HHS Public Access

Author manuscript

J Exp Psychol Gen. Author manuscript; available in PMC 2020 April 01.

Published in final edited form as:

J Exp Psychol Gen. 2019 April ; 148(4): 619–643. doi:10.1037/xge0000594.

From the structure of experience to concepts of structure: how the concept 'cause' is attributed to objects and events

Anna Leshinskaya and **Sharon L. Thompson-Schill**

Department of Psychology, University of Pennsylvania

Abstract

The pervasive presence of relational information in concepts, and its indirect presence in sensory input, raises the question of how it is extracted from experience. We operationalized experience as a stream of events in which reliable predictive relationships exist among random ones, and in which learners are naïve as to what they will learn (i.e., a statistical learning paradigm). First, we asked whether predictive event pairs would spontaneously be seen causing each other, given no instructions to evaluate causality. We found that predictive information indeed informed later causal judgments, but did not lead to a spontaneous sense of causality. Thus, event contingencies are relevant to causal inference, but such interpretations may not occur fully bottom-up. A second question was how such experience might be used to learn about novel objects. Because events occurred either around or involving a continually present object, we were able to distinguish objects from events. We found that objects can be attributed causal properties by virtue of a higher-order structure, in which the object's identity is linked not to the increased likelihood of its effect, but rather, to the predictive structure among events, given its presence. This is an important demonstration that objects' causal properties can be highly abstract: they need not refer to an occurrence of a sensory event per se, or its link to an object, but rather, to whether or not a predictive relationship holds among events in its presence. These learning mechanisms may be important for acquiring abstract knowledge from experience.

Keywords

statistical learning; causal inference; causality; object kinds; relational learning

GENERAL INTRODUCTION

A cardinal and puzzling quality of concepts is that their content often lacks any physical referent: there is nothing we can point to as serving the meaning of words like *believe*, *nutritious*, or *communicate;* yet these concepts denote properties that often belong to people and physical objects. One proposal for the source of such concepts is that they refer to a relational structure among certain sensory qualities, rather than those qualities themselves

Address correspondence to: Anna Leshinskaya, Department of Psychology, University of Pennsylvania, 425 S. University Ave, Stephen A. Levin Bldg., Philadelphia, PA, 19104, alesh@sas.upenn.edu.

AUTHOR NOTE

This work was presented at the Annual Meeting of the Cognitive Science Society (Philadelphia, PA, 2016) and the Annual Meeting of the Vision Sciences Society (St Pete Beach, FL, 2017).

(Carey, 2009; Chatterjee, 2008; Gopnik & Meltzoff, 1997; Jones & Love, 2007; Kemp, Tenenbaum, Niyogi, & Griffiths, 2010; Markman & Stilwell, 2001). For example, *communication* may denote a reliably predictive relation between speaking and receiving a reply; *belief* may denote a reliably causal relation between seeing, knowing, and acting. The elements alone, without the correct structure, would not be sufficient, suggesting that such concepts package relational information itself. What makes such meaning abstract is that it is only indirectly present in the information arriving to the senses: it must be inferred, and thus its representation depends on the operation of our minds (Dennett, 1987). The present research investigates such operations: how we extract relations from sensory experience in the temporal domain, how these relations can characterize concrete objects, and whether and how they can affect conceptual interpretation.

A wealth of research on lexical semantics and category learning confirms that relational structure is part of the meaning of concepts. This is particularly notable for the "cause" relation and its importance in verb meanings (Garvey & Caramazza, 1974; Pinker, 1989)—*pushing* would not be the same if the pusher applied force to himself rather than the pushee, or if the pushee fell over simultaneously with the pusher's actions rather than because of them. Thus, the structure of a relation matters for meaning, holding constant the tangible entities or qualities entering in that relation. Relatedly, adults incorporate explicitly presented causal information when learning new categories of entities (Ahn, 1998; Ahn, Kim, Lassaline, & Dennis, 2000; Genone & Lombrozo, 2012; Murphy & Medin, 1985; Rehder, 2003b, 2003a; Sloman, Love, & Ahn, 1998), and artifact categories critically rely on functional properties, which refer in part to what those objects cause (Bechtel, Jeschonek, & Pauen, 2013; Futó, Téglás, Csibra, & Gergely, 2010; Hernik & Csibra, 2009; Keil, Smith, Simons, & Levin, 1998; Kelemen & Carey, 2007; Kemler Nelson, Frankenfield, & Morris, 2000; Träuble & Pauen, 2011; Truxaw, Krasnow, Woods, & German, 2006).

The pervasive presence of relational information in concepts, and its indirect presence in sensory input, raises the question of how it can be extracted from experience and conceptualized. Although accounts of relational reasoning tacitly assume that relations are already-formed concepts, our view is that relations inferred from experienced events are inputs to conceptualization. In other words, relational information can be extracted from experience to inform the formation of novel concepts and categories. Prior work on visuospatial relations supports the idea that learners can form relation-based categories from visual evidence (Corral & Jones, 2014; Goldstone, Medin, & Gentner, 1991; Stuhlmueller, Tenenbaum, & Goodman, 2010; Tomlinson & Love, 2010). Here, we focus on temporal-predictive relations, and their relation to the concept *cause.* We ask whether the meaning of the concept *cause* can be said to refer to a pattern of predictive relations.

Prior work on statistical models of causal reasoning can be seen as partial support of this idea. When asked to determine if some event A causes some event B, adults and children appear to use specific statistical criteria for assessing causality: that the covariation between A and B is strong, independent, and unique compared to other, alternative potential causes (Buehner, Cheng, & Clifford, 2003; Cheng, 1997; Gopnik et al., 2004; Griffiths & Tenenbaum, 2009; Penn & Povinelli, 2007; Schulz & Gopnik, 2004; Shanks, 1985; Sloman & Lagnado, 2015; Sobel & Kirkham, 2007; Spellman, 1996). Towards our question of what

relational concepts refer to, this suggests that the concept *cause* denotes (in part) a certain pattern of statistical information, which is in principle possible to extract from the experience of events.

This account is incomplete, however, for several reasons. The first is that these judgments have been approached from only one direction: participants are told at the start to evaluate causality, and are then given evidence to do so. For example, they are told that they are doctors seeing patients, and must determine what causes a disease, or whether it seems likely that a certain virus causes it. We term this a "top down" causal task, because participants are given the concept to apply to evidence. But to represent the meaning of a concept, it is important to be able to do the converse. For example, the proper possession of the concept 'dog' would enable someone to pick out instances given the label, and apply the label when presented with instances. Thus, we ask whether, given the evidence, participants will recognize causality on their own accord. We call this situation a "bottom up task," and reason that if participants use the concept cause to designate a special pattern of the statistical evidence, then when the right pattern of statistical relations is presented, participant should see or recognize causation. This ability is important because real-world learners do not know in advance what they are about to learn (Aslin & Newport, 2012) – and thus, building a causal model of the world would strongly benefit from being able to recognize causation when its signature is observed among events.

A second limitation is that causal reasoning experiments are different from everyday experiences of events. Participants are shown contingency information though clearly demarcated instances in which (for example) a disease occurs with, or without, the presence of a virus. Although this might simulate the situation of doctors seeing clinical cases, or gardeners varying individual flower beds, it is not representative of daily sensory experience, in which events unfold sequentially in time without demarcated trials, there are many candidate events to track, and learners must determine the extent to which X uniquely predicts Y rather than Y, Z, or W. Thus, classic causal reasoning experiments present evidence in a way that is quite different from how we typically observe events, though increasingly more naturalistic paradigms are being employed (Bramley, Mayrhofer, Gerstenberg, & Lagnado, 2017; Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015; Buehner & May, 2009; Greville & Buehner, 2010; Rottman & Keil, 2012). We thus adopted a paradigm which better captured everyday experience, by presenting a relatively large number of events in a continuous sequence. We then asked whether the presence of the right statistical information in such displays would be sufficient to lead to recognition that causality was present. Answering this will help us understand whether the content denoted by the concept *cause* is something we may naturally recognize in the course of everyday observation.

Finally, most formal accounts of the statistical evidence supporting causal judgments do not readily explain how we attribute causality to *objects* as opposed to events. Adults and children readily make attributions of causality to objects, and this critically influences how they categorize artifacts (Gopnik & Sobel, 2000; Kemler Nelson et al., 2000; Träuble & Pauen, 2007, 2011). Kettles *cause* water to boil; telephones *enable* communication. Few such causal powers are mechanically transparent, suggesting that predictive or statistical

information is essential (Gopnik et al., 2004; Gopnik & Schulz, 2004). However, it seems unlikely that objects are treated in the same way as events when learning their causal powers. Under most statistical models of causality, events are expected to attain causal powers by a unique, independent correlation between their presence and an effect. But for many everyday objects, this seems inadequate: experience is replete with objects whose causal effects occur neither frequently nor uniquely in their presence. For instance, coffee makers produce coffee but spend more time standing inert in the kitchen; light switches are present on the wall regardless of whether the light is on or off; poisons are largely stored in a cupboard. On the other hand, mugs appear with coffee, the contents of the room with the light, and so on. An increased presence of coffee, light, or toxic effects cannot explain how these objects attain their causal powers.

We thus test a different possibility, that objects are granted causal powers by virtue of a higher-order structure, in which the object's identity is linked not to the increased likelihood of its effect, but rather, to a lower-order predictive structure among events, given its presence. An event (hitting a switch on the coffee maker) is what elicits the effect (the production of coffee); the coffee maker obtains its causal powers by serving as a broader context for this relationship. In this view, objects condition event relations: *if* something is a nutrient, ingesting it will help you; *if* it is a poison, ingesting it will hurt you. This account would make objects' causal properties a level more abstract than the event relationships which enter those relations directly.

This account closely aligns with the philosophical view of objects as having *dispositions* (Goodman, 1955; Mumford, 1998; Ryle, 1959). A disposition is a conditional relation governing how a thing will behave in various circumstances, rather than a physical feature or the occurrence of any state or event. For example, *fragility* is the propensity to break if dropped, rather than a frequent event of dropping or breaking (Mumford, 1998; Ryle, 1959). This makes it both related systematically to observable experience, but also abstracted from it by virtue of this higher-order relation. When entities are described as agents rather than patients (i.e., targets of an action), they are seen as having an increased disposition to cause events, affecting judgments about responsibility (Mayrhofer & Waldmann, 2015). Furthermore, literature on context-dependent learning further supports the idea that higher-order relations can be readily acquired from evidence, and related to spatial or temporal contexts—for example, how different chambers of a maze may govern whether or not a tone predicts a shock (Gershman, 2016; Urcelay & Miller, 2010, 2014). Here we empirically test whether such higher-order conditionals can inform how participants attribute causality to objects, as a way to account for the distinct way in which we judge objects to have causal powers. We term this prediction the dispositional/conditional view.

In summary, we ask two major questions. First, is presenting patterns of statistical dependence among events sufficient to *recognize* causality when not expressly looking for it, when those patterns are typically sufficient to judge causality in "top down" tasks? Second, can attributions of causality to *objects* be seen as higher-order conditionals among events? We ask these questions in the context of a task designed to better capture spontaneous observational learning from continuously presented events.

Towards this end, we created a learning environment that was adopted from the statistical learning (SL) tradition (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996; Turk-Browne, Jungé, & Scholl, 2005). In these paradigms, which we refer to as SL scenarios, a variety of events occur in succession; events can be speech sounds, or the appearances and disappearances of static visual images. Unbeknownst to the learner, these streams contain regularities in which certain events predict the subsequent appearance of others with a high probability, while other event pairs are not predictive: they are equiprobably followed by any of a number of events. Thus, SL scenarios allow the researcher to precisely control contingency statistics, and build in just the sort of statistical dependencies well known to influence causal judgments in other paradigms (Buehner et al., 2003; Cheng, 1997; Griffiths & Tenenbaum, 2005). However, they allow us to capture several features of naturalistic experience: there is no task demand to infer causality or any instruction; events appear in sequence without trial structure; and participants must extract systematic structure from randomness in the context of a number of different events. Thus, although originally conceived to study speech segmentation, the SL paradigm captures something important about the nature of everyday experience more generally.

However, this type of paradigm has been very rarely used in combination with causal judgments. In a notable exception, the SL paradigm has been used to study the segmentation of motion streams into meaningful actions, and such meaningful actions were judged as more likely to be causes of a separate, prespecified effect event (Buchsbaum et al., 2015). However, the unique question of interest here is whether predictive relations among events in such a scenario will themselves be seen as causal, when no instruction to look for causality is given ahead of time.

In Experiment 1, we presented learners with a continuous stream of events, some of which formed part of strongly predictive pairs. Our first question was whether this exposure to regularities would be sufficient to give learners a sense that the strongly predictive events cause each other, over and above the weak ones, despite no expectation to judge causality ahead of time. We measured spontaneous causal attribution by asking participant to describe what they had learned from the streams, and more solicited causal judgments by asking them to evaluate the extent to which various event pairs seemed to be causally related.

We also addressed our second question, whether causal attributions to objects can be based on higher-order statistics, as predicted by the dispositional/conditional view. To do so, we included in each event stream a continually present object, whose movements or color change could reliably predict the occurrence of another event in the environment. We then asked whether participants would be more likely to agree that the object caused this event, relative to another event that appeared even more frequently in its presence. In Experiment 2 we asked whether such statistical attributes support category learning. Finally, in Experiment 3, we used a convergent and more direct test of the dispositional/conditional view by varying the conditional relations among events in the context of two objects, and asking whether this difference led to distinct causal attributions towards them. Data and experiment materials are made freely available at https://osf.io/2d85n/.

EXPERIMENT 1

Introduction

Statistical models of causality were developed to account for participants' causal judgments on the basis of co-variation among events, such as between the administration of a medicine and a resulting headache across a number of trials (Buehner et al., 2003; Cheng, 1997). At the core of these models is the ΔP formula (Allan, 1980; Rescorla & Wagner, 1972; Shanks, 1985):

$$\Delta P = P(\mathrm{B}|\mathrm{A}) - P(\mathrm{B}|\sim\mathrm{A}) \quad \text{Eq 1.}$$

which captures the idea that a causal relation between A and B should be both strong and unique. Suppose that two events A and B coincide, such that after most occurrences of A, B occurs. However, B also occurs *without* A at a very high rate. One would not represent a strong link between A and B in this case. While formal models of causal inference differ in their details, nearly all of them share the core intuition behind ΔP (Hattori & Oaksford, 2007). This includes the noisy-OR model (Griffiths & Tenenbaum, 2005), power PC model (Buehner et al., 2003; Cheng, 1997), and Causal Support models (Griffiths & Tenenbaum, 2005; Kemp et al., 2010). Even though other models capture additional, nuanced aspects of causal judgments, strong differences in ΔP are expected to affect causal judgments under all of them, all else being equal. That is, participants will judge that medicine A causes symptom B to the extent that the occurrence of B increases when A is present, relative to its occurrence without A, assuming there is sufficient evidence to evaluate these quantities and priors to believe they are possible (see Griffiths & Tenenbaum, 2005 for direct model comparisons)[1].

As discussed in the General Introduction, however, these tasks always provide participants with the goal to evaluate causality at the outset. Participants thus keep in mind the target concept, *cause*, while considering evidence. We were interested in testing whether the same sort of statistical dependency—a high ΔP—would give rise to a sense of causality spontaneously. We thus presented participants with the evidence to see whether the inference work in this direction—from statistical information to the concept *cause*?

We presented statistical information in the context of a visual SL paradigm, in which several different visual events appeared in a sequential stream (Figure 1). Unlike typical SL paradigms, our events were animated changes-of-state that occurred in a persistent visual world with a continually present object in the background. We showed each participant two streams, each cued by a distinct background object, but containing largely the same events. In each stream, one pair of events was chosen to have a strong statistical dependency in one direction, such that B followed A with a high probability, but rarely followed other events. On the other hand, A did not often follow B, and was quite likely instead to follow all other

[1]A recent model called "Informative Transitions" (Derringer & Rottman, 2018) also predicts that participants care about contingency as described by ΔP, but that they specifically compute it over variables that *change* from the previous trial. This model is also in line with our predictions, as every trial in this paradigm involves a change from the last trial, and should thus expect very strong causal ratings among predictive events.

the events. Thus, while A to B transitions had strong dependency, B to A transitions had weak dependency. According to the prior models reviewed above, participants should judge that A causes B. In the second stream, a similar pair of events was again dependent, but in the opposite direction (A followed B). Here, participants should judge that B causes A. This reversal allowed us to control, within participant, any effect of stimulus identity or background expectation about mechanism, and vary only the statistical dependency among events according to formulas in prior models. The key question was to what extent participants would spontaneously see these statistical relations as causal, or whether they required prompting to do so.

