Participants provided judgments about each scenario and order of presentation was counterbalanced across participants. For each scenario, there were two phases: training and generalization. In the training phase, participants were given the seven cards corresponding to the animals in the set and asked to study the cards. Participants were then given a 20 question true/false test based on the information on the cards. Participants were allowed to keep the cards for reference during the test and through the generalization phase. Participants were required to score 85% on the test before they could advance to the generalization phase. If participants did not pass the test within 10 attempts, they were eliminated from the experiment (no participants were eliminated from this experiment). In each generalization phase, participants were presented with a series of 42 questions (all possible pairs) of the form, "Animal A has a disease. How likely is it that animal B has the same disease as animal A?" Participants rated the likelihood for each question on a 1–7 scale, where 1 indicated "very likely" and 7 indicated "very unlikely". Questions appeared in random order.

6.2. Results and discussion

In order to compare people's judgments with model predictions, we need to choose values for the free parameters: the background rate of a species having a disease, and the transmission rate, the probability of passing the disease to predators. These parameters are probabilities and therefore vary between zero and one. To fit the parameters, we performed a grid search over values of the parameters at intervals of 0.1 between 0 and 1. The model demonstrated robust performance across a range of parameter values with best performance at a base rate of 0.1 and a transmission rate of 0.5 (see Fig. 8). These parameter settings were used for all of the data presented in this paper. Fig. 6 shows that the model gives a good fit to human judgments, with correlations between the model predictions and human generalizations (by items) of 0.74 and 0.92 for the island and mammals scenarios, respectively.[1]

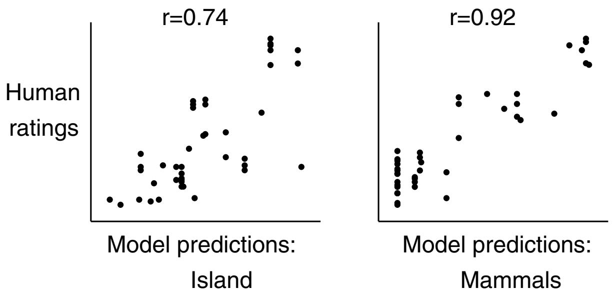

Fig. 6. Scatter plots of human data versus model predictions for the blank webs. Each point indicates an argument. Model predictions and human ratings are plotted along the *x* and *y* axes, respectively. Results for the island and mammals conditions are on the left and right.

Qualitative results were also obtained for two causal phenomena: causal asymmetry and causal distance (see Fig. 7). For the qualitative analyses, results were collapsed over the two scenarios and are based on the participants' mean ratings for items that are relevant to that comparison. To test for causal asymmetry, the average ratings for items involving generalizations up the chain (from prey to predator) were compared to generalizations down the chain (from predator to prey) using a two-tailed *t*-test (and Mann–Whitney's *U*). Generalizations up the food chain ($Mean_{\text{up}} = 2.00$, $Median_{\text{up}} = 1$) were stronger than generalizations down the food chain ($Mean_{\text{down}} = 4.50$, $t(23) = 11.94$, $p < 0.001$; $Median_{\text{down}} = 5$, $U(13,13) = 2, p < 0.001$), as predicted by the model. To test for the causal distance effect, generalizations up the chain were collapsed into four categories based on the distance from the premise to the conclusion. Causal distance predicts that generalization strength should decrease with increasing distance. Because there was only one argument with a distance of 4, statistical significance could not be evaluated for this case. One-link generalizations ($Mean_1 = 2.00$, $Median_1 = 1$) were stronger than two-link generalizations ($Mean_2 = 3.65$, $t(21) = 15.75$, $p < 0.001$; $Median_2 = 3$, $U(13,10) = 130$, $p < 0.001$) and two-link generalizations were stronger than three-link generalizations ($M = 4.65$, $t(10) = 5.46$, $p < 0.01$; $Median_3 = 5$, not enough samples for *U* test).

These results suggest that the model captures the major features of human reasoning about causal transmission. In particular, the predicted qualitative effects are observed and the correlations reflect a high degree of agreement between model predictions and human generalizations using

[1] Kendall τ_b correlations found associations of 0.55 and 0.59. Similar values are obtained for all correlations when performed on medians (rather than means).

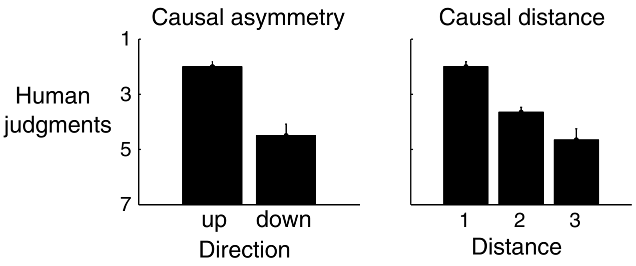

Fig. 7. Qualitative effects from Experiment 2. Results are collapsed across the two scenarios. Error bars indicate two standard errors of the mean.

a minimal number of free parameters (2) relative to the data (42 generalizations). Though the qualitative effects are predicted across almost all parameter settings, quantitative fits reflect much more detailed predictions about the relative ratings for all arguments that are sensitive to the particular parameter settings. We have an intuitive sense for what the values of these parameters ought to be based on our knowledge about the world. Transmission should be probabilistic (as opposed to deterministic), with a reasonable probability of contracting a disease if exposed, and we find that fits between model predictions and human data are best when the transmission rate is in the range of 0.4–0.6 (see Fig. 8). Similarly, the mechanism of transmission should explain the majority of outbreaks, so the background rate should be low, and we find that model fits peak at a background rate of 0.1 and decrease monotonically as the parameter value increases.

We find the strong quantitative and qualitative fits between the model predictions and human generalizations promising. However, these experiments were conducted under highly simplified conditions where only food web knowledge was available. One important difference between participants in these experiments and real-world reasoning is that food web knowledge must be chosen from among many potentially relevant kinds of knowledge, as in Experiment 1, and we turn to an experiment designed to address this issue.

