Title
Design Drives Discovery in Causal Learning

Journal
Psychological Science, 31(2)

ISSN
0956-7976

Authors
Walker, Caren M
Rett, Alexandra
Bonawitz, Elizabeth

Publication Date
2020-02-01

DOI
10.1177/0956797619898134

Peer reviewed

Design facilitates discovery in causal learning

Caren M. Walker[1*], Alexandra Rett[1], & Elizabeth Bonawitz[2]

[1]Department of Psychology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093

[2]Rutgers University – Newark, Department of Psychology, 334 Smith Hall, Newark, NJ 07102

*Correspondence concerning this article should be addressed to Caren M. Walker, Department of Psychology, University of California San Diego, McGill Hall [MC # 0109], 9500 Gilman Drive, La Jolla, CA 92093-0109. Email: carenwalker@ucsd.edu; Telephone: (858) 246-2434.

Word count: 2,000 (excluding Method and Results); Figures: 4

Abstract

We assess whether an artifact's design can facilitate recognition of abstract causal rules. In Experiment 1, 152 three-year-olds were presented with evidence consistent with a relational rule (i.e., pairs of same or different blocks activate a machine) using a machine with one of two designs. In the *standard-design* condition, pairs were placed on top; in the *relational-design* condition, blocks were placed into openings on either side. Experiment 2 assessed whether this design cue could facilitate adults' (*N*=102) inference of a distinct, conjunctive cause (i.e., that two blocks, together, activate the machine). Results of both experiments demonstrate that causal inference is sensitive to design: participants in the design conditions were more likely to infer the a priori unlikely rules. Findings suggest that reasoning failures may result from difficulty generating the relevant rules as cognitive hypotheses, but that artifact design aids causal inference, with clear implications for creating intuitive learning environments.

Keywords: causality; cognitive development; reasoning; inference

How do we infer the causal rules that govern our everyday experiences? When reasoning about causal relationships among objects and events, learners must engage in a mental search to select the most likely hypothesis, or explanation for their observations. For example, prior to activating a novel appliance, you might consider several hypothetical interventions: flipping the 'on/off' switch, depressing the reset button on the circuit interrupter, or perhaps both the switch *and* the button together activate the device. We seem to effortlessly reason about the world, guided by our prior beliefs and experiences. But causal inference—drawing conclusions about causes by observing the occurrence of effects—is almost always underdetermined; there could be infinite hypotheses to consider, and the data from our experience is not sufficient to constrain this space.

How might learners evaluate such a vast array of candidate causes? Recent work suggests that learners may operate "rationally," despite generating only a subset of the most likely causes to evaluate (Bonawitz & Griffiths, 2010). The specific set considered in the context of a particular problem may depend on a variety of factors, including their probability, their relevance, priming, and so forth (e.g., Dougherty & Hunter 2003; Klein, 1993; Schunn & Klahr, 1995; Koehler, 1994). Even young children are sensitive to input that constrains the hypotheses they consider, including information about the problem, how the data were sampled, who generated the evidence, and why (e.g., Bonawitz et al., 2011; Gergely, Bekkering & Kiraly, 2002; Walker, Lombrozo, Legare, & Gopnik, 2014). Accordingly, any input that changes a learner's prior expectations about the most likely causal relationships among events can influence the hypotheses privileged. Here, we propose an environmental cue that has not yet been examined in this context: an object's visible design. If learners use information about design to constrain the hypotheses they consider, the physical features of an artifact may change the salience of various causes, influencing learning and discovery.

Although design has not been specifically examined in the case of causal learning, there are reasons to expect that the physical environment influences causal inference. Indeed, all artifacts include some element of design, and we use these cues to infer their function. For example, if a door has no handle, we infer that we should push, because otherwise a handle would have been added. Norman (1988) includes such constraints as one of several principles of good design that impact reasoning about the intended use of objects. A large literature has explored the ways in which subtle environmental influences, or "nudges," have disproportional effects on choice (Thaler & Sunstein, 2008), impacting hygiene (Holland, Hendriks, & Aarts, 2005), energy use, (Allcott & Mullainathan, 2010), and health (Thorndike et al., 2012). Other applied research has examined whether design can change the way we *learn* in select contexts. Museum designers have used exhibit access, visibility, and affordances to encourage exploration, engagement, and understanding (e.g., adding a knob suggests an object can be moved; Allen, 2004; Shin, Park & Kim, 2014). We go beyond this applied work to consider whether similar cues can influence the salience of certain causal rules.

To illustrate how the design of an object might impact beliefs about its causal structure, consider a novel appliance that has *two* cords. While you may have a strong prior belief that appliances require only a single power source, you might instead form a hypothesis that *both* cords must be connected; otherwise, why would the design include this second cord? This example demonstrates an even more general assumption that an object's features are relevant to its function, constraining the hypotheses that are generated about its mechanism (Norman, 1988).

