Using efficiency to infer the quality of machines

Claudia G. Sehl (claudia.sehl@uwaterloo.ca); **Ori Friedman** (friedman@uwaterloo.ca);
& **Stephanie Denison** (stephanie.denison@uwaterloo.ca)

Psychology Department, University of Waterloo, 200 University Ave. W.
Waterloo, Ontario, Canada N2L 3G1

Abstract

When assessing the quality of a machine, people might consider the machine's outputs—how well it serves its function. Alternatively, people might also consider the efficiency of the machine. We investigated this possibility in two experiments (N = 392). In each experiment, participants saw pairs of machines, one with simple inside parts and one with more complex inside parts. Machines either had the same output or unknown outputs, and people judged which of the two machines was better. When the machines had the same output, participants in both experiments judged that machines with simpler inside parts were better than ones with more complex insides. However, when machines' functions were unknown, people predominantly judged that machines with complex insides were better. Together, our work shows that people consider both parts and functions of machines when inferring quality.

Keywords: simplicity, efficiency, causal process, judgment, cognition

Introduction

Consider two cars that perform identically. They have identical safety ratings, mileage, braking systems, speed, transmissions, reliability, comfort, and so on. You might conclude that the two cars are just as good as one another. But suppose you look under their hoods and find that one car has complex internal workings, whereas the other has just a few parts. Now you might conclude that the car with fewer parts is better since it delivers the same features with fewer parts. So, when judging which of two mechanisms is better, we may consider efficiency—maximal output with minimal input.

In this paper, we ask how people consider the fit between functions and parts when assessing the quality of machines, or how good they are. Investigating this question will extend our knowledge of how people conceive of machines. Prior work shows that people understand that inside parts of machines are causally related to their functions. People infer that if an object's internal properties were transferred to another object, the object's function would also be transferred. Additionally, people make bidirectional inferences about similarity in objects' parts and functions: objects with identical internal parts are inferred to have the same function, and objects with identical functions are inferred to have the same parts (Sobel et al., 2007). The complexity of an object's function is also indicative of the complexity of its inside parts. For instance, when shown a machine with a simple function (makes cupcakes) and another with a complex function (makes cupcakes and soups), people expect that the more complex machine has more complex parts (Ahl & Keil, 2017; also see Erb et al., 2013). Objects are inferred to be complex if they have a greater number of parts, diverse parts, and more connections between parts, suggesting that people consider many cues for complexity (Ahl et al., 2020). Most of these findings have been demonstrated in both children and adults, though we limited our investigation to adults.

How do people consider the fit between functions and parts when assessing quality? As we considered with the cars, they might think that machines with simpler insides are better than those with complex ones. Machines with simpler insides might be judged as better because they are seen as more *efficient*. This proposal relates to the idea that people infer that individuals are more competent if they perform tasks more efficiently. For example, in one recent study, four- to six-year-olds and adults watched two agents each build a tower of blocks (Leonard et al., 2019). Each agent built a tower that was depicted in a picture located near them: one made from ten cubes and the other made from two long blocks. Crucially, the agents finished building their towers in the same amount of time, and tower height was identical. Adults reported that the agent who built the ten-block tower was better than the agent who built the two-block tower, suggesting that people make evaluations based on the efficiency of processes.

Previous research has also found that people often prefer fewer components when judging causal mechanisms and explanations. For instance, if a woman was experiencing exhaustion, weight gain, and upset stomach, people prefer the explanation that she is pregnant, over an alternative explanation that she has mononucleosis, a virus, and lacks exercise (Lombrozo, 2007; Read & Marcus-Newhall, 1993). So, people prefer fewer and simpler causes when determining the origins of complex outcomes.

Another possibility is that people make the reverse judgment. More internal parts, or more complex internal parts, could be a marker of greater quality. This possibility is broadly consistent with previous work indicating that people associate complex inner parts with complex functions (e.g., Ahl & Keil, 2017; Ahl et al., 2020). One final possibility is that people might primarily focus on the output of a machine, irrespective of its parts. They might care about the functions it performs and how well it performs them, and not about whether the underlying mechanism is complex or simple. In the cars example, people might consider the cars to be the

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

same quality since they assess quality depending on the cars' function and not on its parts. However, people often consider an object's inside parts as indicative of its category (Gelman & Wellman, 1991; Keil, 1989) and capacity (Gelman & Kremer, 1991), so it may be unlikely for people to not consider inside parts when making these evaluative judgments.