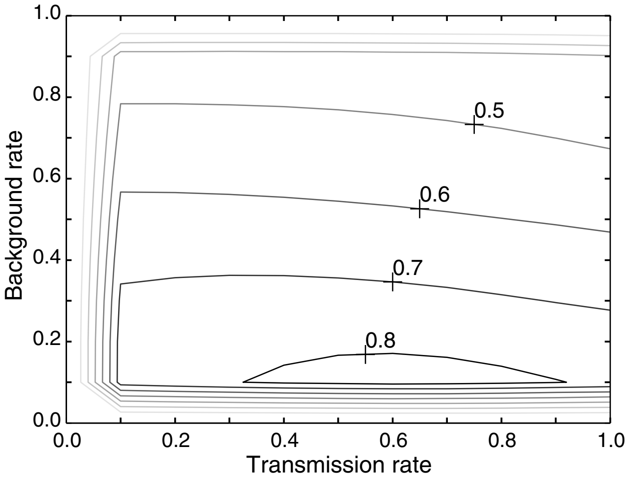

Fig. 8. Plots showing correlations between the model predictions and human data for different parameter settings. Plots reflect averaged results across the two (mammal and island) scenarios. Fits are robust across different values of the parameters, with best fits occurring for low values of the background rate and medium values for the transmission rate.

7. Experiment 3: Contrasting domain theories

In this experiment, we extended the scenarios used in Experiment 2 by replacing the blank labels with names of known biological species (see Fig. 2). Thus, people had knowledge of both taxonomic and food web relations among species to draw upon in reasoning, allowing us to revisit the context-sensitive reasoning found in the first experiment. To do so, we manipulated the kind of property people made inferences about: participants reasoned about either a novel disease or a novel gene. Based on previous work (Shafto & Coley, 2003) and the results of Experiment 1, we expect that inferences about diseases will be guided by knowledge about food web relations, while inferences about genes will be guided by taxonomic knowledge.

To model inferences drawing upon these different kinds of knowledge, we contrast the predictions of two models of domain theories: our Bayesian model of causal transmission, and a previously described Bayesian model of taxonomic reasoning (Kemp & Tenenbaum, 2003). These models both use the same Bayesian inferential machinery and differences in predictions therefore depend on the different priors used by the two models.

7.1. Method

7.1.1. Participants

Forty-two people participated in this experiment, 21 in the disease condition and 21 in the gene condition. Participants either received course credit or a small monetary reward in exchange for participation. Participants were drawn from the same population as in Experiment 2.

7.1.2. Materials

In this experiment, two domain structures were provided for each scenario: a food web and a set of taxonomic relations. Real animal names (e.g. "Wolf") were used instead of the blank animals (e.g. "Animal A"), and color pictures of the animals appeared on the training cards. The food web relations were structurally the same as in the previous experiment (see Fig. 2). The experiment used known animal species, and exploited people's intuitive knowledge about taxonomic relations. Training cards included both immediate food web relations (e.g. "Wolves eat bobcats, squirrels, and woodchucks" and "Wolves are eaten by mountain lions") and immediate taxonomic relations (e.g. "Wolves and fox are both canines"). For the island condition, the taxonomic labels used on the cards were mammals (human and dolphin), sharks (mako shark and sand shark), fish (herring and tuna), and plant (kelp). For the mammals condition, the taxonomic labels used on the cards were canines (fox and wolves), felines (bobcat and lion), and rodents (squirrel, woodchuck and wolverine).[2] Taxonomic questions were also added to the familiarization test (see

[2] Scientifically, wolverines are not rodents. However, we are interested in intuitive taxonomies, and in keeping with the distinctions present in the taxonomies derived from subjects' similarity judgments, we labeled the node corresponding to squirrels, woodchucks and wolverines. An informal survey suggested that "rodent" was the most appropriate label.

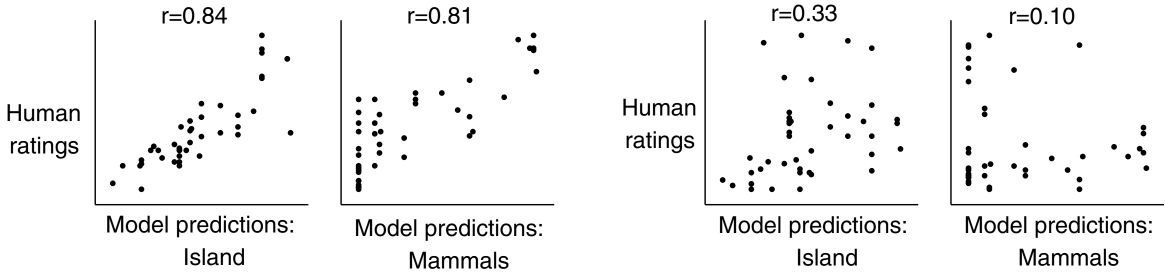

Fig. 9. Scatter plots of human inferences versus transmission model predictions for disease (left) and gene (right) conditions. Model predictions and human generalizations are plotted along the x and y axes, respectively. Each point indicates a single argument. The transmission model predicts reasoning about diseases but not genes.

Appendix B for the full set of questions for the mammals scenario).

7.1.3. Procedure

The experimental procedures were identical to the previous experiment with minor exceptions. First, the pretests included taxonomic questions. Second, participants were randomly assigned to a disease or gene condition. In the gene condition, the induction task contained questions of the form, "Bobcats have gene XR-34. How likely is it that lions have gene XR-34, like bobcats?" The disease condition questions were analogous to the questions in the previous experiment. Also, color images of the animals in the questions appeared on the screen with the question. Three participants failed the pre-test, and were dropped from the experiment. Also, one participant volunteered that she had used the rating scale backwards, and her responses were inverted. All other aspects of the experiments were the same.

7.2. Results and discussion

Model predictions were again compared to human judgments. For the causal model, predictions were derived using the same parameters as in the previous experiment. The taxonomic tree was constructed from similarity ratings obtained from a separate group of participants who went through the same training. Fits for the taxonomic model were robust across parameter values, and the mutation parameter was set to 0.1.

We computed correlations for both models on both property conditions. Results indicate a double-dissociation, with predictions of the causal model fitting generalizations of diseases ($r = 0.84$ and 0.81, see Fig. 9, left panel) but not genes ($r = 0.33$ and -0.10, see Fig. 9, right panel) and the taxonomic model fitting generalizations of genes ($r = 0.91$ and 0.90, see Fig. 10, right panel) but not diseases ($r = 0.27$ and 0.12, see Fig. 10, left panel).[3]

We also analyzed the data for the qualitative effects: causal asymmetry, causal distance, and taxonomic distance. Again, results were collapsed over the two scenarios and are based on the participants' mean ratings for items that are relevant to that comparison. We expect the causal effects to hold only in the disease condition, and the taxonomic effects to hold only in the gene condition.