Existing support for this proposal can be found in the literature examining reasoning about artifacts. Specifically, both children and adults adopt a ‘design stance,’ viewing features of artifacts as reflective of that object’s kind, function, and intended use (Kelemen, 1999; Kelemen & Carey, 2007; Malt & Johnson, 1992). In addition, both preschoolers and adults privilege efficient design (selecting an object with a single feature for a single purpose over one with superfluous features; Kelemen, Seston, & Saint Georges, 2012), and map the quantity and diversity of object functions to infer the complexity of its internal mechanism (Ahl & Keil, 2016). Preschoolers also map the type of effect produced (i.e., discrete vs. continuous) to the mechanism that produced it (a binary “on/off” switch vs. a dial), providing evidence that even children relate the physical structure of an object’s mechanism to its effect (Magid, Sheskin, & Schulz, 2015).

In this prior work, learners made inferences about the design of objects, given information about functions. Here, we ask if they can perform a more challenging task—whether they will be more likely to infer an unlikely causal rule, given an object’s design. In two experiments, we test the novel prediction that manipulating the physical structure of an object will lead learners to privilege certain types of causes. We first present a case in which 3-year-olds (Exp. 1) typically fail to infer an abstract, *same-different* rule. We then present a different case in which adults (Exp. 2) typically fail to infer an unusual *conjunctive causation* rule, in which the combination of two causes produce an effect. In both paradigms, we present learners with one of two machines and assess whether design differences facilitate identification of the relevant rule.

Experiment 1

In Experiment 1, we present 3-year-olds with a relational reasoning task that they systematically fail at this age (Walker, Bridgers & Gopnik, 2016). In this task, children are introduced to a novel toy that plays music when certain pairs of blocks are placed on top. In past research, when 3-year-olds were provided with evidence that the toy is activated by the abstract relation between blocks (i.e., whether pairs are the *same* or *different*), rather than by object kinds (i.e., blocks of a particular shape/color), they failed to make the correct inference. Notably, younger children (18-30-month-olds) successfully infer *same-different* relations in the identical task, suggesting that the later decline is due to a failure to spontaneously generate relational hypotheses, not a lack of competence (Walker & Gopnik, 2014, 2017; Carstensen et al., 2019). This tendency likely results from a learned bias that temporarily privileges the role of objects over the relations between them, leading to a u-shaped developmental trajectory (in the U.S.[1]). This domain therefore provides a case study to explore whether object design influences hypothesis generation.

To assess this, we made one small modification to the original task: Rather than placing pairs of blocks on top of the machine, the blocks were inserted into transparent openings on either side. If the learner treats object design as relevant to causal inference, these features might suggest an affordance: that the machine activates due to the combination of blocks. This may raise the possibility that the *relation between blocks* is relevant.

On the other hand, 3-year-olds' well-documented failure to spontaneously privilege abstract relations suggests a strong prior for object-based hypotheses at this age. To correctly

[1] This dip in performance is *not* observed in China (see Carstensen et al., 2019), suggesting that variation in the learning environment leads to differences in relational responding.

infer the relation in this case, children must integrate information about the object's design with their prior beliefs about likely causes, taking into account why design is relevant, *and* weighing this more heavily than their prior commitments. That said, if children are indeed sensitive to the design of the learning context, this cue may inform the type of hypotheses that are privileged, reducing their tendency to prioritize object-based causes.

Method

Participants

A total of 152 3-year-olds participated in Experiment 1, with 76 children randomly assigned to either the *standard-design* (M = 41.9 months; 36 female) or *relational-design* (M = 41.6 months; 37 female) conditions (see https://osf.io/wtpmd/ to download the data from all experiments). Within each condition, half of the children observed evidence consistent with the *same* relation and half observed evidence consistent with the *different* relation. Sample size was predetermined, and satisfies a power analysis with power > .8, given an alpha of .05 and an effect size of φ = .3 (medium). This choice of sample and effect size was based on the findings from previously published data using the identical standard-design task (i.e., the causal relational reasoning task) and age group (3-year-olds in the U.S.; Carstensen et al., 2019), and is also consistent with other related findings using this method (e.g., Walker et al., 2016). An additional 9 participants were excluded due to experimenter error (3), failure to complete the study (4), parent interference (1), or interference by another child (1). Children were recruited and tested in the lab, at preschools, and at museums. All participants were tested in a quiet, private room with the experimenter.

Materials and procedure

The materials and procedure for the *standard-design* condition replicate those used by Walker et al. (2016, Exp. 1; see Fig. 1). Children were seated at a table across from the experimenter. The experimenter began by placing an opaque cardboard box on the table, saying "This is my toy! Sometimes when I put things on top, the toy will play music, and other times it does not. Should we try some and see how it works?" As in previous research, the machine appeared to activate and play a novel melody in response to certain combinations of blocks. (The machine actually activates a wireless doorbell via hidden button.) The experimenter was blind to the hypotheses of the study.