We used two dependent measures of causal judgment, one more self-generated or spontaneous, and the other more solicited. The more spontaneous measure was a free response, in which participants described what they had learned in each video. Truly spontaneous causal recognition should lead participants to describe the relation between A and B in causal terms. Such descriptions would suggest that the statistical patterns themselves led participants to see causal relations, with no prompting or suggestion that the concept should apply. The second was a solicited causal judgment, in which participants rated how much they agreed with statements saying that A caused B, vs. that B caused A. This is less spontaneous because it is probed, but nonetheless reflects conclusions made *after* the evidence was processed by the participant. If participants accept these causal statements for the strongly predictive event pairs (that A caused B when A predicts B), relatively more than for the weak ones, it would suggest that they were able to make use of predictive statistics learned outside of an explicitly causal context to inform later causal judgments. This would indicate a moderate level of unsolicited inference from statistical patterns to the application of the concept *cause.*

Our paradigm also allowed us to ask about causal attributions to the *objects* in these streams. A different object was present in each of the two streams, and we used the object's movement or color change as the cause in one of these streams. For example, in stream 1, object A tilting might be followed reliably by the light flashing, while the confetti appear unpredictably (as in Figure 1). However, we ensured that confetti were actually more frequent in the presence of the object than the light flash, and much less frequent in the presence of object B. According to the dispositional/conditional view outlined in the General Introduction, it is the *dependent* event, light flash, which should be seen as being caused by the object, even though it does not covary with its presence like the confetti does. Rather, the light flash participates in a disposition: a conditional rule governing how an object will behave in specific conditions, rather than any specific state in which the object appears. We thus probed participants' judgments about what they believed the object caused, and predicted they should judge the light flash to be chosen more than the confetti (in this example).

Methods

Participants—70 participants were recruited and tested using Amazon's Mechanical Turk[2]. The sample size was chosen to be in line with that of related work on causal reasoning (Bramley, Gerstenberg, & Lagnado, 2014; Buchsbaum et al., 2015). Procedures

were approved by the Institutional Review Board of the University of Pennsylvania, and all participants provided electronic consent. Participation was compensated with $3.00 plus a performance-based bonus of up to $4.00. One participant was excluded for previously participating in another experiment involving these stimuli; of the remaining 69, 12 (17%) were excluded for failing to pass an attention check measure (described below). Of the 57 remaining participants included in the sample reported below, 30 were female and 27 were male, and their ages ranged from 21 to 62, with a mean of 36.

Stimuli—The stimuli are shown in Figure 1. They consisted of two novel, geometrical objects which each appeared in several types of animated events, presented as GIFs. Three events were *object based*: tilting (the entire object rotated −10 degrees), part moving (a detachable part on the object either slid up and down for object A or rotated for object B), and color changing (the object gradually changed from white to peach and back). Three others were *ambient*: light flashing (the background changed from gray to yellow and back), bubbles appearing (blue bubbles floated across the scene), and confetti (multi-colored stars streamed across the scene). GIFs were generated by hand-drawing individual frames using Adobe Illustrator and concatenating them into GIF files using Matlab (Mathworks). The GIFs were composed of 12 frames each, with each frame playing for 100ms, for a total event duration of 1200ms. A *static* event showed the object still on the screen for 2400ms. The object shapes (A or B) designated as object 1 and object 2 were randomized, and were given randomly assigned pseudoword names (*sibbie* or *thale*).

The GIFs were shown in continuous sequences, in which event transitions were governed by a pairwise transition matrix that specified the probability of any particular event following another. This was defined in a fixed way across all event assignments, objects and participants (Figure 1B). The transition matrix specified the following regularities: two events were designated as the *cause* and *effect*. If the cause event occurred on trial n, the effect occurred on trial n+1 with a 92% probability. Following other events, however, the effect occurred with very low probability (~.01). To compute ΔP, we used Eq 1, such that causal strength from A to B was given by the transition probability from A to B minus the transition probability from all other events to B[3]. This latter quantity is computed by the proportion of time that, given a non-A event on trial n, B occurred on trial n+1. For the cause-effect relation, the ΔP value was high (.91), as was, correspondingly, the power PC value (.91), and thus by prominent statistical models of causal inference, should lead to strong causal judgments (Cheng, 1997; Griffiths & Tenenbaum, 2005). On the other hand, the relation from effect to cause is much weaker: the probability of the cause on trial n+1 given the effect on trial n is .25, and also relatively high given other events, resulting in a ΔP value of .05 (power PC .06). Thus, these models would predict that participants should judge that the cause produces the effect, but not that the effect produces the cause.

[2]Participants were required to have a 95% approval rating and have an IP address located in the United States.

[3]We take ~A as indicating the set of events in which A is absent, by normal assumptions of logic and probability. Nonetheless, it is important to point out that the set of alternative events to A can at times be a subset (Cheng & Novick, 1990; Shanks, 1995). If there is no appropriate set of alternative events in our scenario, then the probability of B given ~A is 0, and ΔP is still high from cause to effect, and substantially lower from effect to cause. The present design was not intended to be a test discriminating models of causal reasoning, but rather, to present statistical dependencies that would by all such models be considered strong evidence towards a causal link. However, our recent work indicates that ΔP captures learning in exactly this kind of paradigm, such that the probability of B given ~A is indeed computed as calculated here (Leshinskaya & Thompson-Schill, 2018).

Two other events appeared in the sequences and were designated *frequent* and *rare;* these were designed to occur slightly more often and very much less often than the other events, respectively, while the cause and effect had equal frequencies. This was specified by the stationary distribution of the transition probability matrix, in which rare would occur on 3% of trials, frequent on 28%, and cause and effect on 20%. The frequent and rare events were relatively often (70% and 49%) followed by static to avoid giving the impression of a causal relation from frequent to cause (or other events). The frequent vs rare manipulation was used to test whether a frequency difference between events in the presence of object 1 vs 2 would lead to causal attribution—an alternative to the conditional/dispositional view—as well as a basic attention check (see below).

The specific events assigned to each abstract type in the matrix were randomized across subjects and varied systematically between objects. Thus, for a given subject, one pair of events was designated as the *effect* and the *cause* for object 1, and swapped for object 2 (so that the cause became the effect and vice-versa). Any of the 6 non-static events could serve as the effect or cause, with the constraint that if the effect was ambient, the cause was object-based, and vice versa. Two other events were chosen to be *rare* and *frequent*, one of which was object-based and the other ambient; these assignments also swapped between objects. Thus, each participant was exposed to 4 of the 6 possible types of events (plus static).

Sequences were generated by running a weighted walk: each step was chosen according to the transition probabilities specified in the row of the transition matrix corresponding to the previous step. Sequences contained 250 events, which were split into a shorter preview video of 75 events and a longer video of 175 events. To ensure each walk was a good reflection of the requested transition matrix, the walks were generated iteratively and verified until they met several criteria: the mean difference between the generated and requested matrix transition probabilities was below an absolute value of .0002 (reflecting the average difference between the specified transition probabilities and actual transition probabilities between each pair of events); the standard deviation between total counts of the cause and effect events was less than 3.5 (which amounted to a difference of fewer than 5 occurrences); and the rare event occurred at least 4 times. Figure 1 displays the mean, obtained transition matrix across all included subjects. With respect to obtained frequencies: the effect event occurred 1.25 trials more for object 2 than object 1, averaging across subjects, a difference that is unlikely perceptible over a 250-trial sequence. The cause event occurred an average of 1.29 trials more than the effect event, while the frequent event occurred for one object 62.90 more times than for the other object.

Procedure—The task was implemented using JavaScript and presented in participants' web-browsers via the Mechanical Turk interface. Participants were instructed as follows: "You will see animated videos about two different novel objects. Pay close attention to each video and learn as much as you can about the objects and events. You will be asked questions about the videos and will receive a bonus for accuracy of up to $4." They then saw a preview video about each object lasting 75 trials, or approximately 1.5 minutes. Above the video for its entire duration, participants saw text indicating the object's name ("this object is called a sibbie / thale") and a reminder to pay close attention. After the two previews, they saw longer videos about each object, each lasting 175 trials (~4.5 minutes). During these

videos, they were additionally instructed to try to predict what will happen, and hit the 'a' key whenever "something unexpected happened". This was to keep participants engaged and was not a dependent measure. Because the two objects' properties (names, shapes, and event assignments) were randomized, their order of presentation was not systematically linked to any of these properties; 'object 1' thus refers to the first object the participant saw and 'object 2' the second.

Measure of spontaneous causal attribution.—After both videos for an object were complete, subjects were asked to "describe what you have learned about sibbies / thales", for each object; they typed their response into a separate text box for each. This was used to spontaneous ascription of causality. Descriptions were coded for use of the term 'causes' or 'makes'.

Attention check.—Attention check questions asked participants to judge the relative frequency of certain events. The *rare event* question asked participants to identify which event occurred more than once, but the least often, from a list of 6 event descriptions, which included events that never occurred in either video. Only one option could be selected.

After the second object only, participants saw a *relative frequency question*, asking "Which event occurred more often for this object that the previous one?" and shown a list of 6 event descriptions; they could choose just one.

Participants were included if, on the rare event questions for both objects, they did not choose an event that never occurred, by the reasoning that they would be tempted to select this option only if they had not watched the video in its entirety. They could also be included if they were accurate on the relative frequency question for object 2 and the rare event question for object 1. These criteria were determined using a pilot sample, to balance having a reasonable rate of inclusion while not including participants with overly low accuracy on the other learning measures.

Familiarity forced-choice test.—Following all videos and attention check questions for both objects, participants were given a forced-choice test to assess how well they had learned the pairwise predictive relationships among the events. This follows typical procedure in statistical learning paradigms (e.g., Turk-Browne et al., 2005), and is a useful measure because it allows a clear definition of chance. Questions were blocked by object with order of objects randomized. On each trial, two videos were played sequentially, one on the left and one on the right, and participants had to choose which one was more typical or familiar. Presentation side was randomized for each trial. Each of the two videos showed a pair of events. One was a strong (high transition probability) pair, and the other a weak (low transition probability) pair (except filler items, see below). To respond, participants clicked on one of two buttons that appeared above each video after they both finished playing. They had the option to replay the videos before making their selection, but not to change their answer or to move on without answering. No feedback was given. Participants' accuracy was computed as the average across four forced-choice trials.

Because there was only one strongly predictive event pair (the cause-to-effect transition), the question set was composed of the cause-effect pair as the high probability pair compared to various weaker pairs (effect to cause, effect to frequent, and cause to frequent). The critical pair was the comparison between cause-effect and effect-cause videos, and was presented 4 times; the other questions were presented twice each. To avoid cuing participants to the right answer on critical pairs, via their relative frequency within the other test items, filler questions presented the same (non-critical) questions with the effect-to-cause event as the "high probability pair" in place of the cause-to-effect pair, making them balanced in frequency. Filler trials did not have a correct answer and were not analyzed. Questions were presented in randomized order.

Sentence acceptability test.—Participants were asked to rate verbal statements on a 1–5 scale, in which 1 = Definitely False, 2 = Likely False, 3 = Unsure, 4 = Likely True, and 5 = Definitely True. This scale was chosen so that responses could be compared to a meaningful midpoint (ratings above 3 suggest acceptance). Furthermore, by framing the question in terms of degree of belief, rather than probability, this allowed statements to be rejected because they were poor descriptions, rather than simply rarely occurring phenomena. The object's image appeared above the question as a mnemonic, and its name was used in the question text. Questions were shown in randomized order but blocked by object, with order of objects randomized and interspersed in randomized order with the forced-choice test.

The questions were of four kinds: *order questions, unordered relation questions, causal questions, and frequency questions*. These were used to probe the nature of participants' interpretation of the statistical evidence they were exposed to.

Order questions described a predictive relationship between a pair of events using the term "after", for example: "After the light flashes, sibbies tend to tilt". Such questions were asked about the cause and effect events. *Causal questions* described the relations using causal language, for example, "Light flashing causes sibbies to tilt," and was asked about the cause and effect events in both directions (cause-to-effect and effect-to-cause). These were used to probe the possibility that participants saw statistical dependencies as predictive, but not causal.

Causal questions were also asked in a 'simple' format, for example, "Sibbies cause the light to flash". These were asked about the effect event if it was ambient, and about the frequent event if it was ambient. (It did not make sense to ask the question of an object-based event such as tilting). This allowed us to probe judgments attributing causal properties to the object itself, as opposed to an event explicitly.

Unordered relation questions probed knowledge of association without demand to retrieve order; for example, "Light flashing and sibbies tilting are strongly related", and were asked about cause and effect events as well as frequent and effect events. This was used to assess whether participants represented directionality, given prior work suggesting that associative representations could be symmetrical (Asch & Ebenholtz, 1962; Endress & Wood, 2011; Kahana, 2002).

Finally, *frequency questions* asked subjects to judge event frequency and compare relative frequency of the frequent and rare events. Simple statements read, for instance, “Confetti often tends to appear around Thales”. Relative statements read, for example, “Sibbies’ changing color is more frequent than confetti appearing around them,” for both frequent and rare events in both directions. This served as the basis for the attention check. Altogether, this formed a total of 12 questions per object.

Statistics presented are two-tailed, planned comparison t-tests, with an alpha of .05. Effect sizes (Cohen’s *d*) and confidence intervals (95%) are reported for tests of principal hypotheses.

Results

Familiarity forced-choice test—Participants were reliably accurate on the critical trials of the forced-choice test, which required them to discriminate between cause-to-effect and effect-to-cause event pairs (e.g., tilting followed by bubbles vs. bubbles followed by tilting), for the object learned first ($M = 70\%$, $SE = 0.05$, $t(56) = 4.05$, $p < .001$) and the object learned second ($M = 67\%$, $SE = 0.05$, $t(56) = 3.49$, $p < .001$), with no difference between objects ($t < 1$). Thus, participants had reliable access to the predictive statistics between the relevant events, allowing us to query how they interpreted them conceptually.

Spontaneous causal attribution—Participants were asked to describe in their own words “what they learned about” each object immediately following each video presentation. This allowed us to query their spontaneous, conceptual interpretation of the stimuli. We found that 35/57 participants accurately noticed and described the predictive pattern between cause and effect, for at least one object; and 32 of these described it in the correct direction. 20 participants accurately described the predictive patterns for *both* objects, and 14 of these described both of them in the correct direction. Critically, however, only 4 participants used any kind of causal language in their descriptions. This suggests that while the majority of participants had explicitly noticed the predictive dependencies, very few of them saw these statistics as causal. It should be noted that participants who did not describe the correct predictive patterns were also not reliably accurate on the forced-choice familiarity test, for either object alone or collapsing across objects (ps > .3). This suggests that these participants were not simply reluctant to offer descriptions, but likely failed to learn the statistics.

Sentence acceptability test—Analogously to most causal reasoning experiments, we asked participants to provide causal judgments by rating the extent to which they believed event A caused event B. These judgments were solicited, unlike the spontaneous causal judgments, but unlike in past experiments, were given as a surprise measure—participants were not expecting to make such judgments while viewing the evidence.

Causal statements about the effect and cause.: These statements described the relationship between cause and effect events in causal terms (e.g., “Thales’ tilting causes bubbles to appear”), and were evaluated on a 1 – 5 scale where 5 indicated “Definitely True”, 1 indicated “Definitely False”, and 3 indicated “Unsure”. We thus compared

participants' ratings to a mean of 3 to test acceptance or rejection. As shown in Figure 2B, such statements were strongly accepted when they were in the correct direction (object 1: M = 3.68, SE = 0.16, CI [3.37, 3.99], $t(56) = 4.43$, $p < .0001$, $d = .59$; object 2: $M = 3.58$, SE = 0.15, CI [3.29, 3.87], $t(56) = 3.97$, $p < .001$, $d = 0.53$), but rejected when they were in the reverse, incorrect direction (from effect to cause), for object 1 ($M = 2.54$, $SE = 0.16$, CI [2.23, 2.85], $t(56) = -2.95$, $p = .005$, $d = -0.39$), though neither accepted nor rejected for object 2 ($M = 2.74$, $SE = 0.16$, CI [2.42, 3.05], $t(56) = -1.67$, $p = .10$); however, there was no difference between objects ($t < 1$) and the effect held overall ($t(56) = 2.86$, $p = .006$, d = −0.22). Causal statements in the right direction were accepted significantly more than causal statements in the reverse direction for both objects (object 1: $t(56) = 4.74$, $p < .001$, $d = 0.63$; object 2: $t(56) = 3.21$, $p = .002$, $d = 0.43$).