To test causal asymmetry, we compared generalizations from prey to predators to generalizations from predators to prey (see Fig. 11). There was a significant difference based on direction for diseases, with inferences up the chain ($Mean_{\text{up}} = 2.86$, $Median_{\text{up}} = 2$) stronger than inferences down the chain ($Mean_{\text{down}} = 4.30$, $t(25) = 11.98$, $p < 0.0001$; $Median_{\text{down}} = 4.5$, $U(13,13) = 0$, $p < 0.0001$). Results from the gene condition indicate no asymmetry, ($Mean_{\text{up}} = 3.43$, $Mean_{\text{down}} = 3.68$, $t(24) = 0.56$, $p > 0.50$; $Median_{\text{up}} = 4.5$, $Median_{\text{down}} = 5$, $U(13,13) = 71.5$, $p > 0.20$).

To test causal distance one-link generalizations were compared to two-link generalizations, which were compared to three-link generalizations. Results from the disease condition indicate a significant effect: one-link generalizations ($Mean_1 = 2.86$, $Median_1 = 2$) were rated more likely than two-link generalizations, ($Mean_2 = 4.51$, $t(21) = 15.75$, $p < 0.0001$; $Median_2 = 4.5$, $U(13,10) = 114$, $p < 0.005$) and two-link generalizations more likely than three-link generalizations ($Mean_3 = 5.33$, $t(10) = 2.54$, $p < 0.05$; $Median_3 = 5.75$). Results from the gene condition indicate no causal distance effect, as predicted: one-link generalizations ($Mean_1 = 4.56$, $Median_1 = 4.5$) were not different from two-link generalizations ($Mean_2 = 4.92$, $t(21) = 0.86$, $p > 0.40$; $Median_2 = 5.75$, $U(13,10) = 88$, $p > 0.05$), and two-link generalizations were not different than from three-link generalizations ($Mean_3 = 5.83$, $t(10) = 0.85$, $p > 0.40$; $Median_3 = 6.5$).

The qualitative prediction of the taxonomic model, that inference strength should decrease with increasing taxonomic distance, was also tested. To test the effect of taxonomic distance, we collapsed arguments from both scenarios into three groups based on distance in the taxonomic hierarchy, and labeled the groups with numbers representing ordinal distance in the taxonomy (see Fig. 2). Pairs such as herring and tuna, dolphin and human, wolf and fox, and squirrel and wolverine were labeled with as distance one pairs. Herring and mako, and wolf and wolverine were labeled as distance two pairs. Kelp and herring, and lion and wolf were labeled as distance three pairs. Results from the gene condition show a significant effect of taxonomic distance. Generalizations between distance one pairs ($Mean_1 = 2.62$, $Median_1 = 2$) were stronger than generalizations between distance two pairs ($Mean_2 = 4.97$, $t(50) = 24.81$, $p < 0.0001$; $Median_2 = 5$,

[3] Kendall τ_b correlations found associations of 0.66, 0.55, 0.28, and 0.03 for the transmission model and 0.79, 0.45, 0.30, and 0.14 for the taxonomic tree model. Similar numbers were obtained when analyses were conducted on medians (rather than means).

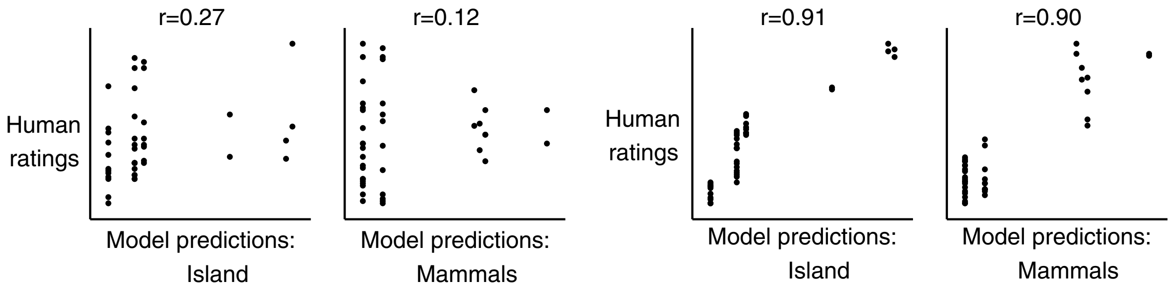

Fig. 10. Scatter plots of human inferences versus versus taxonomic model predictions for disease (left) and gene (right) conditions. Model predictions and human generalizations are plotted along the *x* and *y* axes, respectively. Each point indicates a single argument. The taxonomic model predicts human reasoning about genes but not diseases.

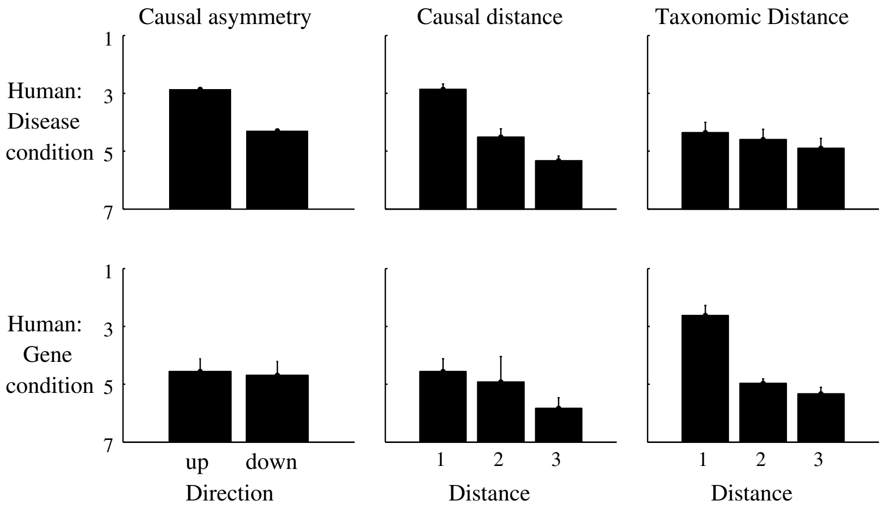

Fig. 11. Qualitative effects for the disease and gene conditions (on top and bottom rows, respectively). Error bars indicate two standard errors of the mean. Effects predicted by the causal model are observed in the disease condition, but not in the gene condition. Conversely, the effect predicted by the taxonomic model is observed in the gene condition but not in the disease condition.