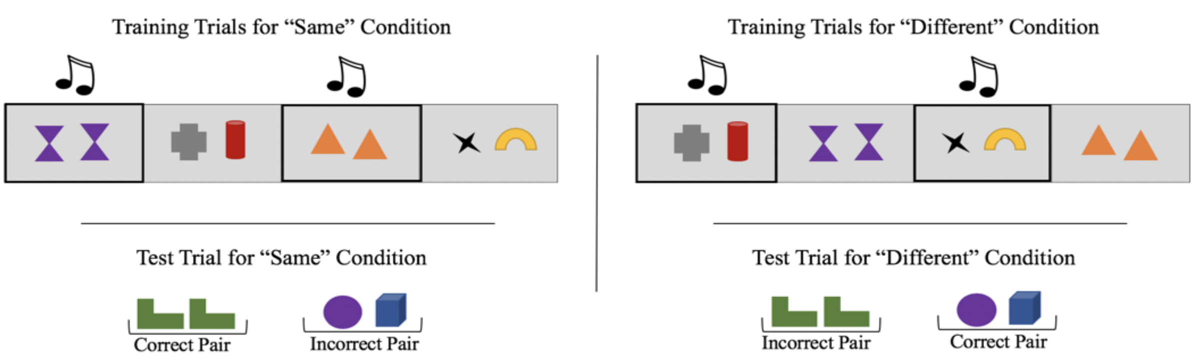

Figure 1. Schematic illustration of evidence presented during training and test trials in the *standard-design* condition. Identical pairs and outcomes were presented in the *relational-design* condition, using the relationally designed machine (see Fig. 2).

Pairs of *same* (2) and *different* (2) painted wooden blocks were used during the training trials. After introducing the machine, the experimenter produced two blocks in either the *same* or *different* relation (depending upon the condition), and said, "Let's try!," putting both blocks on top of the machine, simultaneously. The machine played music and the experimenter said, "Music! My toy played music!" The experimenter then picked up the blocks and simultaneously set them back on the machine, which again played music, saying "Music! These ones made my toy play music!" She then repeated this procedure with a new pair of blocks in the opposite relation. The new pair did not make the machine play music, and the experimenter responded to the first try with, "No music! Do you hear anything? I don't hear anything," and after the second try, said "No music. These ones did not make my toy play music." This pattern was repeated with two additional pairs of blocks, one in each relation. The experimenter always began with a causal pair (identical blocks in the *same* condition and blocks of unique colors and shapes in the *different* condition), and then alternated inert, causal, inert, using novel blocks in each new pair, and randomizing the specific blocks between participants.

After the four training trials, the experimenter said "Now that you've seen how my toy works, I need your help finding the things that will make it play music. I have two choices for you." The experimenter presented the child with two new pairs composed of novel blocks, one "same" pair and one "different" pair. Each pair was presented on a plastic tray, which the experimenter held up, saying, "I have these, and I have these (directing the child's attention to each pair). Only one of these trays has things that will make my toy play music. Can you point to the tray that has the things that will make it play?" The trays were then placed out of the child's reach, on either side of the machine, with each pair set an equal distance from the child. The order and side of presentation of the correct pair counterbalanced between participants. The experimenter recorded the child's first point or reach, scoring the response as correct (1) if the child chose the test pair (*same* or *different*) that corresponded to her training, and incorrect (0) for the opposite pair.

The materials and procedures for the *relational-design* condition were identical to those in the *standard-design* condition with one critical difference: The design of the machine was modified to include two transparent openings located on either side (Fig 2). The openings were constructed using clear, 2”x2” hard plastic boxes. When children observed each of the training trials described above, pairs of blocks were inserted simultaneously into the two openings (one block on either side), rather than placed on top of the machine. To do so, the experimenter picked up one block in each hand, positioned her hands on either side of the machine, and then placed the blocks inside the two openings at the same time, causing the toy to activate and play music. This was the only difference between the two conditions.

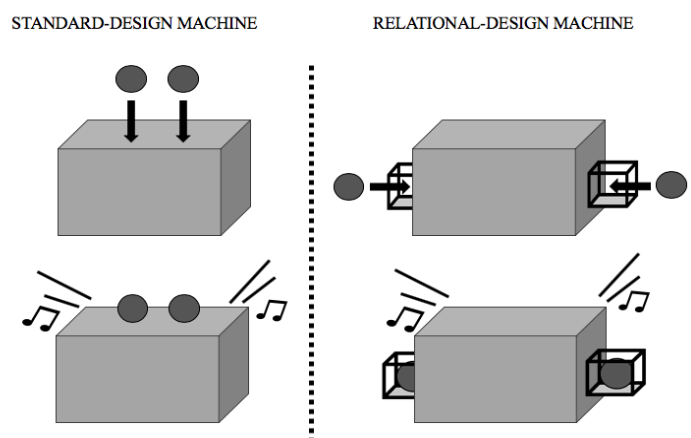

Figure 2. The two different designs of the machine used in Experiments 1 and 2.