Thus, participants attributed causality to strongly predictive events more than to weakly predictive ones, indicating that they were not attributing causality indiscriminately, but were using statistical dependency to do so. This confirms that models based on ΔP are good predictors of solicited causal judgments (though does not contradict the argument that more than just ΔP is needed to account for causal reasoning).

Causal vs. predictive judgments.: One possible explanation for participants' acceptance of the causal statements were that they were assuming that causal statements were simply paraphrases of the non-causal, predictive statements. To assuage these worries, we report findings from a related experiment (Leshinskaya & Thompson-Schill, 2018b). In this experiment, participants were taught predictive information about events in a very similar paradigm, in the presence of similar kinds of objects, albeit with additional feedback, such that they had highly robust representations of event order. At the end of the study, participants evaluated statements phrased as follows: "For two of the objects, their movements predicted the occurrence of another event (e.g., the appearance of a light flash, bubbles, or stars). To what extent did you perceive this relationship as causal? Did these objects seem to cause this event?" Thus, the question already assumed predictive knowledge, which they had already demonstrated, and probed the causal interpretation of that predictive relationship as a conceptually distinct attribute. Although no causal language had been used at any point in the experiment, participants largely accepted this causal interpretation, both for events predicted by object movements ($M = 4.04$, $SE = 0.20$, CI [3.64, 4.44], $t(35) = 5.27$, $p < .001$, $d = 0.88$) and for object movements predicted by events ($M = 3.98$, SE = 0.21, $t(35) = 4.53$, $p < .001$, $d = 0.75$).

Predictive non-causal statements.: Results regarding predictive non-causal statements are reported in Appendix A. The key finding was that participants were equally accurate in responding to ordered predictive statements as for the unordered statements relating cause and effect, $t(56) = 1.45$, $p = .154$. This suggests that participants did not discard directionality information about the predictive relations (c.f., Endress & Wood, 2011; Turk-Browne & Scholl, 2009) and that directionality information was recalled at no greater cost than non-directed association.

Attributions to objects.: We were also interested in causal attributions to the *objects* in the event streams. Participants reliably accepted statements that described the objects as being

causes of the effect event, such as, “Thales cause the light to flash” (object 1: $M = 3.84$, $SE = 0.17$, CI [3.38, 4.29], $t(30) = 3.76$, $p = .001$, $d = 0.68$; object 2: $M = 3.85$, $SE = 0.13$, CI [3.44, 4.25], $t(25) = 4.28$, $p < .001$, $d = 0.84$), shown in Figure 2B. This suggests that participants accepted an attribution of causality to the object itself. Because the effect event (e.g., light flash) occurred equally often in the presence of each object, these judgments were not due to the relative frequency of the effect in the presence of the two objects. On the other hand, the frequent event was more frequent in the presence of one object than the other (28% vs 3%) but was *not* strongly or uniquely predicted by any other event in the sequences. We thus also asked whether participants saw a causal relation between the object and the frequent event. We first ensured that participants were sensitive to these properties of the frequent event; these results are reported in Appendix A under *Frequent Event Validation Measures.*

To assess causal judgment of the frequent event, we used the simple causal statements (such as, “sibbies cause confetti to appear”), for the object whose frequent event was ambient. These results are shown in Figure 2C. We found no reliable acceptance of these causal statements ($t(24) = 1.00$, $p = .33$). Conversely, in the same participants, we found significant acceptance of the cause-to-effect statements ($M = 3.96$, $SE = 0.17$, CI [3.43, 4.49], $t(24) = 3.77$, $p < .001$, $d = 0.75$), which was reliably greater than of the frequent event causal statements ($t(24) = 2.63$, $p = .015$, $d = 0.53$).

To ensure that this was not due to differences in these questions’ wording style, we also compared causal judgments of the frequent event to simple causal statements about the effect event, which read, for example, “Thales cause the light to flash”. Acceptance of these statements was also assessed only for the object which had an ambient effect; this reduced the sample size of responses. Nonetheless, participants reliably accepted simple causal statements about the effect ($M = 4.05$, $SE = 0.16$, CI [3.49, 4.61], $t(19) = 3.94$, $p = .001$, $d = 0.88$), and more so than for the frequent causal statements ($t(8) = 3.82$, $p = .005$, $d = 1.27$).

Discussion

Participants watched continuous streams of events, in which one pair of events was strongly predictive in one direction: for example, an object tilting strongly predicted bubbles appearing, but not vice versa. According to statistical theories of causal attribution, participants should judge tilting to cause bubbles, more so than the reverse, by virtue of their relative statistical dependence (Buehner et al., 2003; Cheng, 1997; Cheng & Buehner, 2012; Griffiths & Tenenbaum, 2009; Spellman, 1996). Indeed, when probed to evaluate causal statements, participants’ judgements followed this prediction. Critically, however, our participants had not been told in advance of exposure that they would be making causal judgments. This result suggests that causal inference can, in fact, be informed by statistics learned independently of any explicit goal to evaluate causality. This supports the idea that, to some degree, the presence of the right statistical pattern, observed without the intention to judge causality, can indeed inform the way learners build causal models of their world. The relatively naturalistic design of our task increases our confidence that this kind of learning could take place in everyday observational experience.

However, this conclusion falls alongside an important caveat. Our most unsolicited measure of causal perception—where participants described in their own what they had learned—showed that fully spontaneous causal perception was rare: only 4 participants of 57 gave descriptions with causal language. This suggests that the presence of statistical dependency can inform causal questions when presented, but does not give rise to the perception or recognition of causality on its own. This presents a problem for the strong hypothesis that the concept *cause* is immediately recognized when such dependency is detected. Although statistical dependency was indeed reliably detected, as shown by forced-choice questions and the unsolicited, freeform responses, it was suspended of a causal interpretation until the possibility was suggested by our questions.

Thus, towards our question of whether statistical contingency can drive participants to see causation, we conclude statistical contingency does not obligatorily give rise to a sense of causality, but enables it optionally. Prior to encountering causal questions, participants maintained robust predictive knowledge, suspended from causal belief. Once the question arose, they were able to put it appropriately to use.

One might argue that using the concept *cause* in self-generated descriptions requires a higher level of confidence than circling “Likely True” on a 5-point causal scale. That may be, but the latter—a solicited judgment on a numerical scale—has been taken as a principal measure of causal inference in much prior work. One might restate our findings as indicating that predictive information is insufficient for generating strong, unsolicited causal judgment, but sufficient for weaker, or less confident, solicited judgment. The point nonetheless holds: predictive information alone does not give rise to a *strong* sense of causality, at least absent of strong prior expectations for seeing it, as here. However, to the extent that solicited numerical ratings are a reasonable measure of causal judgment (as they have been treated in the literature), predictive information can be appropriately informative.

The second question we investigated was how causality might be attributed to objects. At face value, statistical models of causal attribution might predict that objects are granted causal properties much as events are: by virtue of their co-variation with an effect. According to the dispositional/contingent view, however, the continual presence of objects is not itself so much a causal event, but is rather a conditioning context for causal relations among other events in its presence (Mumford, 1998; Ryle, 1959). These alternatives have been rarely tested because in typical paradigms, events and objects are not explicitly distinguished—for instance, ‘medicine’ is both an object and the event of taking it (as noted by Mayrhofer & Waldmann, 2015). We found support for the dispositional/conditional view, though not to the exclusion of the former.

When an effect event was contingent on an object-based event, such as the object’s movement or color change, the object was granted causal powers: participants reliably accepted that “thales cause the light to flash” when light flashing depended on thales’ movement or color change. This was true even though the light flash occurred equally often in the presence of the other object (and thus, its ΔP value low with respect to object presence). Critically, participants accepted this statement more than “thales cause confetti to appear”, when confetti was an event which did covary between the two objects: it occurred

frequently around thales, but less frequently around the other object. However, this event was not contingent on the object's movement (or other event). This implies that causal attribution to an object can arise without statistical dependence on the object's presence; instead, statistical dependency on an object-based event is sufficient and perhaps more effective. This supports the dispositional/conditional view.

It is also worth contrasting our view from that of *enabling conditions.* For example, oxygen "enables" fire, even when it is judged to be caused by a match being struck. Are objects in our experiment enabling conditions in this way? Formal and empirical work (Cheng & Novick, 1991) suggests otherwise: enabling conditions are only given causal powers when they covary with effects, albeit across contexts. For example, in a room without oxygen, fires would no longer occur. On the other hand, a factor that does not covary with an effect is seen as causally irrelevant, rather than an enabling condition. The objects in our experiment, by design, did not covary with their effects: light flashing occurred equally often in the presence of both objects. The difference was that light flashing only *depended on* movement in the presence of one of them. We thus argue that an account of our data requires representing relations hierarchically: top-level factors (here, objects) enable lower-order relations (rather than event occurrences), but are nonetheless attributed causal powers.

We acknowledge a few limitations to our conclusions. First, this finding does not preclude the possibility that in other situations, objects can be seen to cause their effects by virtue of their presence alone (e.g., a table may be said to causally enable support), or that a stronger demonstration of such a dependence could not lead to causal judgment. Second, in all cases, the effect was dependent on an object-based event—an event that took place spatiotemporally on the object, such as a movement or color change. We do not know what kinds of attributions participants would make if the effect dependent on another, non-object based event. Along the same lines, we cannot determine here how the nature of the stimuli influences the kinds of statistics learners attend to. For example, the fact that the dispositional statistic was ascribed to the "object" in our paradigm may not have anything inherently to do with it being a bounded shape, or even that it appeared for long durations. Rather, it may be the very fact that its appearance predicted a relation which led to this style of attribution (Gershman, 2016). We leave these questions for future research.

Finally, we showed that learners were highly sensitive to predictive directionality—essential for spontaneous causal learning to get off the ground in event stream learning scenarios. Not only was directionality discriminated, but it was effectively free: to the extent that we could determine, performance was no better when participants only needed to indicate that cause and effect events were related, than when asked to indicate their direction (results reported in Appendix A). This was not guaranteed by prior findings, which have shown that event contingencies can often be encoded symmetrically (discarding order information), at least in cases where there are no temporal boundaries (pauses) segmenting the cause-effect pair from the stream (Endress & Wood, 2011).

Overall, our findings so far suggest that adult learners possess mechanisms to spontaneously extract information from continuous event streams in the right form to be useful for causal

reasoning, including its use to attributing causal properties to objects. However, two open questions motivate the subsequent two experiments.

The first concerns the nature of the attributional judgment to the objects. We relied on the acceptability of the statement “thales cause the light to flash” to indicate that a causal property is indeed assigned to the object itself. However, participants could have seen this as a short-hand paraphrase of the statement “thales tilting causes the light to flash”, or accepted it because they inferred such a meaning on behalf of the experimenter. In the subsequent experiments, we push this kind of attribution further by measuring object categorization (Experiment 2) and asking participants to determine what they think an object causes, and measuring what they choose (Experiment 3).

EXPERIMENT 2

Introduction

In Experiment 1, participants might have learned that tilting predicts bubbles, and ascribed causality to this predictive structure. However, learning a causal relation between two specific events is not per se the same as identifying the abstract relation “cause”. If what learners extracted was indeed an abstract relational representation, then it should be the case that the same relation with some different participating events should be recognized as similar: it should generalize to a visibly different object whose movements also predict, say, bubbles. Here, we are concerned with the conceptual property “causes X”, where X is one of our specific events; we expect participants to group together objects whose movements also predict X, while placing objects whose movements follow X in a different category. Our aim in this experiment was to establish that this is indeed how participants encoded their experiences.

Simultaneously, participants’ categorization of objects on the basis of predictive relations would bolster the notion such predictive structure was attributed to their identity. If causation is assigned to events, not objects, it would not support object categorization. Broadly, then, if statistics among events are indeed the kind of thing that could affect object property and category learning, and if they do so at an abstract, relational level, it would be supportive of our account of how abstract properties can be learned in bottom-up fashion from experience.

To this end, we presented learners with similar displays as in Experiment 1, but here, objects were given one of two category labels, and participants were told that their objective was to learn about these categories. They were then asked to categorize a test object. We asked whether participants would use the directionality of predictive events to make this decision. If so, it would suggest that this form of relation is seen as relevant to object category membership, that relations can be extracted from experience, and were spontaneously represented in abstract relational form. We also compared category formation on the basis of relational properties to those based on object-relative event frequency, given suggestions in the literature that causal properties may have precedence over ones based on frequency (Ahn, 1998; Gopnik & Sobel, 2000; Rehder, 2003a). Thus, we created two conditions. In the Contingency condition, objects were assigned labels on the basis of similar predictive structure among their events: objects with the same label both caused X, rather than reacted

to X. In the Frequency condition, objects were assigned labels on the basis of event frequency (i.e., the same events were relatively more frequent vs. rare). We then measured whether participants would extend the label appropriately to a third object with the characteristics of one of the categories.

Methods

Participants—80 participants were recruited and tested using Amazon Mechanical Turk (following similar procedures as in Experiment 1). Procedures were approved by the Institutional Review Board of the University of Pennsylvania, and all participants provided electronic consent. Participation was compensated with $2.00 plus a performance-based bonus of up to $3.00. None of the participants had previously performed any similar experiment, which was verified by checking their unique worker ID against a master list of worker IDs of participants in other experiments, and the possibility of which was minimized by only allowing workers who had not previously participated in similar tasks to view it. Of these participants, 37 were female and 43 were male, and their ages ranged from 18 to 74, with a mean of 35.

Stimuli—Stimuli were similar to those used in Experiment 1. Four distinct object shapes were used (Figure 3A). These were all shown, separately one at a time, in the context of animated sequences of events. There was a preview video (50-event, ~1.25 minute) and a longer video (200-event, ~4.5 minute) for each object. As before, the sequences involved 5 different events. This always included static, as well as 4 others selected at random from a pool of 6 (light, bubbles, confetti, tilt, part-move, and color change), and assigned also at random to an abstract type: cause, effect, frequent, and rare, which designated their transition properties as specified in a 5 by 5 transition matrix, depicted in Figure 3B. There were always two object-based events: one was either the cause or the effect and the second was either frequent or rare; the others were ambient.

As before, sequences were generated by a weighted walk governed by the transition matrix. The transition matrix specified that the effect followed the cause 94% of the time, but followed other events < 5% of the time. Transitions to the rare event were all low to ensure it was the least frequent (7% of trials), while the cause, effect and frequent events were equally frequent (22% of trials), as determined by the steady state values of the transition matrix. All of these properties were verified by iteratively generating and checking the walks (as in Experiment 1). We additionally ensured that the frequency of the rare events was consistent among the four objects so that this could not be used as a cue for categorization.

There were two categories of objects, which differed in terms of their sequence properties. Each category was given a label, either *sibbies* or *thales*, assigned randomly for each participant. There were three training objects and one test object, creating two exemplars of one category and one of the other. The choice to present three rather than four objects was primarily to reduce testing time and discourage participants from quitting the task. The 4 object shapes were assigned to the four exemplars in randomized fashion.

Categories were created by swapping the assignments of some of the events to different positions in the transition matrix. In the Contingency Condition, the identity of the cause and

effect events swapped between the two categories. Thus, for example, sibbies might both tilt prior to the light flashing, while thales might tilt after the light flashes. In the Frequency condition, the identity of the frequent and rare events swapped between the two categories. Thus, sibbies might be commonly surrounded by bubbles, while thales might be more commonly surrounded by confetti.

The test object matched category 1 for half of the participants, and category 2 for the other half, in its statistical properties. To ensure that these sets of participants received unbiased and equated stimuli, we yoked participants such that two were given identical training materials and shape assignments, with only the statistical properties of the test object varying. Concretely, for one assignment of events to event types, one assignment of object shapes to object types, and a single set of training object walks, two subjects were tested, whose experience differed only in terms of the sequence properties of the test object they saw. These test objects had the same shape, but inherited the sequence properties of category 1 (with two exemplars) or category 2 (with one exemplar). Any biases towards classifying the test object to the category with fewer or more exemplars would thus cancel out, and could not allow above-chance performance alone.