Taken together, if people are sensitive to inside parts when judging between two machines with identical functions, they may consider the machine with simpler insides to be better than the one with complex insides, as it has a more efficient design. However, without knowing the machines have identical functions, people may instead judge that the machine with complex insides is better, as they may assume it has more complex functions (Ahl & Keil, 2017).

We examined this question across two experiments. In both, participants judged which of two machines was better—one with simple inside parts or one with complex inside parts. We manipulated complexity in two ways: number of circuits versus number of parts on a single circuit. We had no a priori predictions about how these different indicators of complexity would impact judgments and our analyses of this variable were purely exploratory.

Experiment 1

Materials and data for both experiments are available online here.

Methods

Participants We tested 196 adults (M_{age} = 37.56, range = 19-73, 69 female; 125 male; 2 other). Data were excluded from an additional 34 participants who failed at least one of the three post-test comprehension checks. Participants in both experiments were recruited from CloudResearch (http://www.cloudresearch.com/). They were located throughout the United States and had a HIT rate above 95%.

Procedure Participants completed four trials, where they saw pairs of machines (inspired by Ahl & Keil, 2017; see Figure 1a). In each pair, there was a machine with only a few components ('simple') and one with many components ('complex'). For two trials, machines differed in complexity based on the number of circuits ('number' trials). In the other two trials, machines differed by the number of parts and wiring in each circuit ('wiring' trials). The distinction between these trials was not described to participants in each experiment.

The function of the machines was manipulated between-subjects. In the Same Output condition, participants were told that the machines made identical objects, like cupcakes, beach balls, hamburgers, and teddy bears. In the Unknown Output condition, participants were only told that the machines made things, and were not shown the machines' output. Participants rated which machine was better on a scale from 1 (definitely this [left] machine) to 7 (definitely this [right] machine).

The four trials were randomized for each participant. The location of the simple machine in each pair was counterbalanced in the number and wiring trials. After the test trials, participants completed three comprehension checks.

The first comprehension question asked participants to identify the judgments they made: which machine was better, was broken, is the smartest, or is familiar. The second question asked which of four machines was not shown. The third question showed a sample circuit and asked whether it was part of a machine or what the machine made. The order of the options was randomized. Participants were excluded if they failed to answer one of the test questions, or incorrectly answered at least one post-test comprehension check.

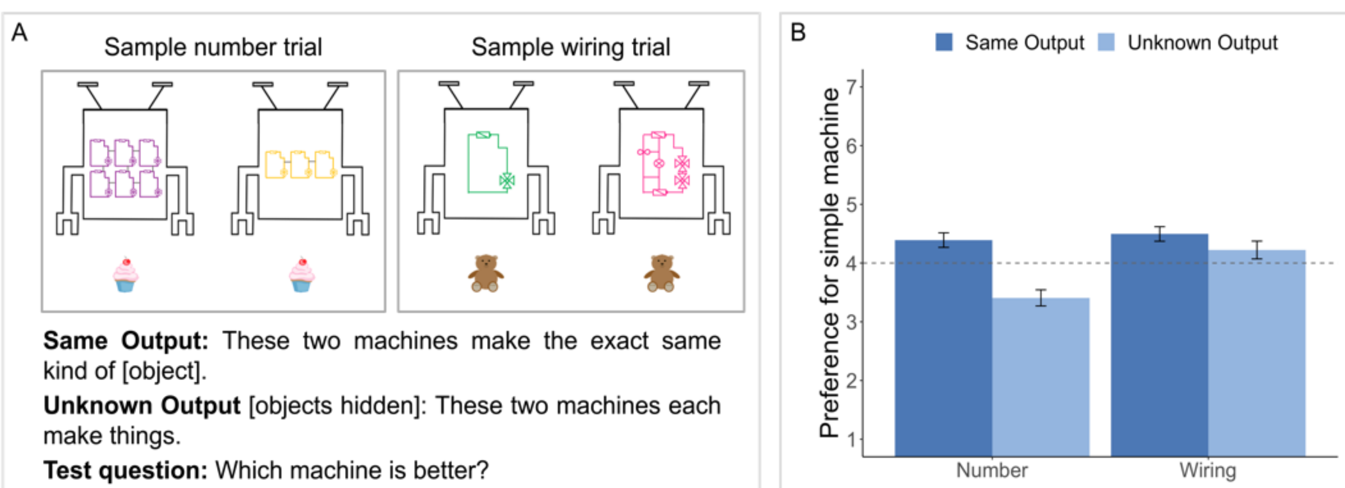

Figure 1. A) Number and wiring trials from Experiment 1 in the Same Function Condition. B) Results from Experiment 1. Error bars represent ± 1 standard error.