Results

The number of correct responses given in each condition is reported in Table 1. Replicating previous work (Walker et al., 2016), 3-year-old children in the *standard-design* condition were equally poor for both *same* (53%, 95% CI [37%, 69%]) and *different* (40%, 95% CI [24%, 55%]) trials, $p = .357$ (two-tailed, Fisher’s exact, OR = .587, 95% CI [.236, 1.46]); overall, children in this condition responded at chance (46%, 95% CI [35%, 58%]), $p = .567$

(two-tailed, exact binomial, OR = .852). Three-year-olds in the *relational-design* condition also showed no difference between *same* and *different* trials (p > .999, Fisher's exact, OR = 1, 95% CI [.388, 2.58]); however, critically these children succeeded in selecting the test pair that was consistent with their training (66%, 95% CI [54%, 76%]), p = .008 (exact binomial OR = 1.94). Comparing performance across conditions, children in the *relational-design* condition significantly outperformed those in the *standard-design* condition (p = .022, Fisher's exact, OR = .446, 95% CI [.219, .897]) in inferring the correct relations.

Table 1. *Frequency of Responses in Experiment 1.*

	Relational-Design Condition			***Standard-Design*** Condition		
	Same	*Different*	**Total**	*Same*	*Different*	**Total**
Correct Response	25	25	**50**	20	15	**35**
Incorrect Response	13	13	**26**	18	23	**41**

Experiment 2

Given this early sensitivity to design, we next examined whether the effects of this same manipulation would extend to impact more entrenched beliefs in adults. Adults perform at ceiling on the standard *same-different* task used in Experiment 1.[2] We therefore selected a different causal reasoning task that adults are known to fail. Specifically, adults fail to infer the conjunctive relation (i.e., that two objects, together, cause a machine to activate), and instead favor the more typical disjunctive relation (i.e., that only one object is needed), despite evidence for the conjunctive rule (Lucas & Griffiths, 2010). Interestingly, preschoolers—who have weaker prior commitments—are *better* able to infer conjunctive rules from the evidence, outperforming

[2] In a pilot study with 38 adults on Mechanical Turk, 100% correctly selected the correct test pair (*same* or *different*) using the standard design.

adults on this task (Lucas, Bridgers, Griffiths, & Gopnik, 2014, see also Gopnik, Griffiths, & Lucas, 2015; Gopnik et al., 2017). These prior results suggest that evaluation of evidence alone may be insufficient for discovery of this relation. In order to facilitate this inference, design must do more than provide additional data—it must help the learner to generate a hypothesis they may not have considered otherwise.

If design is sufficient to influence the hypotheses adults entertain, we predict that they *will* infer this rule in the *relational-design* condition. Of course, an alternative explanation for our predicted result is that, regardless of the evidence, seeing two openings encourages adults to test two blocks. To control for this, we also include a condition in which participants observe evidence that statistically favors the disjunctive rule, but using the relationally-designed machine. Adults in this relational-design control condition should infer that only one object is needed, thus allowing us to rule out the alternative explanation.

Method

Participants

A total of 102 adults (M_{age} = 21.3, *SD* = 2.3, 77 females) participated in Experiment 2, with 34 individuals randomly assigned to each of three conditions: *standard-design* (conjunctive rule), *relational-design* (conjunctive rule), and a control condition—a *relational-design* (disjunctive rule). The target sample size was predetermined, based on effect sizes from prior work (condition N's from Lucas et al. 2014 Exp. 1 =38; Exp. 2 = 28). Adults were recruited from a pool of undergraduate psychology majors at a public university in California and were given course credit in exchange for their participation. An additional 7 participants were excluded due to experimental error (6) or failure to attend to the experimental procedures (1).

Materials and Procedure

Except for our introduction of the relationally designed machine, the materials and procedure for all conditions were based on Lucas et al. (2014). Materials included nine uniquely-shaped purple wooden blocks (triangle, cylinder, cube, sphere, star, pentagon, heart, T-shape, semi-circle). The blocks were broken into sets of three (e.g., triangle, cylinder, cube), with two sets used during training trials and one set used during the ambiguous test. As in Experiment 1, the *standard-design* condition used the standard machine, in which blocks were placed on top to activate it, and the two *relational-design* conditions used the relational machine, in which blocks were inserted into the transparent openings on the sides to activate it (Figure 2). One additional *standard-design* machine was used for the familiarization phase that was a different color (pink) and shape (cylinder) than those used in Experiment 1. All machines were activated by a wireless doorbell controlled by the experimenter. The experimenter was blind to the hypotheses of the study.