$U(16,36) = 576, p < 0.001$) and generalizations between distance two pairs were stronger than generalizations between distance three pairs ($Mean_3 = 5.33$, $t(66) = 3.08$, $p < 0.01$; $Median_3 = 5.75$, $U(32,36) = 745$, $p < 0.05$). Results from the disease condition show no significant differences ($Mean_1 = 4.36$, $Mean_2 = 4.60$, $t(50) = 0.46$, $p > 0.40$; $Median_1 = 4.5$, $Median_2 = 5.5$, $U(16,36) = 319.5$, $p > 0.05$; $Mean_3 = 4.90$, $t(66) = 1.22$, $p > 0.20$; $Median_3 = 6$, $U(32,36) = 696$, $p > 0.05$).

Fig. 12 shows model fits to human data across a range of parameters. Fits for the causal transmission model peak for medium values of the transmission rate and low values of the background rate, in agreement with our intuition that transmission should be probabilistic and suggesting that model fits are best without much noise. Similarly, the taxonomic model performs best for low values of the mutation parameter, suggesting that the model structure is important to the fit of the model. The sensitivity of model fits to variations in the parameter settings emphasizes the fact that different parameter settings imply different priors and different predictions about the relative strengths of arguments.

These results provide further support for our causal model. Across four data sets, the model shows strong correlations with human reasoning about novel diseases, with correlations in all cases greater than 0.75. The causal model also predicts two qualitative phenomena observed in all three experiments. Additionally, the best-fitting values for the parameters correspond to our intuitions about disease transmission in the world.

The observed differences between people's reasoning about genes and diseases provide additional evidence that a single kind of knowledge cannot account for reasoning in the domain of biology (see also Heit & Rubinstein, 1994; Shafto & Coley, 2003; Smith, Shafir, & Osherson, 1993). Previous models have only captured reasoning about taxonomic properties relying on combinations of similarity and taxonomic knowledge, but human reasoning is more varied than can be accounted for with this information alone. Though in this paper we have pitted predictions of the causal transmission and taxonomic models against each other, this mutually exclusive approach to reasoning is also likely to be too simple to account for human inferences. Our qualitative and quantitative results have shown that, in our experiments, people tended to draw upon either causal or taxonomic knowledge. However, our main point is that both kinds of knowledge are necessary, and for this purpose it is sufficient that people draw primarily on food web or taxonomic knowledge to guide inferences about these properties. It is unlikely that human reasoning always

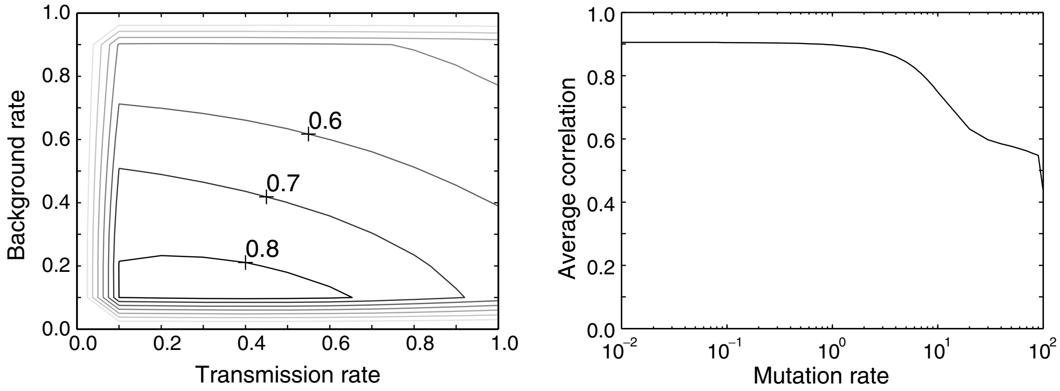

Fig. 12. Plots showing correlations between the model predictions and human data for different parameter settings of causal transmission (left) and taxonomic tree (right) models. Plots reflect averaged results across the two (mammal and island) scenarios. Fits for the transmission model peak for low values of the background rate and medium values of the transmission rate. Fits for the taxonomic model decrease with increasing mutation rate.

draws upon one single kind of knowledge, and we are working toward methods of integrating qualitatively different kinds of knowledge (Kemp, Shafto, Berke, & Tenenbaum, 2006). Additional knowledge structures are also probably necessary. Exactly how many different kinds of knowledge are sufficient to account for human reasoning will vary by domain and the experience of the person or group, but we are optimistic that a small set of models will emerge that are important across a wide range of domains.

More generally, these results highlight a strength of our Bayesian approach: the ability to support inferences based on different kinds of prior knowledge in a single inferential framework. Because our models capture knowledge about different kinds of relations among entities, they generate different prior distributions over hypotheses. Combining these domain-specific beliefs with domain-general inference results in qualitatively different generalizations. Our Bayesian approach provides a natural framework in which to explore the effects of different prior beliefs on inferences.

8. General discussion

We have presented a Bayesian model of reasoning about causal transmission. In Experiment 1, we showed that undergraduates spontaneously use knowledge about food web relations to support reasoning about causal transmission. In Experiment 2, we showed that the model provides good quantitative fits to human generalizations over food webs, and predicts two qualitative phenomena. In Experiment 3, we highlighted the importance of different domain theories in human reasoning. We showed a double-dissociation between the predictions of models of causal transmission and taxonomic knowledge and human generalizations of diseases and genes. The causal transmission model provides good quantitative and qualitative fits to human reasoning about diseases, but not genes. Similarly, the taxonomic model provides good quantitative and qualitative fits to human reasoning about genes, but not diseases. Together these results provide strong support for our model of causal transmission, illustrate the importance of different kinds of knowledge in inductive reasoning, and demonstrate the promise of our theory-based Bayesian framework as an approach to modeling human inductive reasoning.

More broadly, we have shown that for two different kinds of properties, a prior that approximately describes how properties of that kind actually vary over categories in the world provides a good fit to people's judgments, while priors that describe the distribution of other kinds of properties do not fit human judgments well. Our evidence suggests that people are able to understand the appropriate domain theories and use them in the appropriate contexts to guide inferences – at least for two kinds of properties that are likely to be of some ecological significance.