Procedure.: Participants were instructed that their task was to learn about two categories of objects, sibbies and thales, and randomly assigned to the Contingency or Frequency condition. They saw a short, preview video about each of the three training objects (~1 minute), in randomized order, followed by longer videos (~4.5 minutes), in the same randomized order. During the preview phase, they were asked to watch attentively and learn as much as they could. During the longer videos, they were asked to press the ‘a’ key whenever something unexpected happened, where unexpected was not defined, and to continue to learn about this type of object. The name of the object was written above the video for its duration. Following each longer video, participants were given a freeform question asking them to describe what they learned about that object and what might make it a sibbie or thale. Finally, they were given a fourth longer video (~6 minutes). Ahead of this video, they were told that they will see a video about a new object and to “try to determine whether this object is a sibbie or a thale”. This instruction persisted above the video for its duration. After the video, a freeform question asked participants to describe “in what ways this object is similar to or different from sibbies and thales”. Subsequently, they were asked whether this object is a sibbie or a thale, and to rate its similarity to sibbies on a 5 point scale from Very Dissimilar to Very Similar, and to rate its similarity to thales on a similar scale.

Attention check.: Following each video, participants were asked to select from a list of event descriptions those which *never occurred* during the last video. This list included the 4 which occurred and 2 which never occurred. Unfortunately, a large proportion of participants failed this attention check, suggesting that the event descriptions might have been ambiguous. We thus did not use any attention check and included all participants.

Freeform response coding.: Participants’ freeform responses were coded for whether they accurately noticed any of the contingency or frequency information.

Statistics presented are two-tailed, planned comparison t-tests, with an alpha of .05. Effect sizes (Cohen's *d*) and confidence intervals (95%) are reported for tests of principal hypotheses.

Results

The principal measure of interest was the classification accuracy of the test object. These results are shown in Figure 4. In the Contingency condition, where categories were based on the direction of dependence between two events, participants were significantly accurate (Binomial test against 50%, 25/40 passing, $p = .037$). They were also significantly accurate in the Frequency condition, where categories were based on relative event frequency (Binomial test, 26/40 passing, $p = .021$). There was no difference in success rates between these conditions (Chi square, $p > .9$). We also measured similarity ratings of the test object to the two categories. Across conditions, participants rated the test object as more similar to the matching object category ($M = 4.47$, $SE = 0.13$) than to the non-matching category ($M = 3.88$, $SE = 0.14$), CI [0.15, 1.05], t(79) = 2.67, $p = .009$, $d = 0.50$, though the effect within each condition alone was marginal (Contingency condition: CI [−0.06, 1.26], t(39) = 1.85, $p = .071$, $d = 0.49$; Frequency condition: CI [−0.04, 1.24], t(39) = 1.90, $p = .065$, $d = .50$). There were again no differences between conditions (indeed, the means were identical; $p = 1$).

Participants entered freeform responses describing what they had learned about each object following its (longer-video) presentation. These were scored to measure whether participants explicitly noticed the relevant properties (predictive direction for the Contingency condition, and frequency for the Frequency condition, although both statistics were present in both conditions). Participants often described both kinds of statistics as well as other observations about the objects, such as their shapes, manners of movement, etc. We also coded participants' descriptions of how the test object was similar to or different from the objects in two categories, entering a score of 1 if they mentioned the category-relevant statistic (even if they also mentioned other factors), and 0 otherwise. A score of .5 was given in ambiguous cases.

These measures allowed us to compare the Contingency and Frequency conditions on difficulty: were participants equally likely to pick up on one kind of statistic than another? We found this to not be the case: participants were equally likely to notice/describe the relevant statistic in each condition (25/40 subjects in the Contingency condition; 24/40 in the Frequency condition; Chi Square, $p = 1.00$). They were also equally likely to describe the relevant statistic when describing differences and similarities of the test object to the training objects (20 in the Contingency condition; 21 in the Frequency condition; $p > .80$). This suggests that learning of the relevant information in the two conditions was similar, and thus, test object classification performance similarity was not masking underlying differences in learning that were then counteracted by differences in categorization relevance[4].

[4]Participants rarely described the category-irrelevant statistic (5/40 in the Contingency condition; 13/40 in the Frequency condition) but this difference was marginally significant, Chi Square = 3.512, $p = .061$. This difference between conditions might suggest that contingencies were more salient than frequencies overall, but it still does not imply that the conditions were differentially difficult.

Finally, we asked whether participants who did not explicitly notice the statistical regularities for any object (n = 16) were still able to use them successfully to classify the novel object, perhaps by the use of an implicit impression or other intuitive strategy. We found that they were not able to do so, in either condition alone or collapsing across conditions (ps > .10). This suggests that the representations driving successful performance in this task were explicit.

Discussion

Participants were taught two categories of objects based on the statistics of events surrounding or involving them, in a presentation largely similar to Experiment 1. Each exemplar exhibited two regularities: one strong predictive relation between two events (the 'cause' and the 'effect'), and one frequency difference, in which one event ('rare') was less frequent than the others. In the Contingency condition, object category labels grouped the objects exhibiting the same predictive relation, while in the Frequency condition, object categories grouped objects by the identity of the rare event. The test was to categorize a novel object, which matched one of the two categories on the relevant statistic. We found that participants in both conditions learned the relevant statistics and successfully used them to categorize the test object.

Although this experiment was very simple, it was essential in establishing that the information acquired by our participants was, indeed, the kind that could serve as the basis of category formation, in two ways: first, that predictive event information was encoded at a sufficiently abstract, relational level, and second, that it was attributed to objects and thus relevant to their identity. We rejected the alternative possibility that, when asked to learn "about object categories" with minimal guidance in an SL scenario like ours, learners would not attend to or use event contingencies. Instead, a substantial proportion of participants described such properties, and used them in classification.

The role of causality in categorization of novel objects has been previous shown, when this information is presented explicitly to participants (Ahn, 1998; Rehder, 2003b, 2003a; Sloman et al., 1998). Most of this work investigates causal relations among an object's features, however, rather than between their actions and effect on the world. Relatedly, however, Lien and Cheng (2000) argue that it is rational to group different causes of the same effect into a common category in order to create a coherent model of the world: if two causes both result in an effect, overall predictive strength is maximized if those causes are grouped into a common class. We add to this body of work by showing that this phenomenon can arise even when causal structure is observed passively on the basis of statistics in event streams and not made explicit to the participants.

Past findings have also indicated that causal information is privileged over the frequency of features in category learning (Ahn, 1998; Rehder, 2003b, 2003a; Sloman et al., 1998), and we thus expected that contingency-based categories would be better learned than frequency-based ones. However, this is not what we found: test object classification was significant in both conditions, and we detected no difference between them. Of course, it is possible that we lacked power to detect an effect of condition, or that such effects arise only when these two sources of information conflict and individual subjects must choose between them.

However, it also remains possible that the privileging of causal information does not extend to the kind of learning scenarios presented here, where it was less explicitly causal. This remains an open question for future research.

EXPERIMENT 3

In our final experiment, we aimed to substantiate the claim from Experiment 1 that objects can gain causal properties by virtue of the dependency structure among events in their presence, without an increased occurrence of the event they cause. We addressed the potential shortcoming in Experiment 1 that participants might have accepted causal statements (such as “sibbies cause the light to flash”) due to their beliefs about what the experimenter meant by object causation, rather than due to their own beliefs, or that they saw it as a short hand for other causal statements presented in the experiment (“sibbies’ movement causes the light to flash”). Here, we gave participants the top-down goal to judge object causality, and evaluated whether they used event contingency to do so.

The participants’ task was simple: they were asked to determine what they think each object causes, if anything. They then saw two objects in turn, each embedded in an event sequence. These sequences both contained the same set of events, but the dependency structure among the events varied (Figure 5). For object 1, the hypothesized effect event (say, bubbles) appeared reliably following one of the object’s movements, while the other (‘random’) events were neither strongly nor uniquely predicted by any event. For object 2, one of the events that had been random for object 1 became the effect, such that now it reliably followed one of the object’s movements (for example, confetti became predictable while bubbles became unpredictable). The critical comparison is between the two events which are predictable for one object but random for the other: we expect that participants will judge that object 1 causes bubbles, but not stars, and the reverse for object 2.

These two events were matched on two important properties. First, within and across objects, the subjective (perceived) frequencies of the hypothesized effect and the other-object effect were empirically matched. In addition, both the hypothesized effect and the other-object effect followed the object’s movements equally often—just with different reliabilities. Each object exhibited a second, distinct movement (‘non-causal movement’) which was highly frequent, but followed by equiprobably by all of the non-effect events. The number of times participants saw the other-object effect follow this movement was very close to the number of times they saw the effect follow the other movement. However, only the latter was uniquely predictive.

In summary, we anticipated that participants would judge that both objects had causal properties, but that the effects they caused would vary: only the highly predictable event would be seen as each object’s effect. If participants did not take predictive event structure into account, they should not ascribe different causal effects to the two objects. As described in Experiment 1, these predictions arise from a dispositional/contingent view of how causality is attributed to objects (Mumford, 1998), in that it suggests that objects’ causal properties are *dispositions* that are cashed out in contingencies: here, that the reliable contingency between the object moving and the appearance of bubbles will lead to

participants' judgments that the object causes bubbles. This should be true despite the fact that the bubbles are not particularly frequent events when the object is present. That is, the object causes bubbles not by virtue of being often seen with bubbles, but by virtue of being linked to a conditional relation: that, if it moves in just the right way, bubbles reliably appear.

Methods

Participants—82 participants were recruited and tested using Amazon's Mechanical Turk following similar procedures as previous experiments. Procedures were approved by the Institutional Review Board of the University of Pennsylvania, and all participants provided electronic consent. Participation was compensated with $2.00 plus a performance-based bonus of up to $1.50. Five participants reported experiencing technical glitches during the task and were excluded. None participated in related experiments.

As described below, not all participants passed a learning criterion, but results are described both including (Appendix B) and excluding these participants (Results). In the accuracy-filtered sample (n = 48), 30 were female and 18 were male, and their ages ranged from 21 to 71, with a mean of 37. In the overall sample, 45 were female and 32 were male, with ages ranging from 21 to 71 and a mean of 27. A power analysis on data from Experiment 1, using the comparison of causality ratings for the cause-effect relation vs. the frequent event, indicated that a sample of 30 would be sufficient to detect a similar effect with 80% power (Cohen's d = 0.52). A target sample size of 48 (following performance-based filtering) was chosen in order to also complete two sets of counterbalancing conditions.

Stimuli—Stimuli resembled those in Experiment 1. Two novel objects appeared one at a time in 250-trial (~ 4 minute) animated sequences composed of 6 different events (Figure 5). Two of the events were object-based (tilting and part-moving), and four were ambient (light flash, bubbles, confetti, and leaves falling). It was assumed that learners would expect the effect to be ambient; the design therefore manipulated the statistical properties of the ambient events and then assessed which of them would be seen as the effect. Object shapes were assigned in randomized fashion. Objects here did not have any names in order to simplify memory demand: each had a different color and was referred to by its color (e.g., "the green object").

One of the two object-based events was designated as the "cause" and the other as the "non-causal movement" (randomized across participants), and the four ambient events were designated as the "effect", "random 1", "random 2" and "random 3" in counterbalanced fashion across participants (creating 24 counterbalancing conditions, which were used twice). Event assignments were varied systematically between the two objects. Random event 1 of object 1 became the effect of the object 2, so is henceforth referred to as the "other-object effect". Random event 2, which stayed random in both objects, was used in the probe questions as a control (see below), and is henceforth termed the "random" event.

The predictive relations among all events are shown in the transition matrix in Figure 5 (bottom), which governed the weighted walks used to generate the sequences. Notably, the cause was followed with a ~95% probability by the effect, and the effect was not preceded

by any other event. The 3 random events were equally likely to follow the effect, the non-causal movement, and each other, and were thus neither strongly nor uniquely predicted by any event. No static events appeared (except as the very first and last events). All events had an equal probability of repeating (10%). Thus, neither repetition structure nor static event (pause) occurrence could be used as cues to causality. The same transition matrix was used for both objects; their only difference was an exchange of the effect event and random event 1.

The individual frequencies of the 6 events were such that the cause and the effect each occurred 7% of the time, and the others each occurred 20–21% of the time (on average across subjects, and as specified by the steady state of the transition matrix). Notably, therefore, the effect appeared *less* frequently than any other ambient event in the presence of its respective object, and hence, less frequently than in the presence of the other object. This was done in order to match effect and non-effect events for *subjective* frequency, as detailed in the Pilot section below. As a result, "chunk" (pair) frequency among event movement-ambient event pairs was highly similar: there were 17 vs. 14 instances of the cause – effect pair vs. the non-causal-movement – other-effect pair, on average. These values are likely within the range of discriminability over a 250 event sequence, and thus, chunk frequencies were reasonably well matched between effect and non-effect events.

Procedure—Participants were given the following instructions: "You will watch short (4-minute) animated videos about two objects. Your task is to learn what each object causes, if anything. After the videos, you will be asked about what you have learned and various questions about the videos. Pay close attention throughout as there will be additional questions about the videos." They then saw the 250-event sequence for object 1, followed by a freeform response box asking them what they had learned about this object. They then saw a force-choice familiarity test, as in Experiments 1 and 2, which was composed of the following critical questions: cause - effect vs. effect - cause, cause - other-object-effect, cause - random, and non-causal-movement - random. These were presented in randomized order amongst filler questions that balanced the number of times each kind of event pair was shown. The four non-filler question response accuracies were averaged into an overall score. They then saw a 250-event sequence for object 2, including a reminder to pay attention to the entire video, and similar tests. Performance on the familiarity forced-choice responses was used as inclusion criteria for the accuracy-filtered sample. These participants were required to have an overall average performance above 50% on non-filler questions.

Following all videos and familiarity tests, participants saw a set of conceptual questions. Questions were blocked by object and object blocks were presented in the same order as the learning videos. A picture of the relevant object was shown at the start of its question block, and its name was given by reference to its color ("green object", "blue object", etc.).

Binary causal questions.: The first page always revealed the question, "Which of the following do you think this object caused? Check all that apply." A list of events was shown in randomized order, naming the effect, the other-object-effect, and the random event; these were followed by the options "something else" and "it had no causal effect". These

questions allowed us both to determine what participants thought the object caused, as well as allowing the belief that there was nothing that it caused.

Continuous Causal Questions.: The second page instructed participants to "evaluate how true these statements seem to you", and listed a set of statements that participants could evaluate on a scale from 1 ('Definitely False') to 5 ('Definitely True'), with 3 indicating 'Unsure'. Statement acceptance in absolute terms was evaluated by comparing responses to a rating of 3. Three statements were shown describing the causal properties of the object (for example, "The green object seemed to cause the multi-colored stars to appear"), each with a different effect: the effect, the other-object effect, and the random event. Because all statements were shown simultaneously, it encouraged participants to express their relative beliefs about these different causal possibilities, and the continuous scale allowed finer grained measures of belief. Thus, it served as a complementary measure to the binary questions.

Following the causal questions, participants were asked to evaluate relative event frequency. This ordering decontaminated the causal questions from the influence of frequency questions. These questions were used to validate the assumption that participants did not believe the effect event was more frequent than the other events, neither relative to those that appeared in the presence of its object, nor relative to its occurrence in the presence of the other object.

Object-relative frequency questions.: The first set of frequency questions asked participants to evaluate statements about the relative frequencies of events between the two objects. These read, for example, "When the green object was present, leaves falling happened more often than when the blue object was present." A 5-point scale was shown below each, just as in the continuous causal questions. The object named first in the question was the one being queried in that block; hence, symmetrical statements were asked in the other object's block. There were 3 questions, which probed the effect event, the other-object effect, and the random event.

Event-relative frequency questions.: On a subsequent page, participants were asked to evaluate statements concerning the relative frequencies of the events to each other, within the presently probed object. These read, for example, "When the green object was present, leaves falling happened more often than blue bubbles floating up". Responses were collected with a similar 5-point scale. Events presented for comparison were the effect vs. other-object effect, and the effect vs. the random event, in both directions, creating 4 questions.