Familiarization task. We began with a warm-up task that included a backward blocking paradigm (Gopnik & Sobel, 2000) to introduce participants to a causal procedure in which object labels could be ambiguous. Participants in both conditions were shown a machine (standard design) and were told, “We are going to try to figure out which things are *wugs*. You can’t tell that something is a *wug* just by looking at it, but only *wugs* have *wugness* inside of them. Luckily, I brought my *wugness* machine. The way that this machine works is it turns on and plays music when there is *wugness* to be detected.” Participants were then shown two objects, placed on top of the machine simultaneously. The machine activated, playing music. Both objects were then removed and one of them was placed back on top of the machine. Again, the machine activated. This provided evidence that one object was a wug, which also served to explain away the initial activation, leaving the “wugness” of the second object ambiguous (e.g.

see Griffiths et al, 2011). Participants were then asked whether each object was a *wug*. Afterwards, the familiarization machine and all materials were removed, and participants proceeded on to the main experimental procedure.

Training events. Participants were then introduced to a new machine (*standard-design* or *relational-design*, depending on condition) differing in shape and color from the familiarization machine. The experimenter explained, “Now, we are going to try to figure out which things are *blickets*. You can’t tell that something is a *blicket* just by looking at it, but only *blickets* have *blicketness* inside them. I brought my *blicketness* machine. The way that the *blicketness* machine works is it turns on and plays music when there is *blicketness* to be detected.”

Prior to observing objects on the machine, participants were shown a single object and asked whether or not it was a *blicket*. Specifically, they were asked, “First, before we try any objects in my machine, do you think this is a blicket?” This allowed us to empirically estimate how likely participants were to judge that a particular object was a *blicket* when no evidence was available, and ensure that their subsequent responses were significantly different from baseline probability that objects were *blickets*.

Participants were then shown three objects, and observed a set of training events in which the experimenter placed objects either alone or in pairs on (or in) the toy. In the *standard-design* condition, objects were placed on top. In the *relational-design* conditions, objects were instead inserted into the openings on either side, as in Experiment 1[3]. In some cases, the objects would cause the machine to activate (play music).

---

[3] When a training event included a single object, the experimenter placed the object in one side, which was counterbalanced between sets.

The training events followed directly from the conjunctive or disjunctive training used in Lucas et al. (2014); see Figure 3. Depending on condition, participants either observed evidence that two blocks together caused the machine to activate (conjunctive rule), or that only a single block was needed (disjunctive rule). Identities of the individual objects that activated the toy and the order of the sets were counterbalanced.

Ambiguous test events. For the ambiguous events, participants were introduced to three new blocks (D, E, and F), and observed the events illustrated in Figure 3 (identical to Lucas et al., 2014)[4]. Critically, if viewed without the initial training, the events were designed to be ambiguous between a disjunctive (F) or a conjunctive (D & F) causal rule. However, if participants learned the form of the relationship required to activate the blicketness machine during the training trials, they could then apply this knowledge to disambiguate the evidence and draw the appropriate inference about the causal status of each of the blocks. That is, individuals in the conjunctive rule conditions should apply the label to *both* D and F, but not E, because both would be required to have blicketness inside to jointly activate the machine. In contrast, only object F should be labeled a blicket in the disjunctive (relational-control) condition, because if only one block is necessary to activate the machine, the evidence of the initial trials rules out D as a candidate cause. In all conditions, participants could explain-away the E block's association with activation (on the sixth ambiguous event), by virtue of causally identifying the other blocks from the other trials. Finally, after observing the ambiguous events, participants were asked to

[4] When the triplet (D & E & F) was demonstrated with the *relational-design*, the experimenter placed two objects in one opening and one in the other, with object placement counterbalanced between participants.

infer whether or not each of the new objects was a blicket. This served as our critical test measure.