Results

In both experiments, responses were coded from 1 (definitely complex machine) to 7 (definitely simple machine). Data were analyzed using linear generalized estimating equation models (GEE; independent correlation matrix), with condition and trial type as predictors (see Figure 1b). There was a main effect of condition, as participants were more likely to choose the simple machines in the Same Output condition (M = 4.44; SD = 1.84) than the Unknown Output condition (M = 3.81; SD = 1.93), Wald χ^2 (1) = 8.23, p = .004. There was also a main effect of trial type, as participants chose the simple machines more in wiring trials (M = 4.38; SD = 1.91) than number trials (M = 3.96; SD = 1.89), Wald χ^2 (1) = 18.63, p < .001. There was a significant condition by trial interaction, Wald χ^2 (1) = 11.12, p < .001.

This interaction resulted because participants in the Unknown Output condition were more likely to select the simple machine in wiring trials than number trials, Wald χ^2 (1) = 20.06, p < .001. However, in the Same Output condition, there was no significant difference in trial type, Wald χ^2 (1) = 0.89, p = .345. For number trials, those in the Same Output condition chose the simple machine more than those in the Unknown Output condition, Wald χ^2 (1) = 16.51, p < .001, while there was no significant difference of condition for wiring trials, Wald χ^2 (1) = 1.24, p = .265.

Single-sample tests using intercept-only GEEs revealed that judgments significantly differed from chance, all ps < .001. The simple machine was chosen significantly above chance for all trials, except the complex machine was favored in the number trial when the function was unknown.

Discussion

Across both trial types, simple machines were rated as better than complex ones (with the exception of the Unknown function number trial). This judgment may have emerged in the Same Output condition since simple machines were seen as more efficient than complex ones—they produced the same output with fewer parts. However, in the Unknown Output condition, when machines' functions were not known to be the same, it was impossible to judge one machine as more efficient than the other. Instead, participants could only judge the machines based on their internal parts. We might have expected participants in this condition to select the complex machine as better, as people infer machines with complex insides to have more complex functions (Ahl & Keil, 2017). Nonetheless, complex machines were only preferred in the number trials, but not wiring trials where instead the simple machine was preferred.

In Experiment 2, we sought to address some limitations from this experiment by modifying the procedure. First, we introduced a cover story so that the comparison of machines' quality was more natural. Second, to ensure participants considered the output of the machines to be identical in the Same Output condition, we also emphasized that the machines had the same performance. So, participants were told that the machines made the same products at the exact same speed. Finally, in the first experiment, machines in the number trials only differed by the number of circuits, while machines in the wiring trials differed by the number and the diversity of parts. This inconsistency may have been what caused a stronger simplicity preference in the wiring trials, as there were two internal complexity cues. In the next experiment, we better controlled the complexity manipulation by ensuring that there was only one cue for complexity for each trial type.

Experiment 2

Methods

Participants We tested 196 adults (M_{age} = 38.05, range = 20-69, 87 female; 107 male; 2 other). Data were excluded from an additional 34 participants who failed at least one of the three post-test comprehension checks.

Procedure To introduce the task, participants saw a blue machine from the 'blue company', and an orange one from the 'orange company' (see Figure 2a). Participants read that they were going to see the inside parts that make the machines go. Similar to Experiment 1, participants saw four pairs of machines and rated which was better on a seven-point scale. Two pairs were number trials where machines differed by the number of circuits, as in Experiment 1. The other two pairs were wiring trials, where machines differed by the number of wired connections within a circuit. The four trials were randomized for each participant, and the location of the simple machine in each pair was counterbalanced. Participants again rated which machine was better on a 7-point scale. After, participants completed similar comprehension checks as in the previous experiment.

Test trials differed from Experiment 1 in a few ways. The color of the machines pertained to the company color introduced in the cover story. In the Same Output condition, the machines were described as making the exact same items (e.g., pens, chocolate bars, toothpaste, or pinwheels), and as making them at the exact same speed. Lastly, the circuitry in the wiring trials were made to only differ by one cue, number. Diversity of parts was not examined in this experiment.

Results

There was a main effect of condition, as participants were more likely to choose the simple machines in the Same Output condition (M = 4.93; SD = 1.77) than the Unknown Output condition (M = 3.34; SD = 1.76), Wald χ^2 (1) = 47.46, p < .001. There was also a main effect of trial type, as participants chose the simple machines more in the wiring trials (M = 4.30; SD = 1.85) than number trials (M = 4.09; SD = 2.00), Wald χ^2 (1) = 8.60, p = .003. There was a significant condition by trial interaction, Wald χ^2 (1) = 4.46, p = .035 (see Figure 2b).