Piloting—Two pilot samples were used to approximate the point of subjective equality of event frequency. In the first, 24 participants (19 included based on accurate forced-choice performance, as here), saw largely similar stimuli and questions, but all of the events had equal frequency. On continuous causal questions, participants rated the effect event (M = 4.00, SE = 0.20) as more likely to be caused by the object, relative to the other-object effect (M = 3.05, SE = 0.25), CI [0.36, 1.53], $t(18) = 3.41$, $p = .003$, $d = 0.95$. However, they also rated the effect event (M = 3.84, SE = 0.30) as having appeared more frequently than the other-object-effect (M = 2.74, SE = 0.33), CI [0.30, 1.91], $t(18) = 2.90$, $p = .010$, $d = 0.81$,

which was empirically false. They also believed that each object's effect event was more frequent in its presence than in the presence of the other object (M = 3.92, SE = 0.21, CI [3.49, 4.36], $t(18) = 4.45$, $p < .001$, $d = 1.02$). Thus, participants were not simply unsure about frequency, nor randomly guessing about it, but were systematically incorrect, such that they perceived caused or strongly predictable events as more common. This is in line with prior findings that predictable events are attentionally enhanced, aiding in their recognition, memory and perceptual discrimination (Barakat, Seitz, & Shams, 2013; Otsuka & Saiki, 2016; Zhao, Al-aidroos, & Turk-Browne, 2013). This is a natural explanation for why they were judged as more frequent, as they would be over-represented in memory.

Even though the direction of influence is clearly from higher predictability to increased frequency perception (since only the former varied), it is nonetheless possible that frequency perception, and not predictability, is what influenced causality judgments in this sample. Thus, to avoid subjective frequency perception as an explanation of causal ratings, subsequent experiments reduced the frequencies of the effect (and the cause) relative to the other events. A second, though small, sample of 10 participants (5 included) determined that a relative frequency of 1/2 was still insufficient to compensate for frequency misperception effects ($t(4) = 3.09$, $p = 0.037$, for relative frequency judgment). On this basis, we reduced the frequency to 1/3 (7% vs. 21%) in the present, reported experiment.

Statistics presented are two-tailed, planned comparison t-tests, with an alpha of .05. Effect sizes (Cohen's d) and confidence intervals (95%) are reported for tests of principal hypotheses.

Results

Familiarity forced choice.—Forced choice accuracy was used to verify that participants had learned the relative contingencies among events, prior to probing their causal attributions. We report subsequent results from participants who reached an overall average performance above 50% on this test. In Appendix B, we report details about this measure and report the remaining results from the unfiltered group, which align nearly perfectly.

Continuous causal questions.—The key question in this experiment was which event(s) participants would believe each object causes. We report results from the binary causal questions in Appendix B under *Binary Causal Questions.* Continuous questions asked participants to evaluate the veracity of various statements about objects' causal properties, on a 1–5 scale with 3 indicating unsure. Participants saw a statement describing the object's appropriate effect, the other object's effect, and a random event, as being caused by the object. This served as a measure of participants' beliefs about what each object caused, and we predicted that participants should judge that each object caused its putative effect over and above other events.

A 2 (object) by 3 (event type) ANOVA revealed a main effect of object ($F(1,47) = 5.84$, MSE = 5.84, $p = .041$, partial $\eta^2 = 0.01$) and a main effect of event type ($F(2,188) = 39.00$, MSE = 37.42, $p < .001$, partial $\eta^2 = 0.13$), with no interaction ($F < 1$). The main effect of object indicated that object 1 (M = 3.56, SE = 0.15) was given overall higher acceptability ratings than object 2 (M = 3.27, SE = 0.17; $t(47) = 2.09$, $p = .041$) Given the absence of an

interaction with question, we report planned comparisons collapsing over object (though effects hold in the two objects individually as well).

As shown in Figure 6, causal acceptability of the appropriate effect event was significantly higher than unsure ($M = 4.12$, CI [3.84, 4.41], $t(47) = 7.99$, $p < .001$, $d = 1.15$), whereas it was not reliably so for the other-object effect ($M = 2.96$, CI [2.59, 3.33], $t(47) = -0.22$, $p = .823$) nor for the random event ($M = 3.16$, CI [2.76, 3.55], $t(47) = 0.80$, $p = .427$). Consequently, we found significantly higher causal ratings for the effect event compared to the other-object effect (CI [0.75, 1.58], $t(47) = 5.64$, $p < .001$, $d = 1.02$), and compared to the random event (CI [0.56, 1.38], $t(47) = 4.73$, $p < .001$, $d = 0.82$). This supports the predictions and the findings from the binary causal question.

Comparisons between causality and frequency—We first validated that participants did not believe that the effect event was more frequent than other events, and report these tests in Appendix B, *Validating Assumptions about Frequency Perception.* To statistically establish that frequency perception could not account for causal judgment, we performed a comparison between frequency and causality judgments. First, we found that participants' acceptance of the causal statement about the effect ($M = 4.12$, $SE = 0.14$) was significantly higher than their acceptance of the object-relative frequency statement about the effect ($M = 3.07$, $SE = 0.17$), CI [0.68,1.43], $t(47) = 5.64$, $p < .001$, $d = 0.98$, as shown in Figure 6. That is, they more strongly endorsed statements such as, "The green object causes confetti to appear" than they endorsed "When the green object was present, confetti appeared more frequently than when the blue object was present". Second, we found that causal discrimination was significantly stronger than frequency discrimination: the difference between accepting the effect event and the other-object effect event ($M = 1.17$, $SE = 0.21$) was greater than the difference between accepting the object-relative frequency statements about the effect event and a rating of 'unsure' ($M = 0.07$, $SE = 0.17$), $t(47) = 4.20$, $p < .001$, $d = 0.83$. Finally, participants accepted the causal statement about the effect ($M = 4.12$, $SE = 0.14$) significantly more than they accepted that the effect event was more frequent than the other-object event ($M = 2.86$, $SE = 0.16$), $t(47) = 6.35$, $p < .001$, $d = 1.20$, also shown in Figure 6. Likewise, causal discrimination ($M = 1.17$, $SE = 0.21$) was significantly more accurate than event-relative frequency discrimination ($M = -0.14$, $SE = 0.16$), $t(47) = 5.02$, $p < .001$, $d = 1.01$).

Discussion

When asked to evaluate the causal properties of objects, participants reliably made use of a strong and unique statistical relationship between event pairs – one of the object's movements and an ambient event – to decide what each object caused. Although this strongly predictable event was objectively less frequent than the other events in the object's presence, participants overwhelmingly selected it as the object's effect.

Indeed, even having recognized that one of the unpredictable ("random") events occurred more frequently than the predictable one, participants still chose the predictable, rare event as the effect. Furthermore, the effect was not the only event to follow an object movement: unpredictable events all followed another one of the object's movements, and these

movement-event pairs each occurred nearly as often as the cause-effect pairs (on average, only 3 instances fewer over 250 events). Thus, the only source of this difference is that many different events could follow one movement, while only one event, the effect, reliably followed another. In sum, participants relied on unique predictive relationships among individual events to assign causality to objects, against the grain of frequency information. This supports the dispositional/contingent hypothesis, that participants would use event contingency to assign causality to objects, rather than event frequency in the object's presence.

Event predictability also enhanced the perception of frequency, skewing it away from the evidence: predictable events were systematically perceived as being more frequent than they truly were. This was found in pilot experiments, where predictable events were seen as more frequent than unpredictable events, even when matched for frequency, and in the main experiment, where infrequent predictable events were seen equally as frequent as more frequent, unpredictable events. This corroborates past findings on the salience of predictable events—such events are better noticed and remembered (Barakat et al., 2013; Otsuka & Saiki, 2016). An event which is more salient and memorable will naturally be over-estimated in frequency. Overall, then, not only did participants disregard available frequency information for causal judgment, but predictability itself skewed frequency information to be in line with it. This suggests that event-event predictive information may have precedence over frequency information.

As in Experiment 1, we do not claim that objects can never appear to cause events that occur frequently around them. Our frequency manipulations were not particularly dramatic. If bubbles had appeared 90% of the time that one object was on the screen, and 0% in the presence of the other object, perhaps participants might have accepted that the first object caused the bubbles to appear. Our point was not to rule out that adults ever use frequency, but rather to show that it is not necessary, and that participants are indeed sensitive to a more higher-order structure. Further, the latter appears more plausible as a source of causal knowledge for real-world objects—that a coffee maker enables making coffee not by its presence, but by affecting the relations among others events surrounding it. Although this point is simple and intuitive, it helps explain how certain conceptual properties can appear so abstract: because they rely on a higher-order prediction between an object and an event-event relation, rather than a prediction between two sensory qualities (the object and an event).

GENERAL DISCUSSION

The findings across the three experiments presented here illustrate three main points. Experiment 1 demonstrated that predictive information, in the form of transition probabilities between events, was able to inform later causal judgment, without giving rise to a strong sense of causation spontaneously. This indicates that knowledge relevant to causal inference can be acquired spontaneously from predictive statistics; that is, without an exogenously supplied goal to infer causality. However, prior to the demand to make a causal judgment, causal interpretation of predictive statistics was suspended. This implies a

moderate extent to which causal inference works in the "bottom up" direction, that is, from statistical evidence to the application of the concept *cause.*

Experiment 2 illustrated that predictive event information acquired in this fashion was encoded at an abstract, relational level, and that it was seen as sufficiently relevant for categorization that participants were willing to use this information to classify objects, with similar ease as using co-occurring events to do so. In other words, participants were able to attach to a category label the abstract relation "causes X" and extend it to a new object. This served to bolster the finding from Experiment 1, that predictive structure among events in the presence of an object was a relevant source of evidence for assigning (causal) properties to that object itself.

Experiment 3 further supported this claim, in demonstrating that objects could be granted causal powers specifically on the basis of predictive structure among events in their presence, holding object-effect co-variation constant. We now discuss these findings with respect to prior work.

Bottom-up Causality

Prior work on causality has rigorously explored the statistical conditions under which learners will infer causality on the basis of contingency (Buehner et al., 2003; Cheng, 1997; Gopnik et al., 2004; Griffiths & Tenenbaum, 2009; Rehder, 2014; Spellman, 1996), but have always done so in the context of a "top-down" task, in which their explicit goal was to evaluate causality. Our question was to what extent causal inference can work in the other direction: given the statistical evidence, would participants see causality? This is critical to understanding how contingency information experienced outside of explicit causal reasoning contexts can be used: whether contingency information can be represented with a suspension of a causal interpretation—and reinterpreted afterwards. Indeed, this is what we found: causal interpretation was not obligatory, but fully supported, by environmental contingency statistics[5].

This argument is strongly in line with the view of causality as a conceptual label which can be optionally applied to evidence, but where evidence is represented in a distinct vocabulary —that of co-variation or prediction (Cheng, 1997; Cheng & Buehner, 2012). However, it is inconsistent with a strong view that the concept *cause* denotes a pattern of statistical dependency, such that when the right statistical pattern is observed, causality is recognized. As in our opening example, we expect both directions of inference to occur for concepts like *dog.* This suggests that to elucidate the meaning of the concept *cause*, more elaborated models may need to distinguish between criteria used to *recognize* vs. *evaluate* causation. We expect that recognition may have stricter criteria, and perhaps may rely relatively more on intervention data.

It is important to emphasize that our interest is in how adults use the concept *cause*, and that a model of the meaning of this concept is a rather different question than how well adults

[5]We do not intend to claim that this holds for causal perception of mechanical events (e.g., Schlottmann & Shanks, 1992), which appear more obligatory.

reason in normative ways about causality, or how they distinguish true causality from spurious cases. Covariation is a useful guide towards what *could* be a causal relationship, which ultimately, intervention evidence will bear out (Hattari & Oaksford, 2007; Pearl & Mackenzie, 2018). Just as we use heuristic cues like eyes and movement to infer whether something is alive (Carey, 2009), we use covariation to detect what might be a causal fact—but can be proven wrong if further inferences are not borne out.

Causal Categorization

Our finding that learners are willing to form categories of novel objects on the basis of predictive information in an SL scenario fills an important gap between experiments on statistical learning, which have not explored how event statistics inform object knowledge[6], and experiments on object categorization, which present relational properties explicitly (i.e., verbally), where these properties are already attributed to the objects and are at the right level of abstraction (Ahn, 1999; Goldwater & Gentner, 2015; Jones & Love, 2007; Markman & Gentner, 1993; Rehder, 2003b; Rehder & Ross, 2001; Sloman et al., 1998). Without explicitly specifying that an object causes bubbles, for example, it is possible that a predictive relation between an object's movements and bubbles appearing are seen as a fact about the environment, not relevant to the category membership of the object. However, we found that event statistics were used to assign properties to objects, sufficiently well for learners to categorize novel entities.

This finding bolsters prior work on relational category learning in the visuo-spatial domain (Christie, Gentner, Call, & Haun, 2016; Corral & Jones, 2014; Stuhlmueller et al., 2010; Tomlinson & Love, 2007), which demonstrates that adult learners are able to form categories on the basis of relational qualities present in static images (such as 'above' or 'brighter than'), in supervised contexts. An important open question for both prior work and ours is to what extent such categories are formed without supervision (i.e., when exemplars are not labeled).

Finally, it supports and extends past work using observed events in more ostensive scenarios —for example, work in the blicket detector paradigm (Gopnik & Sobel, 2000; Gopnik, Sobel, Schulz, & Glymour, 2001; Nazzi & Gopnik, 2003)—and in more operant-like paradigms where participants can interact with objects to learn their relations to other objects (Kemp et al., 2010; Tenenbaum & Niyogi, 2003), which also demonstrate relational category learning on the basis of event contingencies. Our paradigm uniquely isolated predictive relations among events as a source of object property knowledge, in absence of any physical interaction by an actor, and substantially fewer top-down cues towards the relevant statistics.

This line of research is important, given that many object kinds are organized around relational properties: particularly so, artifact categories, whose function properties rely on causality (Bechtel et al., 2013; Futó et al., 2010; Hernik & Csibra, 2009; Keil et al., 1998;

[6]In that 'objects' and 'events' are conflated in these studies, since events are the appearances and disappearances of visual symbols or patterns. Our view is that real-world objects are more likely to persist in time, while events are briefer state changes around them. This distinction has allowed us to treat them separately.

Kelemen & Carey, 2007; Kemler Nelson et al., 2000; Träuble & Pauen, 2011; Truxaw et al., 2006). The extent to which predictive relations among observed events are a source of relational properties (like causality) has not been determined: for example, function learning could rely more on physical/mechanical reasoning, or reasoning about actors' intentions (Bloom, 1996; Kelemen, Seston, & Georges, 2012; Matan & Carey, 2001), than on predictive relations. Simultaneously, predictive relations are pervasive aspects of experience readily available to learners. Thus, the extent to which they contribute to learning of real-world categories of objects is a ripe possibility (for similar ideas, see Johnson, Shimizu, & Ok, 2007; Ullman, Harari, & Dorfman, 2012; Wellman, Kushnir, Xu, & Brink, 2016).

How Objects get their Causal Properties

While it has been well established that events are judged to cause other events on the basis of contingency statistics, we suggest that objects (stationary entities that tend to persist in time) obtain their causal properties in a distinct fashion: by a higher level of conditioning over top of predictive event relations. Our findings supported this view, by demonstrating that the occurrence of an event need not be more frequent in the presence of an object for that event to be selectively seen as that object's effect. Rather, a reliable contingency between an object-based event (a movement or color change) and another event was, on its own, highly effective evidence for object causality.

This is strongly in line with the philosophical view that objects possess dispositions that can be expressed as contingencies, and this can give rise to causal attribution (Mumford, 1998, Ryle). Contingencies are probabilistic rules governing how an object is expected to behave in various circumstances; here, this is expressed as the prediction that after the object moves a certain way, a specific result will obtain. A disposition is not the observation that an object tends to move or that an event happens around it, but the dependent between the two. We showed that a specifically dispositional quality enabled causal judgments.

This result puts learning objects' causal properties closely in line with work on context-dependent and hierarchical learning. For example, physical and temporal contexts can serve as 'occasion setters' for the retrieval of another association, e.g., between a shock and tone (Urcelay & Miller, 2010, 2014). Context-based occasion setting is thought to be the result of learners inferring the latent structure governing their world, to best explain the evidence they are confronted with (Coutanche & Thompson-Schill, 2012; Gershman, 2016; Gershman, Blei, & Niv, 2010; Gershman & Niv, 2012). Furthermore, visuo-motor mappings are spontaneously encoded hierarchically by task context (Collins & Frank, 2013). It is possible that attributing causal properties to objects relies on similar, or even shared, mechanisms, where objects act as contexts. This is a direct prediction for future research.