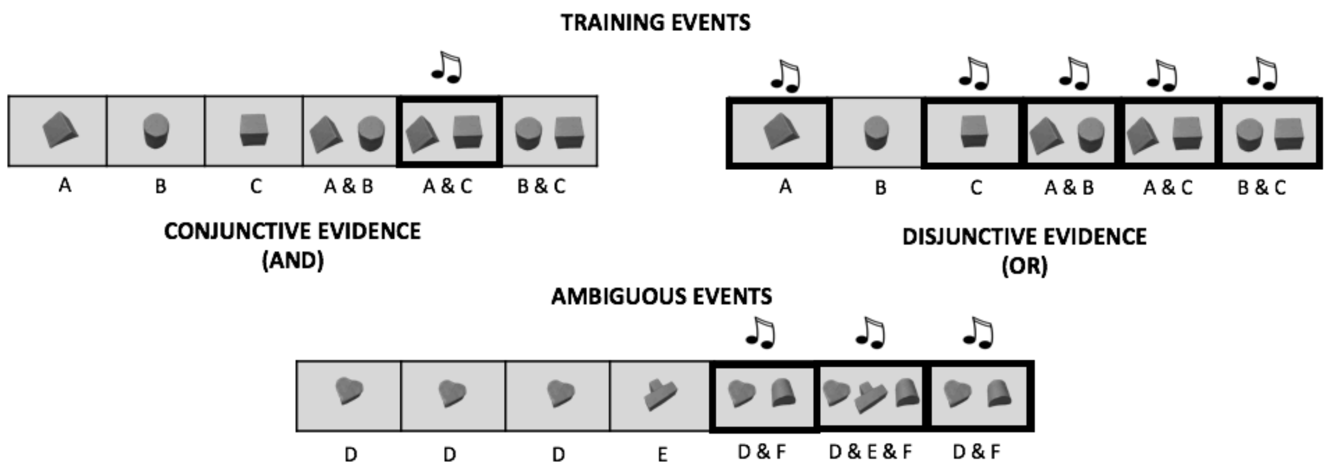

Figure 3. Conjunctive and disjunctive training events and the ambiguous events in Experiment 2. The evidence in the ambiguous event trials is insufficient to infer whether both D&F or just F are blickets, without information about whether the machine operates via conjunctive or disjunctive rules. The training events (between subjects) disambiguates these trials, but requires that the participants consider the correct rule initially from this evidence.

Results

Familiarization and Baseline Questions

The majority of participants made the backward blocking inference during familiarization (89%, 95% CI [82%, 95%]), $p < .001$ (exact binomial, OR = 8.09), with no difference across conditions ($p = .616$, *ns,* Fisher Exact), suggesting that participants had no trouble reasoning causally about the machines and are able to use information to screen-off ambiguous blocks (as is required for block "E" in the primary experimental task). Second, in the primary experimental task, only 24% of participants inferred that the baseline object was a blicket, indicating that adults assumed blickets to be relatively rare when compared to chance (50%), (p<.001, exact binomial, 95% CI [16%, 33%], OR = .316), with no differences across conditions (p = .956, *ns,* Fisher Exact). This baseline measure also served as a comparison point, for which we could assess participant responses in the test condition.

Table 2. *Frequency of Labeling Each Object a "Blicket" in Experiment 2*

Condition	Standard-Design, Conjunctive Rule			Relational-Design, Conjunctive Rule			Relational-Design, Disjunctive Rule		
Object	D	E	F	D	E	F	D	E	F
Frequency	6	2	27	15	4	30	0	3	34

Effects of Design Manipulation on Judgements

Results appear in Figure 4, with the exact number of adults who labeled each object a "blicket" reported in Table 2. There was no difference among the three conditions in their tendency to label E a blicket (*standard-design* [conjunctive rule]: M = 6%, 95% CI [0%, 14%]; *relational-design* [conjunctive rule]: M = 12%, 95% CI [.10%, 23%]; *relational-design* [disjunctive rule]: M = 9%, 95% CI [0%, 18%]), p = .771 (Fisher Exact). Also, as expected, participants in both the *standard-design* (conjunctive rule) and *relational-design* (conjunctive rule) conditions labeled F a blicket more often than expected by chance[5] (*standard-design* [conjunctive rule]: M = 79%, p < .001, 95% CI [66%, 93%], OR = 11.9; *relational-design*

[5] Chance was determined using the baseline measure (24%).

[conjunctive rule]: *M* = 88%, $p < .001$, 95% CI [77%, 99%], OR = 23.2, with no difference between groups[6], $p = .512$ (OR = 1.94, 95% CI [.512, 7.38]).

Critically however, there was a significant difference in the tendency to label D a blicket across conditions, $p < .001$ (Fisher Exact). Planned comparisons revealed that although participants in the *standard-design* (conjunctive rule) condition rarely labeled D a blicket (*M* = 18%, 95% CI [5%, 30%]), which was not different from chance (binomial, $p = .546$, OR = .695), those in the *relational-design* (conjunctive rule) condition *did* label D a blicket (*M* = 44%, 95% CI [27%, 61%]) more often than expected by chance (binomial $p = .014$, OR = 2.49), with a significant difference between these groups, $p = .034$ (Fishers exact, OR = 3.68, 95% CI [1.21, 11.2]). In line with our predictions, these findings suggest that the design of the relational machine served to support adult inferences of the conjunctive form. Replicating what would be expected for disjunctive rule inference, participants in the *relational-design* (disjunctive rule) condition *never* labeled D a blicket (*M* = 0, 95% CI [0%, 0%], $p < .001$, binomial), which is significantly different from the *relational-design* (conjunctive rule) condition reported above, $p < .001$ (Fisher exact), Phi = -.54. This demonstrates that the relationally designed machine did not simply increase adults' tendency to label more than one object a blicket, and therefore cannot explain the improved performance in the *relational-design* (conjunctive rule) condition.