This interaction resulted because participants in the Unknown Output condition were more likely to select the simple machine in wiring trials than number trials, Wald χ^2 (1) = 8.80, p = .003. However, there was no significant

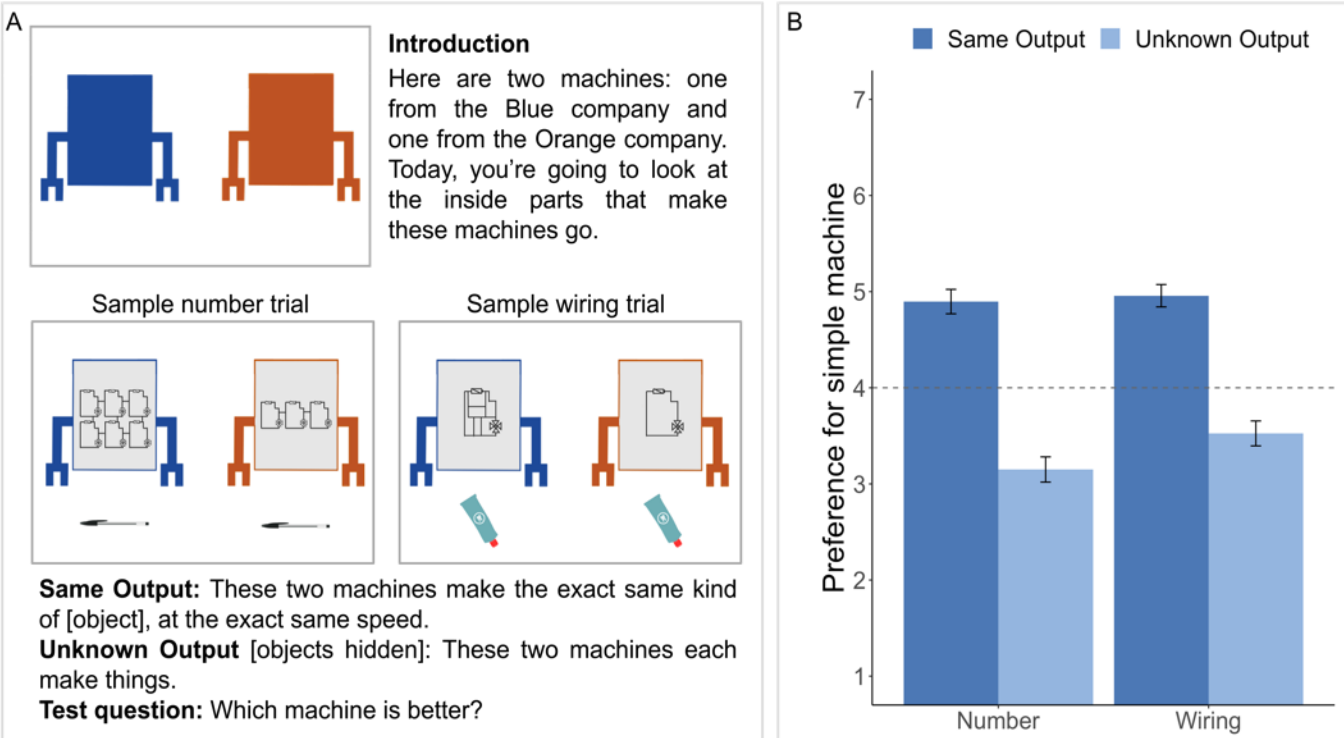

Figure 2. A) Number and wiring trials from Experiment 2 in the Same Function Condition. B) Results from Experiment 2. Error bars represent ± 1 standard error.

difference in trial type in the Same Output condition, Wald χ^2 (1) = 0.61, p = .435. For both number and wiring trials, those in the Same Output condition chose the simple machine more than those in the Unknown Output condition, ps < .001.

Single-sample tests using intercept-only GEEs revealed that participants in the Same Output condition chose the simple machine significantly above chance for both trial types, both ps < .001. However, participants in the Unknown Output condition chose the complex machine significantly below chance for both trial types, both ps < .001.

Discussion

Simple machines were preferred over complex ones when machines had equal function and speed. However, complex machines were preferred when the function and speed were unknown. We consider explanations for this response pattern below.