The important implication of our finding is that the ability to encode higher-order contingencies may be an essential part of how even very concrete things can come to have non-concrete properties. If causing bubbles was represented only by virtue of a link between an object shape and the appearance of bubbles, the property "causes bubbles" would have a more direct pointer to something in the world. Such a pointer was insufficient to explain the kind of representations our participants acquired. In Experiments 1 and 2, one object would cause bubbles while another reacted to it, but all appeared with bubbles equally. In

Experiment 3, bubbles also appeared (subjectively) equally with both objects, but were reliably predicted by an object-based event only for one. As such, it was insufficient to represent a pointer to a sensory event to distinguish the objects' causal properties. Instead, it was essential to link the objects to the presence or absence of *relation*, which itself is not in the world—despite its precise definition in statistical evidence. It is learners' use of sophisticated statistical reasoning machinery that can close this gap.

Conclusion

Our aim in this work was to understand some aspects of the cognitive machinery that translates statistically definable aspects of event experience into content useful for conceptual inference—here, the application of the concept 'cause', its assignment to objects, and the formation of novel object categories—in a more or less bottom-up fashion. Our findings make several points about how we do so: that the concept cause is optionally but not obligatorily applicable to pure predictive evidence; that it can be assigned to objects on the basis of higher-order conditioning; and that objects' causal properties can be represented without making reference to the occurrence or non-occurrence of specific events, but to their predictive relational structure alone. Altogether, these findings provide insights into how we employ statistical inference to create abstract representations out of concrete streams of events.

ACKNOWLEDGEMENTS

This work was supported by NIH grants R01EY021717 and R01DC015359 to S.L.T-S. We also thank Domonique Roberts-Mack and Cristina H. Leon for assistance with programming and data analysis.

APPENDICES

Appendix A.: Additional Results for Experiment 1

Ordered predictive statements

These statements described event order in non-causal terms (e.g., "After thales tilt, bubbles tend to appear"). Participants reliably accepted statements that described cause events preceding effect events, for both objects (object 1: $M = 4.07$, $SE = 0.13$, $t(56) = 8.10$, $p < .001$; object 2: $M = 3.84$, $SE = 0.15$, $t(56) = 5.55$, $p < .001$), with no difference between objects ($t = 1.34$, $p = .19$). They accepted these statements more than statements describing effects preceding causes for both object 1 ($M = 2.84$, $SE = 0.17$; $t(56) = 5.09$, $p < 0.001$) and object 2 ($M = 2.93$, $SE = 0.18$; $t(56) = 3.17$, $p = .002$), which they neither accepted nor rejected (ts < 1).

We found that non-causal, predictive statements were accepted to a greater degree than analogous causal statements (object 1: $t(56) = -3.04$, $p = .004$, $d = -0.40$; object 2: $t(56) = -2.32$, $p = .024$, $d = -0.31$). However, there was no difference in how well participants *discriminated* between cause-to-effect vs. effect-to-cause directionality within causal vs. non-causal statement sets; i.e., the difference in ratings between the statements differing in direction (ts < 1). Furthermore, directionality discrimination was highly correlated between causal and non-causal statements (object 1: $r(55) = 0.77$, $p < .001$, object 2: $r(55) = 0.84$, p

< .001), suggesting that individuals made use of their predictive knowledge to make these judgments.

Unordered predictive statements

Although accuracy on ordered predictive statements was high, it was not at ceiling. We can thus ask whether participants performed better on an unordered version of these statements, such as, “Thales tilting and bubbles appearing were strongly related”, which did not require recalling the directionality of the relationship. This would suggest that they discarded directionality information. As shown in Figure 2A, participants strongly accepted unordered statements for both objects (object 1: $M = 4.25$, $SE = 0.13$ $t(56) = 9.71$, $p < .001$; object 2: $M = 3.96$, $SE = 0.15$, $t(56) = 6.43$, $p < .001$), and they accepted them more strongly than similarly phrased, unordered statements describing a strong relationship between unrelated events (the effect and the frequent event), for object 1 ($M = 3.11$, $SE = 0.16$; $t(56) = 5.68$, $p < .001$) and object 2 (M = 3.02, SE = 0.16; $t(56) = 4.09$, $p < .001$), which they did not accept nor reject (ts < 1). However, importantly, we did not find an advantage for the unordered statements (relating cause and effect) over the ordered ones: they were no more likely to be accepted for either object or overall ($t(56) = 1.45$, $p = .154$). This suggests that participants did not reliably discard directionality information about the predictive events (c.f., Endress & Wood, 2011; Turk-Browne & Scholl, 2009).

Frequent Event Validation Measures

To ensure that participants did not see the frequent event as dependent on other events, we used accuracy on the forced-choice questions which contrasted cause-effect pairs with cause-frequent pairs and with effect-frequent pairs; participants had to select cause-effect as more typical than the others, above-chance on average over 4 trials. The resulting means in the included group were 93% and 95% for objects 1 and 2, respectively. Exclusion was done on a per-object basis, so that only the accurate object(s) were included in subsequent analyses; this left 33 responses for object 1, and 26 for object 2.

To establish that these participants also recognized that the frequent events for each object were more frequent for one object than the other, we used the object-relative frequency question that was part of the attention check measure. Because participants were already excluded partly on this basis, it was unsurprising that these participants were 86% correct on this test (in which chance was 1/6). This demonstrated their knowledge that the frequent event of the second object appeared more often in its presence than that of the first object. To ensure they also recognized this of the frequent event for object 1, we used the relative frequency judgment, in which participants had to accept that the frequent event occurred more often than the rare event. We only included object 1 data from participants who gave this statement a rating above 3 (either “Likely True” or “Definitely True”), leaving 21 responses for object 1.

Appendix B.: Additional Results from Experiment 3

Forced-Choice Test Data

In the unfiltered sample, participants were largely accurate, performing reliably above chance (50%) on the forced choice questions for object 1 ($M = 0.65$, $SE = 0.04$, $t(76) = 3.57$, $p < .001$) and object 2 ($M = 0.70$, $SE = 0.04$, $t(76) = 5.63$, $p < .001$), with no difference between objects ($t(76) = -1.37$, $p = .17$). In the filtered sample ($n = 48$), participants' accuracies on the forced-choice test were 84% and 89% on objects 1 and 2 respectively, with no difference between objects ($t(47) = -0.92$, $p = .36$). We chose to report data from both samples. Under one view, only including accurate participants is essential for validity: participants must have picked up on the statistical information in order for us to properly assess whether they use or it or not for causal judgment. Further, because participants were instructed to attend to the entire video and to prepare for questions, which were possible to fully anticipate for object 2, poor performance was an indicator of inattention. Under another view, however, participants' inattention to event contingencies could be driven by their belief that it is not relevant to causation, and thus, they should be included in the evaluation of our hypothesis. We thus report data using both approaches, which align almost perfectly.

Binary causal questions

This measure asked participants to select what they believed each object caused from a list of events, giving them the option to select or not select any of a number of response options. We predicted that participants would be more likely to select the putative effect more often than any other option, for both objects.

A 2 (object) by 5 (response option) ANOVA revealed no effect of object ($F(1,47) = 0.37$, $p = .54$), but a significant effect of response option ($F(4,376) = 42.47$, $MSE = 6.86$, $p < .001$, partial $\eta^2 = 0.26$), and no interaction. Planned comparisons, collapsing across objects, revealed that the appropriate effect event ($M = 81\%$, SE = 0.05) was chosen significantly more often than the other-object effect ($M = 44\%$, SE = 0.06), CI [0.24, 0.51], $t(47) = 4.78$, $p < .001$, $d = 0.96$. It was also chosen more often than the random event ($M = 52\%$, SE = 0.06), CI [0.15, 0.43], $t(47) = 4.35$, $p < .001$, $d = 0.76$, and more than the 'something else' option ($M = 22\%$, SE = 0.05), CI [0.47, 0.72], $t(47) = 9.49$, $p < .001$, $d = 1.70$, and more than the 'nothing' option ($M = 14\%$, SE = 0.04), CI [0.51, 0.85], $t(47) = 7.98$, $p < .001$, $d = 2.15$. Thus, participants indeed chose the hypothesized effect event over and above other options.

Results in the Unfiltered Sample

On binary causal questions, a 2 (object) by 5 (response option) ANOVA revealed no effect of object ($F < 1$, $p = .40$), but a significant effect of response option ($F(4,608) = 36.31$, $MSE = 7.03$, $p < .001$), and a marginal interaction ($F(4,608) = 2.17$, MSE = 0.421, $p = .071$). Planned comparisons were thus performed within each individual object, and revealed mostly similar patterns: the effect event was chosen significantly more often than the other object's effect (object 1: $t(76) = 3.20$, $p = .002$; object 2: $t(76) = 2.59$, $p = .010$), more often than the random event for object 1 ($t(76) = 2.01$, $p = .048$), though not for object 2 ($t(76) =$

1.52, $p = .132$), but more often than the 'something else' option (object 1: $t(76) = 9.33$, $p < .001$; object 2: $t(76) = 6.31$, $p < .001$) and more often than the 'nothing' option (object 1: $t(76) = 7.78$, $p < .001$; object 2: $t(76) = 3.60$, $p < .001$).

On continuous causal questions, a 2 (object) by 3 (event type) ANOVA revealed a main effect of object ($F(1,76) = 7.34$, MSE = 10.91, $p = .008$) and a main effect of event type ($F(2,304) = 29.72$, MSE = 24.68, $p < .001$), with no interaction. Participants accepted the effect event as being caused by the object ($M = 3.86$, $SE = 0.12$, CI [3.63, 4.08], $t(76) = 7.50$, $p < .001$, $d = 0.85$), but we did not find this for the other-object effect ($t < 1$, $p = 0.411$) nor for the random effect ($t(76) = 1.70$, $p = .093$). Correspondingly, participants' causal ratings were significantly higher for the effect event compared to the other-object effect (CI [0.45, 1.04], $t(76) = 5.02$, $p < .001$, $d = 0.68$), and compared to the random event (CI [0.33, 0.91], $t(76) = 4.26$, $p < .001$, $d = 0.56$).

We also found that causal discrimination was significantly stronger than frequency discrimination, ($t(76) = 6.15$, $p < .001$, $d = 0.82$; and $t(76) = 3.94$, $p < .001$, $d = 0.89$), and that participants accepted the causal statement about the effect significantly more than they accepted that the effect event was more frequent than the other-object event, $t(76) = 6.08$, $p < .001$, $d = 0.89$. Likewise, causal discrimination was significantly more accurate than event-relative frequency discrimination, $t(76) = 4.14$, $p < .001$, $d = 0.68$.

Validating Assumptions about Frequency Perception

Event-relative frequency questions.

These questions were used to rule out that for any object or any event, participants believed the effect was more frequent than other events, as an alternative account of why they may have judged it to be what the object causes. To this end, participants were asked to judge whether the effect event occurred more frequently than other events, within the context of each object. Our assumption was verified: accuracy-filtered participants did not reliably accept any statements indicating that the effect event was more frequent than the other-object effect event (ts < 1, ps > .400), individually and collapsing across objects. Neither did participants accept the reversed statements—that the other-object effect was more frequent than the effect (ps > .148; combined across objects, $p = .344$). Thus, subjective frequency was well matched for the effect and other-object effect events, although objective frequency was in fact three times as high for the other-object effect. This corresponds with our pilot findings (see Methods), in which participants judged predictable events (the effect events) as more frequent than unpredictable ones, even when their frequencies were objectively matched. Among the unfiltered group, we likewise found no reliable acceptance that the effect occurred more often than the other-object effect (all ts < 1; ps > .59). However, these participants did accurately accept that the other-object effect was more frequent than the effect ($M = 3.31$, $SE = 0.12$, $t(76) = 2.67$, $p = .01$). Since their predictive knowledge was by definition less precise, this effect is fully consistent with the above findings.

Object-relative frequency questions.

As above, these questions were used to rule out that participants' differential causal judgments were due to differential perceptions of frequency, but this time by comparing the perceived frequencies of events across objects (for example, "When the green object was present, leaves falling happened more often than when the blue object was present"). We again used planned t-tests to evaluate participants' acceptance of these statements in each individual object. We found that none of these statements was reliably accepted (all *t*s < 1, ps > .376*).* This was also true in the unfiltered group (*p*s > .290).

REFERENCES

Ahn W-K (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. Cognition, 69(2), 135–78. [PubMed: 9894403]

Ahn W-K (1999). Effect of causal structure on category construction. Memory & Cognition, 27(6), 1008–1023. [PubMed: 10586577]

Ahn W-K, Kim NS, Lassaline ME, & Dennis MJ (2000). Causal status as a determinant of feature centrality. Cognitive Psychology, 41(4), 361–416. [PubMed: 11121260]

Allan LG (1980). A note on measurement of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Society, 15(3), 147–149.

Asch SE, & Ebenholtz SM (1962). The Principle of Associative Symmetry. Proceedings of the American Philosophical Society, 106(2), 135–163.

Aslin RN, & Newport EL (2012). Statistical learning: From acquiring specific items to forming general rules. Current Directions in Psychological Science, 21(3), 170–176. [PubMed: 24000273]

Barakat BK, Seitz AR, & Shams L (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. Cognition, 129(2), 205–211. [PubMed: 23942346]

Bechtel S, Jeschonek S, & Pauen S (2013). How 24-month-olds form and transfer knowledge about tools: The role of perceptual, functional, causal, and feedback information. Journal of Experimental Child Psychology, 115(1), 163–79. [PubMed: 23465335]

Bloom P (1996). Intention, history, and artifact concepts. Cognition, 60, 1–29. [PubMed: 8766388]

Bramley NR, Gerstenberg T, & Lagnado DA (2014). The order of things: Inferring causal structure from temporal patterns. Proceedings of the Cognitive Science Society, 36.

Bramley NR, Mayrhofer R, Gerstenberg T, & Lagnado DA (2017). Causal learning from interventions and dynamics in continuous time. In Proceedings of the 39th Annual Meeting of the Cognitive Science Society.

Buchsbaum D, Griffiths TL, Plunkett D, Gopnik A, & Baldwin D (2015). Inferring action structure and causal relationships in continuous sequences of human action. Cognitive Psychology, 76, 30–77. [PubMed: 25527974]

Buehner MJ, Cheng PW, & Clifford D (2003). From covariation to causation: a test of the assumption of causal power. Journal of Experimental Psychology. Learning, Memory, and Cognition, 29(6), 1119–40.

Buehner MJ, & May J (2009). Causal induction from continuous event streams: Evidence for delay-induced attribution shifts. The Journal of Problem Solving, 2(2), 42–80.

Carey S (2009). Origin of Concepts. Oxford: Oxford University Press.

Chatterjee A (2008). The neural organization of spatial thought and language. Seminars in Speech and Language, 29(3), 226–238. [PubMed: 18720319]

Cheng PW (1997). From covariation to causation: A causal power theory. Psychological Review, 104(2), 367–405.

Cheng PW, & Buehner MJ (2012). Causal learning In Holyoak KJ & Morrison RG (Eds.), The Oxford Handbook of Thinking and Reasoning (pp. 210–233). Oxford: Oxford University Press.

Cheng PW, & Novick LR (1990). A probabilistic contrast model of causal induction. Journal of Personality and Social Psychology, 58(4), 545–567. [PubMed: 2348358]

Cheng PW, & Novick LR (1991). Causes versus enabling conditions. Cognition, 40(1–2), 83–120. [PubMed: 1786673]

Christie S, Gentner D, Call J, & Haun DBM (2016). Sensitivity to relational similarity and object similarity in apes and children. Current Biology, 26(4), 531–535. [PubMed: 26853364]

Collins AGE, & Frank MJ (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychological Review, 120(1), 190–229. [PubMed: 23356780]

Corral D, & Jones M (2014). The effects of relational structure on analogical learning. Cognition, 132(3), 280–300. [PubMed: 24858106]

Coutanche MN, & Thompson-Schill SL (2012). Reversal without remapping: What we can (and cannot) conclude about learned associations from training-induced behavior changes. Perspectives on Psychological Science, 7(2), 118–134. [PubMed: 26168440]

Dennett DC (1987). The Intentional Stance. Cambridge, MA: MIT Press.

Derringer C, & Rottman BM (2018). How people learn about causal influence when there are many possible causes: A model based on informative transitions. Cognitive Psychology, 102, 41–71. [PubMed: 29358094]

Endress AD, & Wood JN (2011). From movements to actions: Two mechanisms for learning action sequences. Cognitive Psychology, 63(3), 141–171. [PubMed: 21872553]

Fiser J, & Aslin RN (2002). Statistical learning of higher-order temporal structure from visual shape sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(3), 458–467.