Beyond these specific findings, our work advances understanding of property induction and the role of intuitive theories in several ways, which we will elaborate in the following sections. First, we compare our results with previous empirical work on context-sensitive reasoning. Second, we emphasize the importance of theories in constraining inductive reasoning by contrasting our model with previous formal approaches to property induction. Third, we discuss methods of approximating full Bayesian inference and suggest algorithmic-level implementations of our framework, in the spirit of Marr's levels of analysis (Marr, 1982). Fourth, we contrast our theory-based Bayesian approach with previous formal models of theories. While far from complete, we argue that they capture important aspects of people's theories. We then conclude by outlining some of the remaining challenges in developing a complete model of context-sensitive property induction.

8.1. Empirical studies of property induction

Our results are consistent with previous studies demonstrating context-sensitive reasoning in the domain of biology, and offer insight into the basis of context-sensitive reasoning across domains. In the domain of biology, previous work found that experts but not novices used taxonomic knowledge and knowledge about food web relations to flexibly guide inferences (Shafto & Coley, 2003). Our results suggest that the lack of context-sensitive reasoning by novices was not due to a qualitative difference in experts' and novices' concepts of disease transmission. Rather, our results show that undergraduates use causal knowledge to guide inferences about familiar food web relations (cf. Medin et al., 2003) and novel scenarios, consistent with a multi-level theory of causal transmission where specific knowledge about the existence of a relationship exists at a concrete level, and more abstract knowledge about when and how causal transmission applies exists at an abstract level.

Previous research studying reasoning in the domain of biology has also demonstrated context-sensitive reasoning about properties such as "can bite through wire" (Smith et al., 1993), and anatomical versus behavioral features (Heit & Rubinstein, 1994). Natural extensions of our framework apply to each of these cases. Inferences about strength related properties such as "can bite through wire" may depend on a linear representations of the dimension of strength; because dobermans are stronger than poodles, if poodles can bite through wire, dobermans probably can too. Inferences about the anatomical features in Heit and Rubinstein (1994) can be modeled with a taxonomic tree, as in our Experiment 3. Similarly, behaviors often develop in response to environmental pressure, and inferences about behavioral properties can be modeled by distributions over ecological categories like land predators, ocean prey, aerial predators, etc. (see Shafto, Kemp, Mansinghka, Gordon, & Tenenbaum, 2006). Our framework provides a natural way to model these knowledge-based inferences and extending our framework to handle a wider variety of inductive contexts is an important area of future work.

A Bayesian approach can help explain the computational principles that support context-sensitive reasoning about different properties. This goes beyond the capacities of previous models, capturing a fundamental aspect of human reasoning – drawing upon different kinds of knowledge in different inferential contexts. Importantly, the model also provides strong correlations to people's intuitive judgments in these contexts. Our model, however, is an initial proposal and may need to be refined to fit empirical data more closely. Two challenges will be incorporating phenomena like the inclusion fallacy, where people rate an argument with a more general conclusion (e.g. bird) stronger than arguments with a more specific, atypical member of that category (e.g. ostrich), and violations of screening off, where for a causal chain A → B → C, people judge inferences from A and B to C to be stronger than inferences from B to C. One way general way to incorporate these phenomena into a Bayesian approach is to model more closely the assumptions that people bring to the task; for example, softening the model's interpretation of the meaning of "all" in the case of the inclusion fallacy, and modeling unobserved latent causes (c.f. Rehder & Burnett, 2005) or inferring the causal power from the number of examples for violations of screening off. Such refinements may be able to account for some of the phenomena that have been traditionally difficult for Bayesian approaches, and these are very important areas of future work.

8.2. Previous formal accounts of property induction

Previous formal models of property induction have focused on reasoning based on taxonomic knowledge. In this section, we discuss two representative models, the similarity-coverage (Osherson et al., 1990) and Sloman's feature-based model (Sloman, 1993). It has been argued that neither of these models cannot explain aspects of inductive reasoning, but the notions of a similarity metric or set of features are notoriously flexible and these are appealing languages for providing psychological descriptions of reasoning. It is important to note that neither of these approaches was designed with multiple contexts in mind, but it remains an important question whether the power of theory-based representations necessary, or could the different kinds of reasoning displayed here be explained more simply in one of these models – if only it is allowed to choose an appropriate similarity metric or set of features?

The similarity-coverage model obtains good fits to human reasoning based on taxonomic knowledge; indeed, the fits to our data were quantitatively comparable to those obtained by our taxonomic model (see Tenenbaum, Kemp, & Shafto, 2008). However, similarity-based models do not offer a simple way of accounting for reasoning based on causal knowledge. Even if similarity is allowed to vary based on context (e.g. Heit & Rubinstein, 1994), context-specific similarities cannot naturally account for the causal phenomena presented here because food web relations do not confer similarity in any traditional sense, and similarity does not demonstrate the strong asymmetries characteristic of the inferences in our studies. In contrast, our theory-based approach offers efficient representation for each context through the combination of a single structure with a small number of stochastic parameters (for causal transmission, a web plus two parameters). Additionally, our approach achieves greater generality by considering inferences in terms of hypotheses about possible extensions of the property; for example, allowing our models to make inferences based on knowledge beyond simple similarities (such as causal knowledge).

The feature-based model represents prior knowledge in terms of a category-by-feature matrix, where for each category each feature is either present or absent. Unlike the similarity-coverage model, the feature-based model may be able to account for reasoning about causal transmission given the right set of features. We can see why by observing an analogy between our Bayesian models and a feature-based approach. Hypotheses in our Bayesian models are analogous to features in a feature-based approach: each hypothesis or feature can be identified with a subset of categories in the domain, as the extension of that hypothesis or feature. The stochastic process we give for generating hypotheses according to some specified prior distribution (see Fig. 3) can then be thought as a process for generating features with some specified weights. Consider a category-by-feature matrix where features are sampled from the prior used by our model. For a fixed set of categories, this matrix would represent the same inductive potential as our priors because both would be generated by the same theory. In addition, our Bayesian inference rule (see Eq. (2)) can be interpreted as a measure of weighted feature overlap (see Tenenbaum & Griffiths, 2001). In light of this correspondence, our results and feature-based accounts of inductive generalizations are not incompatible.