[6] There was a significant difference in the tendency to label F a blicket across all three conditions, $p = .014$, this was driven by the fact that 100% of adults in the *relational-design* (disjunctive rule) condition indicated that F was a blicket.

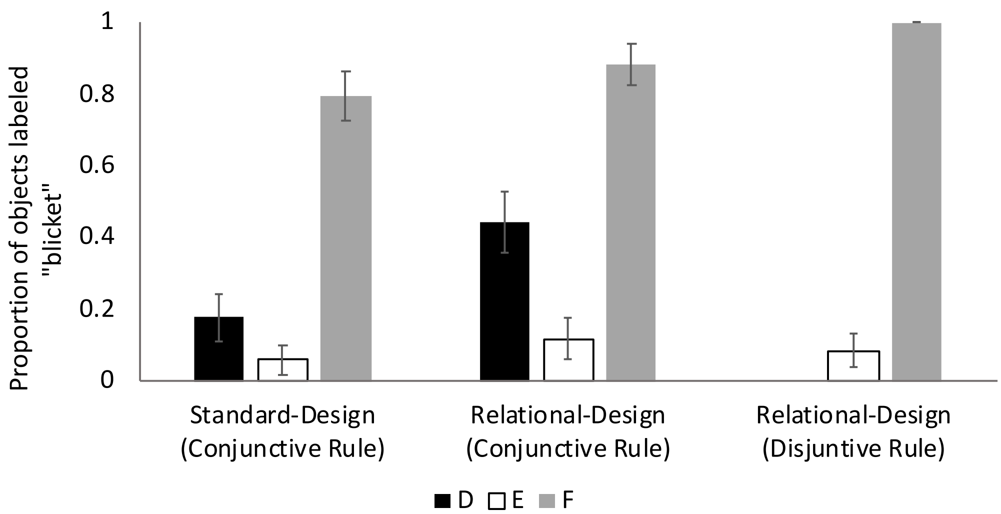

Figure 4. Proportion of objects (D, E, F) labeled "blicket" across conditions (*standard-design* conjunctive rule, *relational-design* conjunctive rule, and *relational-design* disjunctive rule). Error bars represent +/-1 SEM.

General Discussion

Results demonstrate that both children and adults are sensitive to the design of the learning context when reasoning about causal relationships. Although 3-year-olds in the *standard-design* condition failed to recognize the relational hypothesis (replicating prior work), increasing the salience of this hypothesis through the application of a subtle design cue increased their tendency to engage in relational reasoning. In addition to providing evidence for the role of design in constraining causal inference, these data provide additional support for the proposal that children's previous reasoning failures do not result from a lack of competence (e.g., Walker et al., 2016). These results are particularly striking given children's strong prior for causal hypotheses based on individual objects. In order to use the design of the learning context to override it, these children had to make a sophisticated inference: They must have noticed this cue, inferred that an object's design is relevant for its function, and weighed this information more heavily than their prior commitments.

Experiment 2 shows that adults also integrate design features to improve causal inference from evidence. Moreover, results suggest that adults' previous failure to infer the conjunctive rule may be due to their failure to consider it as a possible hypothesis during their evaluation of evidence. This is consistent with prior accounts suggesting that evaluation and generation of hypotheses may represent two distinct processes underlying inductive inference (Kuhn, 1989; Klahr, Fay, & Dunbar, 1993). These experiments provide initial evidence that the design of objects may influence hypothesis generation, changing learning outcomes even for the more entrenched beliefs and biases characteristic of adults. Together, these findings suggest that relatively minor elements of design can change the distribution of a learner's prior expectations, constrain the type of hypotheses generated, and influence learning.

Given that design serves as a constraint for both young children and adults, it is likely a useful cue across the age range. However, future work might examine the nature of these effects in older children and adolescents, and explore a wider range of learning contexts and knowledge domains. Relatedly, future work might consider whether similar effects are found in other cultural contexts – particularly those with fewer artifacts. It is possible that the sensitivity to these cues is itself learned through exposure to complex tools and designed environments. Relatedly, there are a variety of open questions regarding how the particular design modifications used here influenced reasoning. One possibility is that the relational design simply served to disrupt the learner's initial intuitions about the likely causal mechanism, leading them to consider alternatives more broadly. If so, this may have made it more likely for participants to discover the relational hypothesis, albeit indirectly.

Finally, these results have clear implications for the design of learning environments. Our findings dovetail with literature in education that points to the importance of “mise en place” or setting the stage for learning (Weisberg, Hirsh-Pasek, Golinkoff, & McCandliss, 2014). As demonstrated here, both young children and adults are sensitive to design when engaged in causal inference. These findings therefore open up new avenues for work examining how learning environments can be used to constrain reasoning, support belief revision, and guide discovery.