Futó J, Téglás E, Csibra G, & Gergely G (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. Cognition, 117(1), 1–8. [PubMed: 20605019]

Garvey C, & Caramazza A (1974). Implicit causality in verbs. Linguistic Inquiry, 5(3), 459–464.

Genone J, & Lombrozo T (2012). Concept possession, experimental semantics, and hybrid theories of reference. Philosophical Psychology, 25(5), 717–742.

Gershman SJ (2016). Context-dependent learning and causal structure. Psychonomic Bulletin & Review, 24(2), 1–25.

Gershman SJ, Blei DM, & Niv Y (2010). Context, learning, and extinction. Psychological Review, 117(1), 197–209. [PubMed: 20063968]

Gershman SJ, & Niv Y (2012). Exploring a latent cause theory of classical conditioning. Learning & Behavior, 40, 255–268. [PubMed: 22927000]

Goldstone RL, Medin DL, & Gentner D (1991). Relational similarity and the nonindependence of features in similarity judgments. Cognitive Psychology, 23(2), 222–262. [PubMed: 2055001]

Goldwater MB, & Gentner D (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. Cognition, 137, 137–153. [PubMed: 25638033]

Goodman N (1955). Fact, fiction and forecast. Cambridge, MA: Harvard University Press.

Gopnik A, Glymour C, Sobel DM, Schulz LE, Kushnir T, & Danks D (2004). A theory of causal learning in children: causal maps and Bayes nets. Psychological Review, 111(1), 3–32. [PubMed: 14756583]

Gopnik A, & Meltzoff AN (1997). Words, thoughts, and theories. Cambridge, MA: MIT Press.

Gopnik A, & Schulz LE (2004). Mechanisms of theory formation in young children. Trends in Cognitive Sciences, 8(8), 371–7. [PubMed: 15335464]

Gopnik A, & Sobel DM (2000). Detecting blickets: how young children use information about novel causal powers in categorization and induction. Child Development, 71(5), 1205–1222. [PubMed: 11108092]

Gopnik A, Sobel DM, Schulz LE, & Glymour C (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental Psychology, 37(5), 620–629. [PubMed: 11552758]

Greville WJ, & Buehner MJ (2010). Temporal predictability facilitates causal learning. Journal of Experimental Psychology: General, 139(4), 756–771. [PubMed: 21038987]

Griffiths TL, & Tenenbaum JB (2005). Structure and strength in causal induction. Cognitive Psychology, 51(4), 334–384. [PubMed: 16168981]

Griffiths TL, & Tenenbaum JB (2009). Theory-based causal induction. Psychological Review, 116(4), 661–716. [PubMed: 19839681]

Hattori M, & Oaksford M (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: model comparison and rational analysis. Cognitive Science, 31(5), 765–814. [PubMed: 21635317]

Hernik M, & Csibra G (2009). Functional understanding facilitates learning about tools in human children. Current Opinion in Neurobiology, 19(1), 34–8. [PubMed: 19477630]

Johnson SC, Alpha Shimizu Y, & Ok SJ (2007). Actors and actions: The role of agent behavior in infants' attribution of goals. Cognitive Development, 22(3), 310–322. [PubMed: 18591991]

Jones M, & Love BC (2007). Beyond common features: The role of roles in determining similarity. Cognitive Psychology, 55(3), 196–231. [PubMed: 17094958]

Kahana MJ (2002). Associative symmetry and memory theory. Memory & Cognition, 30(6), 823–840. [PubMed: 12450087]

Keil FC, Smith WC, Simons DJ, & Levin DT (1998). Two dogmas of conceptual empiricism: implications for hybrid models of the structure of knowledge. Cognition, 65(2–3), 103–35. [PubMed: 9557380]

Kelemen D, & Carey S (2007). The essence of artifacts: Developing the design stance In Laurence S & Margolis E (Eds.), Creations of the Mind: Artifacts and their representation (pp. 415–449). Oxford: Oxford University Press.

Kelemen D, Seston R, & Georges L Saint. (2012). The designing mind: Children's reasoning about intended function and artifact structure. Journal of Cognition and Development, 13(4), 439–453.

Kemler Nelson DG, Frankenfield A, & Morris C (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. Cognition, 77, 133–168. [PubMed: 10986365]

Kemp C, Tenenbaum JB, Niyogi S, & Griffiths TL (2010). A probabilistic model of theory formation. Cognition, 114(2), 165–96. [PubMed: 19892328]

Kirkham NZ, Slemmer JA, & Johnson SP (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. Cognition, 83, B25–B42. [PubMed: 11869727]

Leshinskaya A, & Thompson-Schill SL (2018). Inferences about uniqueness in statistical learning. In Proceedings of the 40th Annual Meeting of the Cognitive Science Society.

Leshinskaya A, & Thompson-Schill SL (2018b). Transformation of event representations along middle temporal gyrus. BioRxiv, (11 11 2018), 1–40.

Lien Y, & Cheng PW (2000). Distinguishing genuine from spurious causes: a coherence hypothesis. Cogn Psychol, 40(2), 87–137. [PubMed: 10716875]

Markman AB, & Gentner D (1993). Structural alignment during similarity comparisons. Cognitive Psychology, 25(4), 431–467.

Markman AB, & Stilwell CH (2001). Role-governed categories. Journal of Experimental & Theoretical Artificial Intelligence, 13(4), 329–358.

Matan A, & Carey S (2001). Developmental changes within the core of artifact concepts. Cognition, 78(1), 1–26. [PubMed: 11062320]

Mayrhofer R, & Waldmann MR (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. Cognitive Science, 39(1), 65–95. [PubMed: 24831193]

Mumford S (1998). Dispositions. Oxford: Oxford University Press.

Murphy GL, & Medin DL (1985). The role of theories in conceptual coherence. Psychological Review, 92(3), 289–316. [PubMed: 4023146]

Nazzi T, & Gopnik A (2003). Sorting and acting with objects in early childhood: An exploration of the use of causal cues. Cognitive Development, 18(3), 299–317.

Otsuka S, & Saiki J (2016). Gift from statistical learning: Visual statistical learning enhances memory for sequence elements and impairs memory for items that disrupt regularities. Cognition, 147, 113–126. [PubMed: 26688065]

Pearl J, & Mackenzie D (2018). The book of why. New Yok: Basic Books.

Penn DC, & Povinelli DJ (2007). Causal cognition in human and nonhuman animals: a comparative, critical review. Annual Review of Psychology, 58, 97–118.

Pinker S (1989). Learnability and Cognition: The acquisition of argument structure. Cambridge, MA: MIT Press.

Rehder B (2003a). A causal-model theory of conceptual representation and categorization. Journal of Experimental Psychology. Learning, Memory, and Cognition, 29(6), 1141–1159.

Rehder B (2003b). Categorization as causal reasoning. Cognitive Science, 27(5), 709–748.

Rehder B (2014). Independence and dependence in human causal reasoning. Cognitive Psychology, 72, 54–107. [PubMed: 24681802]

Rehder B, & Ross BH (2001). Abstract coherent categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(5), 1261–1275.

Rescorla R, & Wagner A (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II Current Research and Theory, 21(6), 64–99.

Rottman BM, & Keil FC (2012). Causal structure learning over time: Observations and interventions. Cognitive Psychology, 64(1–2), 93–125. [PubMed: 22155679]

Ryle G (1959). Concept of Mind. Chicago: University of Chicago Press.

Saffran JR, Aslin RN, & Newport EL (1996). Statistical learning by eight-month-old infants. Science, 274(5294), 1926–1928. [PubMed: 8943209]

Schlottmann A, & Shanks DR (1992). Evidence for a distinction between judged and perceived causality. The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 44(2), 321–42. [PubMed: 1565803]

Schulz LE, & Gopnik A (2004). Causal learning across domains. Developmental Psychology, 40(2), 162–76. [PubMed: 14979758]

Shanks DR (1985). Forward and backward blocking in human contingency judgement. Quarterly Journal of Experimental Psychology, 37(B), 1–21.

Shanks DR (1995). The Psychology of Associative Learning. Cambridge, UK: Cambridge University Press.

Sloman SA, & Lagnado DA (2015). Causality in thought. Annual Review of Psychology, 66(1), 223–247.

Sloman SA, Love BC, & Ahn W-K (1998). Feature centrality and conceptual coherence. Cognitive Science, 22(2), 189–228.

Sobel DM, & Kirkham NZ (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. Developmental Science, 10(3), 298–306. [PubMed: 17444971]

Spellman BA (1996). Acting as intuitive scientists: Contingency judgments are nade while controlling for alternative potential causes. Psychological Science, 7(6), 337–342.

Stuhlmueller A, Tenenbaum JB, & Goodman ND (2010). Learning structured generative concepts. Proceedings of the 32nd Annual Conference of the Cognitive Science Society, 2296–2301.

Tenenbaum JB, & Niyogi S (2003). Learning causal laws. Proceedings of the 25th Annual Meeting of the Cognitive Science Society, 1152–1157.

Tomlinson MT, & Love BC (2007). Relation-based categories are easier to learn than feature-based categories. Proceedings of the 29th Annual Conference of the Cognitive Science Society.

Tomlinson MT, & Love BC (2010). When learning to classify by relations is easier than by features. Thinking & Reasoning, 16(4), 372–401.

Träuble B, & Pauen S (2007). The role of functional information for infant categorization. Cognition, 105(2), 362–79. [PubMed: 17129581]

Träuble B, & Pauen S (2011). Cause or effect: what matters? How 12-month-old infants learn to categorize artifacts. The British Journal of Developmental Psychology, 29(Pt 3), 357–74. [PubMed: 21848735]

Truxaw D, Krasnow MM, Woods C, & German TP (2006). Conditions under which function information attenuates name extension via shape. Psychological Science, 17(5), 367–71. [PubMed: 16683921]

Turk-Browne NB, Jungé J, & Scholl BJ (2005). The automaticity of visual statistical learning. Journal of Experimental Psychology. General, 134(4), 552–64. [PubMed: 16316291]

Turk-Browne NB, & Scholl BJ (2009). Flexible visual statistical learning: transfer across space and time. Journal of Experimental Psychology: Human Perception and Performance, 35(1), 195–202. [PubMed: 19170482]

Ullman S, Harari D, & Dorfman N (2012). From simple innate biases to complex visual concepts. Proceedings of the National Academy of Sciences, 109(44), 18215–18220.

Urcelay GP, & Miller RR (2010). Two roles of the context in Pavlovian fear conditioning. Journal of Experimental Psychology. Animal Behavior Processes, 36(2), 268–80. [PubMed: 20384406]

Urcelay GP, & Miller RR (2014). The functions of contexts in associative learning. Behavioural Processes, 104, 2–12. [PubMed: 24614400]

Wellman HM, Kushnir T, Xu F, & Brink KA (2016). Infants use statistical sampling to understand the psychological world. Infancy, 21(5), 668–676.

Zhao J, Al-aidroos N, & Turk-Browne NB (2013). Attention Is spontaneously biased toward regularities.

CONTEXT OF RESEARCH

The first author, A.L., has previously investigated conceptual knowledge of attributes that lack a clear physical or sensory basis. Specifically, she has studied the neural representation of the goals of actions, the belief attributes of social groups, and abstract functions of artifacts. The current experiments arose from the question of what, in experience, such conceptual knowledge might refer to—what kind of evidence would people use to infer such sorts of attributes? Starting with causality was a natural point, given the importance of causality in function, and given the rich prior work on causal inference. From the perspective of this question, several gaps in knowledge became apparent. Would people *recognize* causality from contingency evidence, outside of explicit causal reasoning contexts? And in what ways would causality be attributed to objects, rather than events? These two questions formed the basis of the current work, which fills those immediate gaps. Answering them is one step towards the broader question of what we mean when we believe an object has causal properties. The importance of relations—of which contingency knowledge is one kind—in conceptual knowledge is a key part of the research program of the second author, S.L.T-S. Both authors see a promising avenue stemming from the current findings towards understanding how relations are extracted from experience and put to conceptual use, ultimately helping explain how the mind comes to have such abstract ideas as *belief*, *communication*, or *nutrition*.

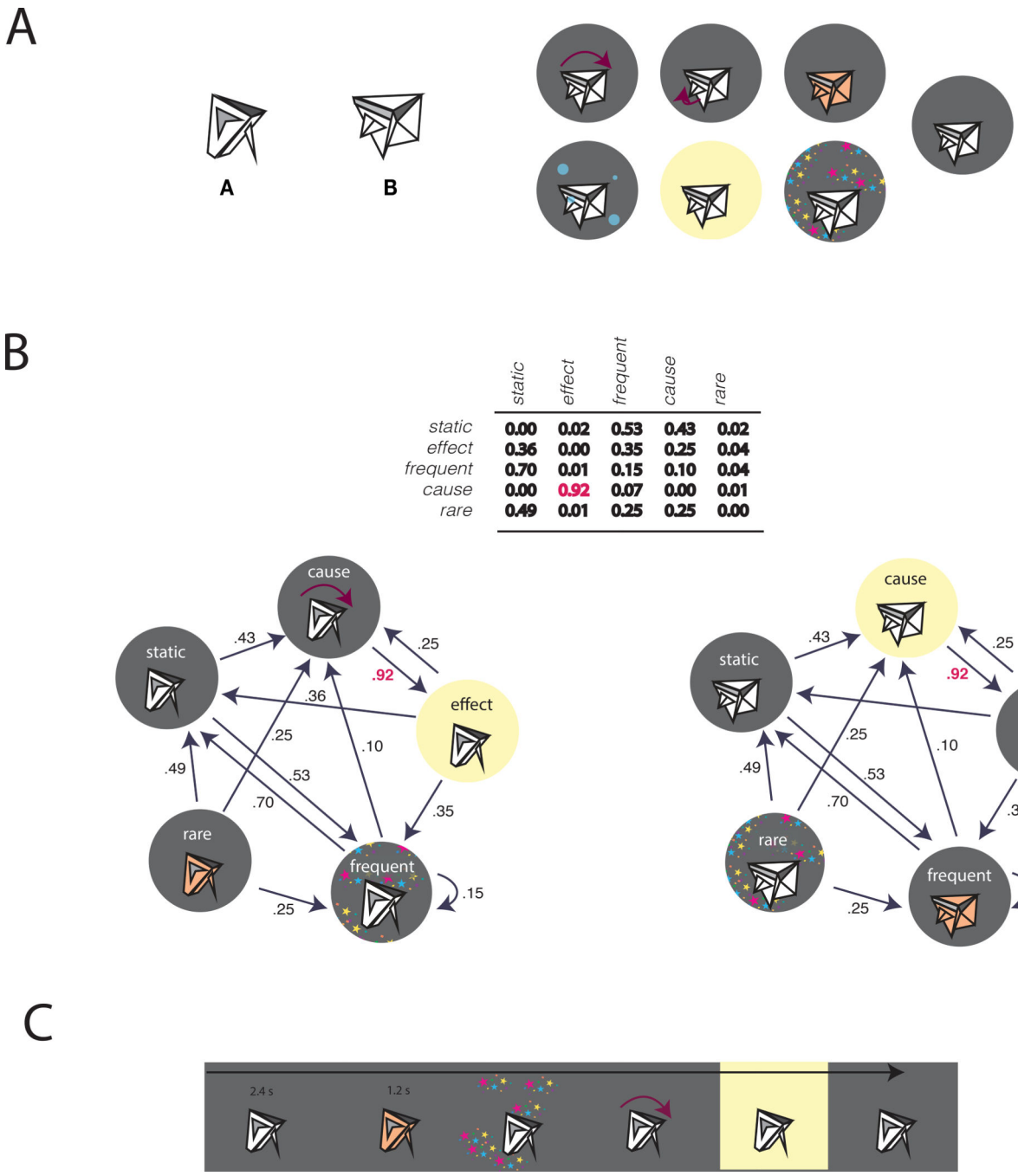


Figure 1.

A. The stimuli used in Experiment 1. Two novel objects (left; A and B) appeared in animated events (right). Three events were object-based (top) and three were ambient (bottom). Specific event assignments to abstract event types were counterbalanced across participants. **B.** Top: transition structure, defined over abstract event types, which governed the sequence of event occurrences. An example assignment is shown in graph form (below). The effect for object 1 became the cause for object 2 and vice-versa. The Frequent and Rare events also swapped between objects. **C.** Sample presentation sequence.