General Discussion

Across two experiments, we found that people are sensitive to inside parts when judging the quality of machines. People judged that machines with simpler insides were better than ones with more complex insides when machines had identical functions. However, people's choices were mixed when the functions of machines were unknown, though they predominantly thought that machines with complex parts were better. Our findings show that people consider both the inside parts and functions of machines when assessing quality.

This research contributes to our knowledge on the types of judgments people make about inside parts. Past work found that people expect functions and parts of machines to match in complexity (Ahl & Keil, 2017), and that objects with identical functions likely have the same parts (Sobel et al., 2007). Based on this work, we knew that people could infer a machine's parts based on its function, and vice versa. Our work extends these findings by showing that people can also compare machines based on the fit between machines' functions and parts. Comparing machines in this way can lead people to make evaluative judgments, such as which machine is better.

Some previous work has found that people consider the efficiency of objects when making various judgments. In one study, three- to four-year-olds and adults explored objects that performed the same function (Keleman et al., 2012). One tool had few parts whereas the other had extra features not necessary for performing the function. For example, two tools could be used to hang rings on a hook, though the more complex tool had extra ridges and attachments. Participants judged that tools with fewer parts were made for the specific function, and would more easily perform the function. The extra parts on the other tools suggested that they might have alternate functions and are more difficult to use. This work shows that people find objects with fewer parts to be better suited to fulfill particular functions. However, one critical difference between this experiment and ours is that participants in the earlier study actually used the objects, and knowledge gained from this experience could have contributed to their judgments. Our work provides additional evidence of people's reasoning about efficient objects, as people came to assess machines' quality without physically exploring their parts and functions.

People's preference for simpler parts is reminiscent of findings from another area of research—people's assessments of explanations. Research in this area has shown that people find simpler explanations more satisfying than complex ones when explaining the same phenomena (Lombrozo, 2007; 2016), and this preference for simpler explanations is even evident in young children (Bonawitz & Lombrozo, 2012). Our findings reveal that people's preference for simpler parts also emerges when evaluating causal mechanisms.

We were interested in exploring whether type of complexity impacted people's assessments of quality, as complexity can be realized in many ways. In Experiment 1, people judged the simple machine as better more strongly in wiring trials than number trials. This may be due to how complexity was realized in each of these conditions. Number trials only differed in the number of circuits, though wiring trials differed by the number *and* diversity of parts. So, judgments may have been stronger in these trials due to there being two cues for complexity. However, a trial effect still emerged in Experiment 2, when we more closely controlled for complexity. In both trials, machines only differed by the number of circuits or connections, and not by diversity of parts. People judged the simple machine as better more strongly in wiring than number trials. Our work provides preliminary evidence that simplicity preferences persist across types of complexity. Future work could more closely examine different types of complexity, such as those identified in previous work (Ahl et al., 2020).

People may not always judge machines with simple parts as better than ones with complex parts. Recent work has found that people have a complexity-matching preference when evaluating explanations (Lim & Oppenheimer, 2020). People rated simple explanations as more satisfying when evaluating explanations for simple phenomena, but rated complex explanations as more satisfying for complex phenomena. This finding suggests that people believe a satisfying explanation should be as complex as the phenomenon itself. A complexity-matching preference may also be apparent when assessing the quality of machines, as in our study. Although people assessed machines with simple parts as better than machines with complex parts, people may hold the opposite judgment when outputs vary in complexity. For instance, when comparing two machines that have complex outputs, people may judge the machine with complex inside parts to be better than the one with simpler inside parts.

Future work could examine the existence of a complexity-matching preference when judging causal processes. One approach might be to manipulate the complexity of machines' inside parts and their outputs. If people have a complexity-matching preference, they would judge complex machines as better when comparing machines with complex outputs, and simple machines as better when comparing machines with simple outputs. This would suggest an exception to people's inference that efficient machines are better. Indeed, if a complex machine had suspiciously few parts, people may be skeptical that the machine actually performs the function equally well as a machine with many parts. Alternatively, if people do not have a complexity-matching preference, such that they prefer machines with simple inside parts irrespective of the complexity of the outcome, this would support the notion that people prefer simple and efficient causal processes.

One factor that may have influenced the judgments in our task is the decision context. People evaluated the quality of machines by comparing their inside parts, but their evaluations may differ when machines are judged on their own. People's evaluations of two options differ when they are jointly considered versus when they are each considered on their own (e.g., Hsee, 1996). Participants in our experiments may have rated the machines equally if they judged their quality separately. This possibility needs to be explored with future work, but could suggest that qualitative assessments based on efficiency are best made in comparative contexts.