The reason we are not satisfied with traditional feature-based models is that they do not represent the abstract knowledge needed to generate the relevant features and feature weights. Real-world inference cannot be accounted for by knowledge about object-feature co-occurrences alone. Consider, for example, the case of a scientist who lives on an island where the local food web is represented by Fig. 2a. Suppose that the scientist has recorded the distribution of one thousand different diseases, and (possibly unknown to him), the distribution closely matches the distribution predicted by our generative model. When asked the disease questions in Experiment 3, the scientist's responses match our model perfectly, but we cannot conclude that he has our theory; perhaps he is using the feature-based model over the data he has collected. Suppose, however, we ask the scientist a counterfactual question: we ask him about an ecosystem that is identical except that now people eat kelp, and makos do not eat tuna. A scientist with our theory of causally transmitted properties will have no trouble, but a scientist without the theory will be lost. While our experiments do not directly address this counterfactual question, we believe that people will respond flexibly when asked to reason about counterfactuals, or when otherwise given information that alters an underlying theory. Indeed, the ability to spontaneously alter a theory and use it to guide inferences is fundamental to the success of our experiments in which participants were asked to make inferences about novel food web relations. Since the species used in these experiments were novel, participants could not rely on a set of previously observed features, and the knowledge they used seems better characterized as a simple theory than as a collection of features.

Reasoning about causal transmission highlights inadequacies in both similarity and feature-based accounts of induction. Models based on similarity cannot naturally represent the asymmetric causal relations that are fundamental to reasoning about causal transmission. Models based on object-feature co-occurrences do not provide a means of abstracting from experience to theories about the world that are fundamental to reasoning about sparsely or previously unobserved instances.

Our theory-based Bayesian framework aims to fulfill the promise of proposals for a general purpose hypothesis-driven framework for inductive reasoning. Previous research has suggested that high-level explanations offered by hypotheses play a role in inference (McDonald, Samuels, & Rispoli, 1996). Other work has suggested that the Bayesian formalism may be used to generate inferences over hypothesis-based representations (Heit, 1998). Our contribution is to formalize how intuitive theories can be integrated with Bayesian hypothesis-based reasoning. We have demonstrated two instantiations of this theory-based framework for inferences based on taxonomies and asymmetric transmission over causal webs. Both our domain theories and Bayesian inferential mechanisms can be extended to other kinds of knowledge and richer inferential settings. We consider this to be a crucial aspect of any framework for investigating human reasoning: if we are to understand human reasoning, we must consider models expressive enough to make predictions across the many different situations that people face.

8.3. Cognitive plausibility and rational inference

Our goal in this paper has been to present a computational analysis of reasoning about causal transmission. Here, we consider the cognitive plausibility of our model and potential algorithmic implementations. There are two aspects of our models whose plausibility may be evaluated independently: representation and inference.

On first glance, the problem of representing prior knowledge in a Bayesian framework seems impossible for all but small cases: for any set of n objects, the number of potential hypotheses is 2^n. Under the most naïve approach, each hypothesis could be explicitly represented along with a number corresponding to our belief in the hypothesis. The amount of memory required to store this information increases exponentially with the number of objects, rendering this approach untenable. However, a central feature of our theory-based priors is representation based on composable graphical units, dramatically reducing the amount of memory required to represent prior knowledge. In the case of web representations, we only need the structure of the web and two parameters. This approach scales reasonably with increasing numbers of objects: regardless of the number of objects, only two parameters and a representation of the food web relations are required.[4] Our representation of prior knowledge is therefore very efficient, and potentially plausible at an algorithmic level of analysis.

There remains the problem of generating inferences from theory-based representations. Exact inference is intractable for Bayesian networks in general (Cooper, 1990). Accordingly, there are a variety of algorithms that have been developed to approximate full Bayesian inference with significantly reduced computational demands. Two such methods with analogs in the psychological literature are likelihood weighting and belief propagation. Likelihood weighting works by formalizing the intuition that people reason by simulating one or a few possible worlds (cf. Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). Belief propagation formalizes the intuition that inferences over a network may be generated by a kind of spreading activation between connected nodes (cf. Collins & Quillian, 1972). Developing a plausible algorithmic level account of reasoning remains an open problem (see Anderson, 1991; Sanborn, Griffiths, & Navarro, 2006), but we believe that the existence of good approximate algorithms is promising in itself, suggesting that there are plausible ways to arrive at approximately rational inferences with limited resources.

[4] The exact relationship between the increase in memory needed as the number of objects increases depends on how many connections a new object has to the existing structure. In the worst case, the maximum number of relations needed for a directed acyclic graph is $(n-1)!$, which still grows more slowly than the naïve case, 2^n. The cases we are interested are generally quite far from the worst case, so the advantage of our approach is typically more marked than is evident in this worst-case scenario.

8.4. Theories of theories

Previous work has described intuitive theories with a variety of forms, from the highly articulated knowledge structures proposed by Carey (1985) to the more skeletal frameworks proposed by Wellman and Gelman (1992). We do not claim to be modeling all or even most of the content of intuitive theories; however, we believe that our models go beyond previous work in capturing some important content of people's intuitive theories.

Previous investigations of theory-like knowledge have addressed the role of theories in organizing knowledge within a category (e.g. Ahn, 1998; Rehder, 2003; Waldmann, Holyoak, & Fratianne, 1995). Unlike our work, which focuses on causal relations between categories, these studies involve causal relations among features of categories, and treat representations of categories as miniature causal theories. Our approach focuses on intuitive domain theories – conceptual frameworks that organize a system of categories and describe how properties are distributed within that domain. Each of these views has precedent in the literature (Carey, 1985; Murphy & Medin, 1985; Wellman & Gelman, 1992), and each provides insight into different roles of intuitive theories.

Our approach, based on modeling intuitions about real-world domains, also contrasts with previous experiments, which have relied on artificial stimuli and features (e.g. Rehder, 2003; Rehder & Burnett, 2005; Rehder & Hastie, 2001). In these previous studies, stimuli were completely novel and features were explicitly chosen such that people would have no a priori beliefs about plausible causal relations. Participants were then taught different kinds of causal relations among the features and, when given different causal relations among features, categorized and reasoned differently. However, because the causal relations were arbitrary, these tasks are unlikely to tap into people's deeply held intuitive theories, with domain-specific knowledge and causal laws, that guide everyday reasoning and are the focus of our work.