Acknowledgments

We are grateful to Mike Frank for his early contributions to the method. We thank Alicia Lunardhi, Valentina Mancuso, Nicky J. Sullivan, and Emily To for their efforts towards data collection. We also thank the Fleet Science Center, the Birch Aquarium at Scripps Institute of Oceanography, and the New Children’s Museum, as well as participating preschools and families. This research was funded by a Hellman’s Fellowship, awarded to C. Walker.

Author Contributions

C.M.W. and E.B. designed the studies, A.R. and C.W. performed the research, A.R. and C.M.W. analyzed the data, and C.M.W., A.R., and E.B. wrote the paper.

References

Ahl, R. E., & Keil, F. C. (2016). Diverse effects, complex causes: children use information about machines' functional diversity to infer internal complexity. *Child Development, 88*(3), 828-845.

Allcott, H., & Mullainathan, S. (2010). Behavior and energy policy. *Science, 327*(5970), 1204-1205.

Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education, 88*(S1), S17-S33.

Bonawitz, E., & Griffiths, T. (2010). Deconfounding hypothesis generation and evaluation in Bayesian models. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2260 –2265). Austin, TX: Cognitive Science Society.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322-330.

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development,83*(4), 1416-1428.

Carstensen, A., Zhang, J., Heyman, G.D., Fu, G., Lee, K., & Walker, C.M. (2019). Context shapes early diversity in abstract thought. *Proceedings of the National Academy of Sciences.* To appear June 24th, 2019. www.pnas.org/cgi/doi/10.1073/pnas.1818365116

Dougherty, M. R., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113*(3), 263-282.

Gergely, G., Bekkering, H., & Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature, 415,* 755.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*(5), 1205-1222.

Gopnik, A., Griffiths, T.L., & Lucas, C.G. (2015). When younger learners can be better (or at least more open-minded) than older ones. Current Directions in Psychological Science, 24(2), 87-92.

Gopnik, A., O'Grady, S., Lucas, C.G., Griffiths, T.L., Wente, A., Bridgers, S., Aboody, R., Fung, H., & Dahl, R.E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. Proceedings of the National Academy of Sciences, 14(30), 7892-7899.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. Cognitive Science, 35(8), 1407-1455.

Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells like clean spirit: Nonconscious effects of scent on cognition and behavior. *Psychological Science, 16*(9), 689-693.

Kelemen, D. (1999). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Sciences, 3*(12), 461-468.

Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. *Creations of the mind: Theories of artifacts and their representation*, 212-230.

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

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*(1), 111-146.

Klein, G.A. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In G. A. Klein, J. Orasanu, R. Calderwood, and C. Zsambok (Eds.), *Decision Making in Action: Models and Methods.* Norwood, NJ: Ablex Publishing Corp., 138–147.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition,20*(2), 461-469.

Kuhn, D. (1989). Children and adults as intuitive scientists. Psychological review, 96(4), 674.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition, 131*(2), 284-299.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science, 34*(1), 113-147.

Magid, R. W., Sheskin, M., & Schulz, L. E. (2015). Imagination and the generation of new ideas. *Cognitive Development, 34*, 99-110.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? Journal of memory and language, 31(2), 195-217.

Norman, D. (1988). *The design of everyday things*. New York: Basic Books.

Schunn, C. D., & Klahr, D. (1995). A 4-space model of scientific discovery. In J.D Moore & J.F. Lehman (Eds.), *Proceeding of the seventeenth annual conference of the Cognitive Science Society* (pp. 106-111). Mahwah, NJ: Erlbaum.

Shin, H., Park, E. J., & Kim, C. J. (2014). Learning affordances: Understanding visitors' learning in science museum environment. In *Topics and Trends in Current Science Education* (pp. 307-320). Springer, Dordrecht.

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

Thaler, R.H., Sunstein, C. R. (2008) *Nudge: improving decisions about health, wealth, and happiness* New Haven: Yale University Press,

Thorndike, A. N., Sonnenberg, L., Riis, J., Barraclough, S., & Levy, D. E. (2012). A 2-phase labeling and choice architecture intervention to improve healthy food and beverage choices. *American Journal of Public Health, 102*(3), 527-533.

Walker, C. M., Bridgers, S., & Gopnik, A. (2016). The early emergence and puzzling decline of relational reasoning: Effects of knowledge and search on inferring abstract concepts. *Cognition, 156,* 30-40.

Walker, C.M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, 25(1): 161-169.

Walker, C.M., & Gopnik, A. (2017). Discriminating relational and perceptual judgments: Evidence from human toddlers. *Cognition, 166,* 23-37.

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition, 133*(2), 343-357.

Weisberg, D.S., K. Hirsh-Pasek, R.M. Golinkoff, & B.D. McCandliss. 2014. "Mise En Place: Setting the stage for thought and action." *Trends in Cognitive Sciences* 18 (6): 276–78.