One concern is that judging machines based on which was 'better' is vague. Participants may have been confused about whether they should consider design efficiency, or other things such as which machine made better objects, was newer, was easier to operate, etc. However, in everyday speech, people often ask for general judgements about quality that do not point to specific evaluative features. For example, people can easily judge which of two songs is better, without needing further description of whether 'better' refers to personal preference, originality, significance, popularity, or legacy.

Future work could examine other kinds of judgments that could be made by comparing machines, other than quality. For example, people may infer that someone who designed a machine with simpler parts is smarter than someone who designed a machine with an identical function but more complex parts. Past work has shown that people use artifacts to make inferences about those who created them (Hurwitz et al., 2019; Pesowski et al., 2020). For instance, when two characters build the same design, people were more likely to see this as evidence of copying when the designs were inefficient than efficient (Schachner et al., 2018). Designs may also suggest how creative or intelligent the designer is (Gosling, 2008). Similarly, people may infer that efficient machines are more valuable than inefficient ones (Gelman & Echelbarger, 2019).

To determine when people begin to consider efficient machines as better quality than inefficient ones, future work could examine a younger population. Children from as young as four years may begin to evaluate machines based on their efficiency, given that they consider complexity in other kinds of inferences (Ahl & Keil, 2017; Sobel et al., 2007). Four-year-olds have adult-like inferences about efficient objects' functions and ease of which they perform those functions (Keleman et al., 2012). Preschoolers also infer that an agent is more competent when they finish a complex task at the same speed as another agent with an easier task (Leonard et al., 2019). Altogether, the capacity to compare processes from the parts and outcomes may emerge early in development.

In conclusion, people consider the fit of machines' parts and functions when making assessments of their quality. Efficient machines were inferred to be better than inefficient ones, suggesting that people know that extra parts are suboptimal. When the efficiency of machines cannot be compared because their functions are unknown, complex machines are rated as better, possibly because they are inferred to have more complex functions.

Acknowledgments

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada awarded to OF, and by a grant from the Social Sciences and Humanities Research Council of Canada awarded to SD.

References

Ahl, R. E., Amir, D., & Keil, F. C. (2020). The world within: Children are sensitive to internal complexity cues. *Journal of Experimental Child Psychology, 200*, 104932.

Ahl, R. E., & Keil, F. C. (2017). Diverse effects, complex causes: children use information about Machines' functional diversity to infer internal complexity. *Child Development*, *88*(3), 828-845.

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*(4), 1156.

Erb, C. D., Buchanan, D. W., & Sobel, D. M. (2013). Children's developing understanding of the relation between variable causal efficacy and mechanistic complexity. *Cognition*, *129*(3), 494-500.

Gelman, S. A., & Echelbarger, M. E. (2019). Children, object value, and persuasion. *Journal of Consumer Psychology*, *29*(2), 309-327.

Gelman, S. A., & Kremer, K. E. (1991). Understanding natural cause: Children's explanations of how objects and their properties originate. *Child Development*, *62*(2), 396-414.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*(3), 213-244.

Gosling, S. (2008). *Snoop: What your stuff says about you.* New York: Basic Books

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576.

Hurwitz, E., Brady, T.F., & Schachner, A. (2019). Detecting social transmission in the design of artifacts via inverse planning. In A. Goel, C. Seifert, & C. Freska, *Proceedings of the 41st Annual Conference of the Cognitive Science Society*.

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

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.

Leonard, J. A., Bennett-Pierre, G., & Gweon, H. (2019). Who is better? Preschoolers infer relative competence based on efficiency of process and quality of outcome. In A. Goel, C. Seifert, & C. Freska, *Proceedings of the 41st Annual Conference of the Cognitive Science Society*.

Lim, J. B., & Oppenheimer, D. M. (2020). Explanatory preferences for complexity matching. *PloS one*, *15*(4), e0230929.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*(3), 232-257.

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*(10), 748-759.

Pesowski, M.L., Quy, A., Lee, M., & Schachner, A. (2020). Children use inverse planning to detect social transmission in design of artifacts. In S. Denison, M. Mack, Y. Xu, & B.C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*(3), 429.

Schachner, A., Brady, T.F., Oro, K., & Lee, M. (2018). Intuitive archeology: Detecting social transmission in the design of artifacts. In C. Kalish, M. Rau, J. Zhu, & T. Rogers, *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, *8*(2), 159-182.