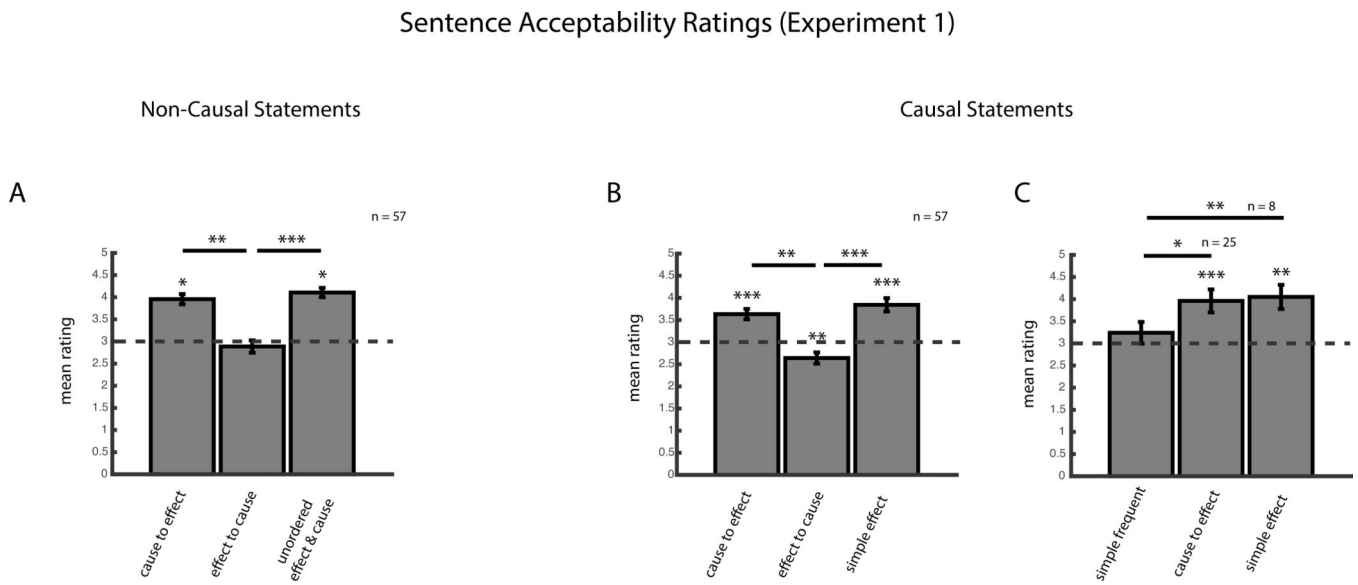


Figure 2.

Experiment 1 results; error bars reflect the standard error of the mean. * $p < .05$; ** $p < .01$; *** $p < .001$. **A.** Acceptability ratings for the non-causal statements, collapsing across objects, showing comparisons to a rating of 3 (Unsure) as well as between statement types. Statements indicating the cause preceded the effect, and that the effect and the cause were related, were accepted more strongly than the reverse relation (that the effect preceded the cause). **B.** Acceptability ratings for the causal statements, showing comparisons to Unsure as well as between statements. Statements indicating that the cause event caused the effect event were accepted, as were statements that the object caused the effect event, more than causal statements in the incorrect direction (from effect to cause). **C.** Acceptability of causal statements about the frequent and effect events, within a subgroup of participants who met the criteria for representing frequency appropriately. These participants accepted causal descriptions of the effect event more than the of the frequent (non-contingent) event.

A

B

	static	*effect*	*frequent*	*cause*	*rare*
static	**0.00**	**0.01**	**0.45**	**0.45**	**0.09**
effect	**0.38**	**0.00**	**0.25**	**0.25**	**0.12**
frequent	**0.66**	**0.04**	**0.12**	**0.10**	**0.08**
cause	**0.00**	**0.94**	**0.05**	**0.00**	**0.01**
rare	**0.4**	**0.00**	**0.30**	**0.30**	**0.00**

Figure 3.

Stimuli and Design for Experiment 2, contingency condition. **A.** Sample event assignment for training and test objects. Each participant saw either Test Object 1 or Test Object 2, which matched either Category 1 or Category 2 (respectively) on predictive relations. **B.** Obtained (actual) transition matrix governing the conditional probabilities among events for each object. Objects differed only in the event assignments to this matrix.

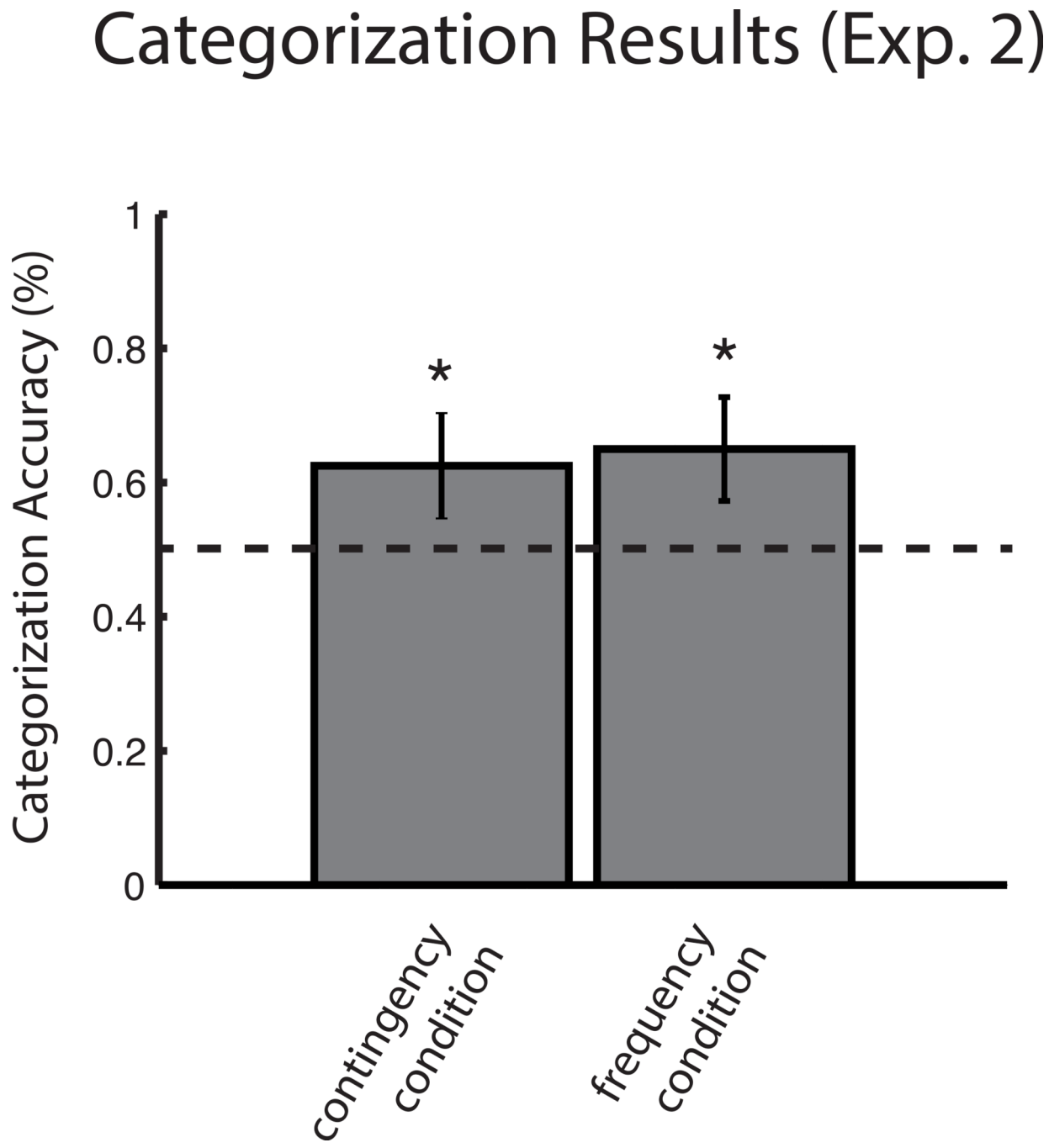
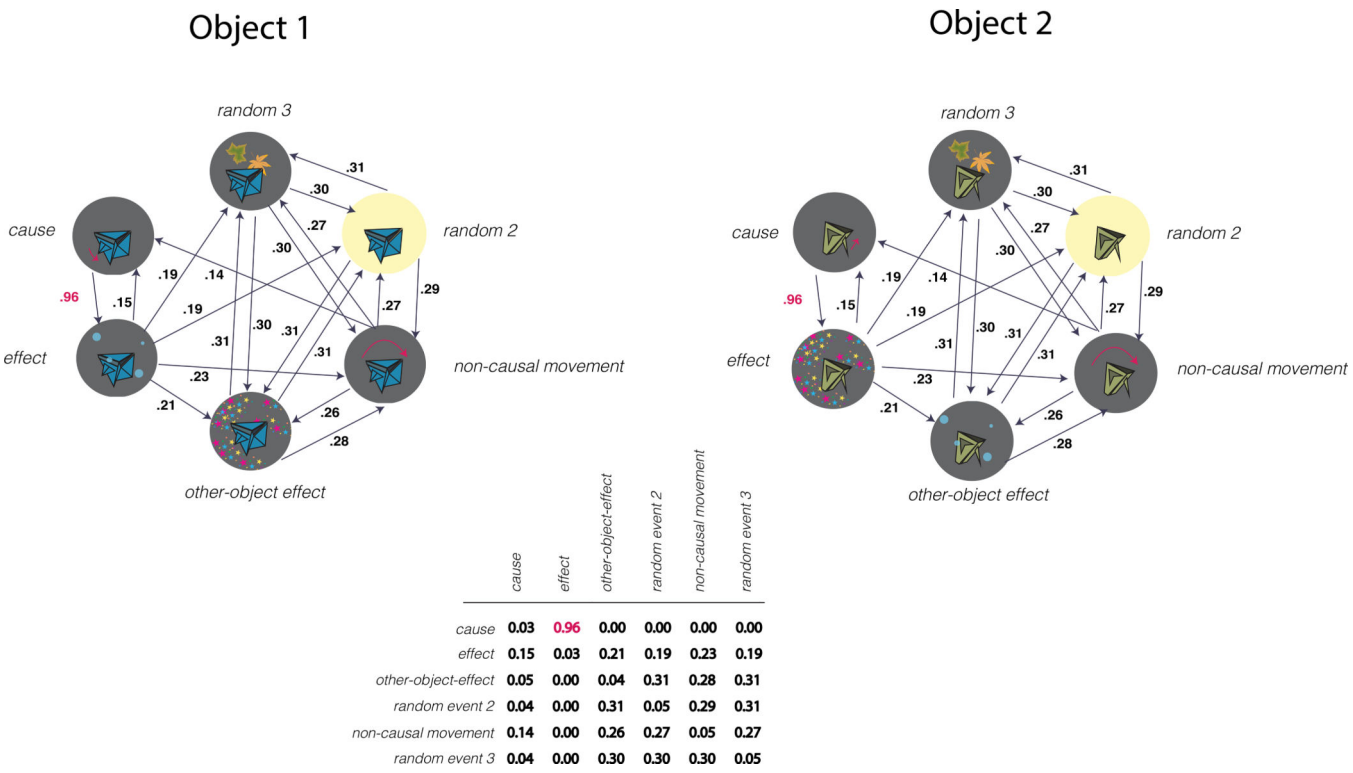


Figure 4.
Experiment 2 results; error bars reflect the standard error of the mean; * $p < .05$. Accuracy in categorizing the test object in each condition was above chance (50%) but there were no differences between conditions.



	cause	effect	other-object-effect	random event 2	non-causal movement	random event 3
cause	0.03	0.96	0.00	0.00	0.00	0.00
effect	0.15	0.03	0.21	0.19	0.23	0.19
other-object-effect	0.05	0.00	0.04	0.31	0.28	0.31
random event 2	0.04	0.00	0.31	0.05	0.29	0.31
non-causal movement	0.14	0.00	0.26	0.27	0.05	0.27
random event 3	0.04	0.00	0.30	0.30	0.30	0.05

Figure 5.

Experimental design for Experiment 3. Above: sample event assignment over two objects. Below: Empirical (obtained) transition matrix governing event transitions among abstractly defined events. Illustration of a sample event assignment over two objects. The effect and other-object-effect events swapped between objects.

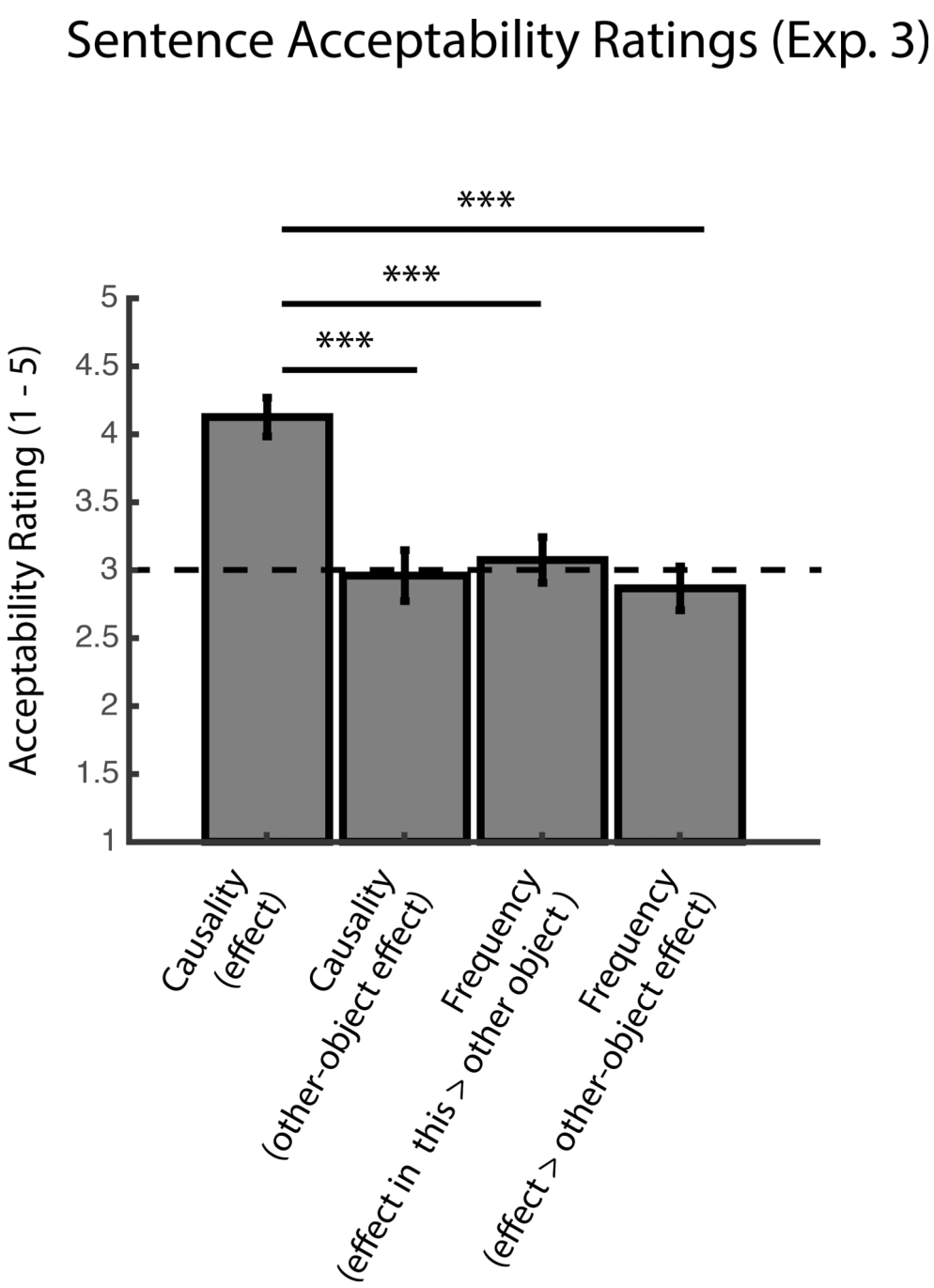


Figure 6.

Experiment 3 results; error bars reflect the standard error of the mean; *** $p < .001$. Sentence acceptability task ratings are shown for the accurate learners (n = 48). Ratings for each statement type are compared to 3 (Unsure) and to each other. Ratings for causal statements about the hypothesized effect (predictable event) were higher than for causal ratings regarding other events, and surpassed beliefs about their relative frequency relative to other events within and across objects. Results are shown for the other-object effect.