Our experiments, in contrast, call on people to use their theories of a real-world domain, and thus provide a means to study the intuitive theories that guide everyday reasoning. We taught participants a novel set of food web relations, but this relational structure alone was not sufficient to predict disease incidence. To perform this prediction task, people must have been able to draw on more abstract, pre-existing knowledge about the domain of biology, and in particular, on their intuitive theories of disease transmission. For example, people must have used prior knowledge to infer that susceptibility to disease depends on a species' location in a directed network of predator–prey relations, that a single exposure is probabilistically sufficient for transmission, and that diseases can be picked-up from within the network or from outside causes. Because we allowed people to use their knowledge about the world to guide inferences, we can conclude that these aspects of our model correspond to people's intuitive theories about disease transmission – abstract knowledge that can be applied to novel problems.

Of course, this work has focused specifically on the role of theories in property induction, and our formalizations are likely simpler than people's theories even in the constrained settings we explored. For instance, evidence suggests that concepts such as resistance and susceptibility play an important role in some people's reasoning about diseases (e.g. Proffitt et al., 2000). The success of our models suggests that capturing only the core elements of theories can go a long way toward predicting people's inferences. Nevertheless, continued work will focus on modeling more of the detailed domain knowledge that supports human reasoning.

We have applied our model to a single case of property transmission, disease transmission among species; however, this basic computational model applies generally to problems of reasoning about causal transmission. As such, the model can be directly applied to problems such as reasoning about the transmission of beliefs, secrets, and fads. Importantly, though the basic model is generally applicable, we expect that aspects of both the model and people's theories should vary when applied to different problems. For example, the values of the transmission rate should be higher when reasoning about the distributions of beliefs than when reasoning about the distribution of secrets (assuming people actually keep secrets). Also, if applied to problems such as reasoning about the transmission of recessive traits from parents to children, we would expect different beliefs about the causal mechanism. In the case of transmission of properties such as having blue eyes, we expect that causes do not act independently: a child must inherit the trait from both parents, as described by a noisy and causal mechanism.

8.5. Towards a model of context-sensitive reasoning

Context-sensitive reasoning is considered a developmental milestone in many domains, and is a fundamental aspect of human intelligence. We have presented models of inference in two contexts, taxonomic properties and causally transmitted properties; however, much work remains to reach the overall goal of fully context-sensitive reasoning about different kinds of properties. First, additional models of different kinds of knowledge need to be developed, and existing models should be applied to different domains. Second, we have assumed that people already have different kinds of theoretical knowledge, for example, that people have knowledge about the structural relations and probabilistic processes implemented in our theories. It is important and necessary to show how multiple domain theories can be learned. Finally, we have assigned which theories apply to which contexts; for example, we specified that a causal theory applies to reasoning about diseases. People, however, infer which knowledge applies in which contexts, and future work will be directed at developing models that match this ability.

Like most important problems, context-sensitive induction should repay investigation before we understand it completely. By attempting to model multiple contexts, we can begin to understand the computational principles that support the scope and flexibility of human reasoning. Our work has emphasized principles like Bayesian inference and the importance of structured representations that apply to both of the contexts we considered, and suggests that modeling several contexts simultaneously may reveal insights that are not obvious when modeling each context in isolation. The ultimate goal of this line of work is to uncover principles that explain the different patterns of inference observed in different inductive contexts. Much work remains, but our formal framework suggests how we can make progress towards this goal.

Appendix A. Formal derivation of qualitative predictions for the causal transmission model

Our experiments test two main predictions of the causal transmission model: *causal asymmetry* and *causal distance*. Recall that the causal transmission model has two parameters. Let t denote the transmission rate and b denote the background rate. Throughout this section, we will assume that all probabilities are strictly greater than 0, which will be true whenever t and b are both greater than 0.

We will show that the causal asymmetry and causal distance effects are always predicted for simple network structures, regardless of the model's parameter values. For more complex network structures these effects may not always be predicted, but our analysis here suggests when they can be expected to hold. We have verified by simulation that these predictions hold for all of the networks used in our experiments, at the best-fitting parameter values. We have also verified that the predictions are robust with respect to changes in the parameters. As long as neither parameter is set to 0 or 1, both causal asymmetry and causal distance effects are always predicted to hold on average. That is, when we average across all relevant arguments, the model predicts that these effects will go in the appropriate direction, even if they do not necessarily go in the predicted direction for every individual argument.

Causal asymmetry. We first analyze the most basic case where the causal asymmetry effect should hold, a two-node causal chain $X \rightarrow Y$. The variable X indicates whether the prey species has a property, Y indicates whether the predator species has the same property, and the arrow denotes a route of causal transmission from prey to predator. Causal asymmetry obtains whenever $p(Y=1|X=1) \geqslant p(X=1|Y=1)$. The right side of this expression can be expanded following Bayes rule to give

$$p(Y=1|X=1) \geqslant \frac{p(Y=1|X=1)p(X=1)}{p(Y=1)}.$$

This inequality holds whenever $p(Y=1) \geqslant p(X=1)$, which will always be true because of the parameterization of the food web model. The predator and prey species are equally likely to acquire the property because of the background process, but the predator has some additional probability of acquiring the property from the prey. More formally, we can expand $p(Y=1)$ by conditioning on the values of X, to give the following condition that must hold for causal asymmetry to be predicted:

$$p(Y=1|X=1)p(X=1)+p(Y=1|X=0)p(X=0) \geqslant p(X=1).$$

Note that $p(X=1)=p(Y=1|X=0)=1-p(X=0)=b$, so the condition becomes

$$p(Y=1|X=1) \times b + b \times (1-b) \geqslant b.$$

This inequality is satisfied if $p(Y=1|X=1) \geqslant b$, which holds for any parameter values. Specifically, if the prey has some property than the probability $p(Y=1|X=1)$ that the predator has that same property is $b+(1-b)\times t$: with probability b, the predator gets the property from the background process, and with probability $(1-b)\times t$, the predator fails to get the property from the background process but does get it through transmission from the prey.

For more complex networks, causal asymmetry may not always hold. It will hold if the network includes more causes of Y, in addition to X. But if there are also causes of X – that is, if the prey itself is a predator – causal asymmetry may not hold. For instance, if X has very many causes while Y has very few causes, and if the transmission and background rates are low, the inequality $p(Y=1) \geqslant p(X=1)$ (and hence causal asymmetry) may no longer hold. Another way causal asymmetry can fail is if we allow the background rate to vary across nodes, and that rate happens to be much higher for the prey than for the predator. Neither of these cases however, applies to the networks and best-fitting parameter values for our experiments.

Causal distance. We first analyze the most basic case where the causal distance effect should hold, a three-node causal chain $X \rightarrow Y \rightarrow Z$. Causal distance obtains whenever $p(Y=1|X=1) \geqslant p(Z=1|X=1)$. The right side of this expression can be expanded by conditioning on Y, and noting that X and Z are conditionally independent given Y:

$$p(Y=1|X=1) \geqslant p(Z=1|Y=1)p(Y=1|X=1) + p(Z=1|Y=0)p(Y=0|X=1).$$

Now, let $q=p(Y=1|X=1)=p(Z=1|Y=1)$; these are equal if X is the only cause of Y, Y is the only cause of Z and the background and transmission rates are assumed to be the same over the whole network. Then the critical inequality becomes

$$q \geqslant q \times q + b \times (1-q).$$

This inequality is satisfied whenever $q \geqslant b$, but as we showed above for causal asymmetry, that is always true; q is just $b+(1-b)\times t$.

Causal distance does not necessarily hold for more complex networks. For example, it can fail if there are paths between X and Z that do not go through Y, or many additional causes of Z that do not influence Y. Exactly when causal distance fails will depend on these structural factors as well as on the values of the background and transmission rates. We have verified experimentally that causal distance holds for the networks and best-fitting parameter values in our experiments.

Appendix B. Pre-test questions: Experiment 3, mammals scenario

True statements:
Mountain lions eat wolves.
Mountain lions eat wolverines.
Wolverines eat fox.
Wolves eat squirrels.

Wolves eat woodchucks.
Wolves eat bobcats.
Mountain lions eat wolverines.
Mountain lions eat wolverines.
Mountain lions eat wolverines, which eat fox.
Mountain lions eat wolves, which eat bobcats.
Mountain lions eat wolves, which eat squirrels.
Mountain lions eat wolves, which eat woodchucks.
Mountain lions and bobcats are both felines.
Wolves and fox are both canines.
Wolverines and squirrels are both rodents.
Squirrels and woodchucks are both rodents.
Woodchucks and wolverines are both rodents.

False statements:
Bobcats eat wolves.
Mountain lions eat woodchucks.
Wolves eat mountain lions.
Wolverines eat mountain lions.
Fox eat wolverines.
Mountain lions eat squirrels.
Fox eat wolves.
Woodchucks eat wolverines.
Mountain lions eat wolverines, which eat squirrels.
Mountain lions eat wolves, which eat fox.
Mountain lions eat bobcats, which eat squirrels.
Mountain lions eat squirrels, which eat fox.
Mountain lions and wolves are both felines.
Wolves and wolverines are both canines.
Bobcats and squirrels are both rodents.
Fox and woodchucks are both rodents.
Woodchucks and mountain lions are both rodents.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts. *Cognition, 69*, 135–178.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409–429.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Cohen, J. D., MacWhittney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Coley, J. D., Vitkin, A. Z., Seaton, C.E., & Yopchick, J. E. (2005). Effects of experience on relational inferences in children: The case of folkbiology. In *Proceedings of the 27th annual conference of the cognitive science society*.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. Gregg (Ed.), *Cognition in learning and memory* (pp. 117–137). New York: Wiley. pp. 117–137.

Cooper, G. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence, 42*, 393–405.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Getoor, L., Rhee, J. T., Koller, D., & Small, P. (2004). Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine, 30*, 233–256.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford University Press.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 411–422.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62–88.

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

Kemp, C., & Tenenbaum, J. B. (2003). Learning domain structures. In *Proceedings of the 25th annual conference of the cognitive science society*.

Kemp, C., Shafto, P., Berke, A., & Tenenbaum, J. B. (2006). Combining causal and similarity-based reasoning. In *Advances in neural information processing systems*.

López, A., Atran, S., Coley, J. D., Medin, D., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251–295.

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition, 59*, 307–335.

Mandler, J. M., & McDonough, L. (1998a). Inductive generalizations in 9- and 11-month olds. *Developmental Science, 1*, 227–232.

Mandler, J. M., & McDonough, L. (1998b). Studies in inductive inference in infancy. *Cognitive Psychology, 37*, 60–96.

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

May, R. M., & Lloyd, A. L. (2001). Infection dynamics on scale-free networks. *Physical Review E*, 64.

McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis-assessment model of categorical argument strength. *Cognition, 59*, 199–217.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychological Bulletin and Review, 10*, 517–532.

Murphy, G. L. (1993). Theories and concept formation. In I. V. Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis*. Academic Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Osherson, D., Smith, E. E., Wilkie, O., L'opez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185–200.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 811–828.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Rehder, B. (2006). When similarity and causality compete in category-based property induction. *Memory & Cognition, 34*, 3–16.

Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264–314.

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

Rips, L. J. (1975). Inductive judgements about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*, 665–681.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th annual conference of the cognitive science society*.

Sanjana, N. E., & Tenenbaum, J. B. (2003). Bayesian models of inductive generalization. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural processing systems 15*. MIT Press.

Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 641–649.

Shafto, P., Kemp, C., Mansinghka, V., Gordon, M., & Tenenbaum, J. B. (2006). Learning cross-cutting systems of categories. In *Proceedings of the 28th annual conference of the cognitive science society*.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*, 1–51.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 213–280.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition, 52*, 1–21.

Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgements of probability. *Cognition, 49*, 67–96.

Solomon, G. E. A., Johnson, S. C., Zaitchik, D., & Carey, S. (1996). Like father, like son: Young children's understanding of how and why offspring resemble their parents. *Child Development, 67*, 151–171.

Springer, K. (1996). Young children's understanding of a biological basis for parent-offspring relations. *Child Development, 67*, 2841–2856.

Spirtes, P., Glymour, C., & Schienes, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Tenenbaum, J. B. (1999) Bayesian modeling of human concept learning. In: M. Kerns, S. A. Soller, T. K. Leen & K. R. Müller (Eds.), *Advances in neural processing systems*.

Tenenbaum, J. B., & Xu, F. (2000). Word learning as Bayesian inference. In: Gleitman, L. & Joshi, A. (Eds.). In *Proceedings of the 22nd annual conference of the cognitive science society*.

Tenenbaum, J. B., Kemp, C., & Shafto, P. (2008). Theory-based Bayesian models of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge University Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629–641.

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

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the aquisition of category structure. *Journal of Experimental Psychology: General, 124*, 181–206.

Wellman, H., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43*, 337–375.