Title
It's Complicated: Children Identify Relevant Information About Casual Complexity

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors
Ahl, Richard E
DeAngelis, Erika
Stephenson, Auburn
et al.

Publication Date
2018

It's Complicated: Children Identify Relevant Information About Causal Complexity

Richard E. Ahl (richard.ahl@yale.edu)[1]
Erika DeAngelis (edeangelis19@amherst.edu)[2]
Auburn Stephenson (stepheau@bc.edu)[3]
Sehrang Joo (sehrang.joo@yale.edu)[1]
Frank C. Keil (Frank.Keil@yale.edu)[1]

[1] Department of Psychology, Yale University, Box 208205
New Haven, CT 06520 USA

[2]16 Barrett Hill Dr., Amherst College, AC #0877, Keefe Campus Center
Amherst, MA 01002 USA

[3] Department of Psychology, Boston College, McGuinn 300, 140 Commonwealth Ave
Chestnut Hill, MA, 02467 USA

Abstract

Mechanistic complexity is an important property that affects how we interact with and learn from artifacts. Previous research finds that children successfully detect complexity contrasts when given information about the functions of simple and complex objects. However, do children spontaneously favor relevant information about an object's causal mechanisms and functions when trying to determine an object's complexity? In Study 1, 7–9-year-olds and adults, but not 5–6-year-olds, favored relevant information (e.g., the difficulty in fixing an object) over irrelevant information (e.g., the difficulty in spelling an object's name) for making determinations of mechanistic complexity. Only in Study 2, in which the relevance contrasts were extreme, did the youngest age group favor relevant over irrelevant information. These results suggest that the ability to detect which object properties imply complexity emerges in the early school years; young children may be misled by features that are not truly diagnostic of mechanistic complexity.

Keywords: causal complexity; relevance; intuitive theories

Introduction

Our daily lives are filled with artifacts of great mechanistic complexity. Such artifacts include entirely new categories of objects that have only recently been invented (e.g., smartphones), as well as older objects whose complexity has increased through new technological innovation (e.g., hybrid engines in cars). The property of artifact complexity was almost wholly absent throughout most of human history but has become far more common in the past several decades (Arbesman, 2016). Now, it meaningfully shapes how we learn from, interact with, and make inductive inferences about the objects we encounter.

For example, adults view complex artifacts as harder to use, and especially harder to repair, than simpler artifacts (Kominsky, Zamm & Keil, 2018). Complexity intuitions influence practical judgments regarding how much there is to learn about a given object and whose expertise to seek if help is needed, enabling us to efficiently gather knowledge and allocate labor. Such intuitions also reflect deep distinctions between the causal structures of simple vs. complex artifacts. Simple artifacts generally lack internal mechanisms. The connections between form, function, and mechanism are rather close for simple artifacts (e.g., a knife's sharpness allows it to cut; Bloom, 1998), and may be discernable through casual inspection, even by infants (Baldwin, Markman, & Melartin, 1993). In contrast, complex artifacts have opaque, numerous, diverse, and internal mechanisms that interact to enable functioning (Kominsky et al., 2018; Gelman, 1988; Simon, 1962). Unlike with many simple artifacts, the observable form of complex artifacts may yield few cues to their functions and mechanisms (Keil, Greif, & Kerner, 2007). Thus, many aspects of the specialized, early-emerging cognitive systems humans possess for understanding the simple artifacts that once comprised our entire material culture (Hernik & Csibra, 2009), which rest on close connections between form, function, and mechanism, may be ill-equipped to handle the new kinds of complex objects that have only recently become common.

Complexity can be overwhelming. Even well-educated adults fail to grasp the mechanistic details of everyday objects, let alone those of advanced technologies (Rozenblit & Keil, 2002). Yet, knowing *that* an entity is complex can motivate further explorations and questions (see Cook, Goodman, & Schulz, 2011; Legare, Sobel, & Callanan, 2017) that yield mechanistic details as well as basic facts about how to make the entity work. Even without comprehensive knowledge regarding the specific entity in question, abstract theories about causal complexity can lead one to seek useful information about relevant features (Wilson & Keil, 1998).

Given the ubiquity of complex objects in the industrialized world, the ability to contend with complexity is important in both daily life and academic settings (Fourez, 1997). Inferences about mechanistic complexity also relate to the understanding of causal relations, one of the most fundamental topics in cognitive science. Imagine that a reasonable adult is trying to determine if an object is mechanistically complex. Such a person would know that complexity emerges from how the object works, rather than features such as its color, cost, or size, and would seek information about causally-relevant properties. The question we explore here is whether children also know which features are relevant to an object's complexity.

Young children possess the basic cognitive abilities and intuitive theories that would support preferences for relevant, mechanistic information when making complexity judgments. By the preschool years, children prefer causal to non-causal information (Alvarez & Booth, 2015) and seek unobserved causes for certain physical events (Muentener & Schulz, 2014). Thus, to the extent that children can identify certain features as causal, even if they are not easily observable, children should be motivated to seek information about such features. Children's privileging of internal properties supports the claim that children possess the ability to consider deep, unseen mechanistic features. Preschoolers view an entity's internal properties as more essential to its category membership and functional ability than its external properties, and therefore appreciate the causal significance of non-visible features (Gelman & Wellman, 1991). Thus, children can favor relevant, causally-central features over surface-level features, even if the latter may be more readily observable.

A new and growing body of literature reveals children's general capacity to reason about how complex artifacts may differ from simpler artifacts. This literature indicates that children, like adults, do use functional information to reason about artifacts' complexity. In the preschool and early elementary-school years, children use a machine's variability, diversity, and number of functions to infer its level of internal complexity (Ahl & Keil, 2017; Erb, Buchanan, & Sobel, 2013). Elementary schoolers' ratings of the difficulty in fixing and using real-world objects strongly correlate with their complexity ratings of such objects, suggesting that children value such features when judging complexity (Kominsky et al., 2018). Additionally, learning mechanistic information about an entity causes elementary schoolers to change their complexity judgments more than learning non-mechanistic information, which further shows that children use mechanistic information when evaluating complexity (Trouche, Chuey, Lockhart, & Keil, 2017).

However, while previous research suggests that children view functional and mechanistic information as relevant to complexity, it does not show that children view it as *more* relevant than other kinds of information. Will children privilege relevant information in the midst of irrelevant distractors? In a complex and noisy real-world environment, information about objects is not constrained to relevant functional contrasts. Rather, it includes a range of observable and unobservable properties. Only some of these properties are actually relevant to complexity, while others (e.g., date of invention) may be correlated with relevant properties but are not relevant by themselves. Compounding these issues is the causal opacity of many complex artifacts, which muddles reasoning about what specific features make them complicated. Children also face general limitations on their ability to distinguish relevant from irrelevant information; those under the age of 7 particularly struggle to do so (Bauer & Larkina, 2017; Johnston & Keil, under revision).

Our current studies investigate children's ability to distinguish relevant from irrelevant information about artifact complexity, with the goal of determining the kinds of information children may seek out or evaluate favorably when making complexity judgments. One naturalistic method of determining what information children seek involves encouraging children to ask questions about an entity (e.g., Greif, Kemler-Nelson, Keil, & Gutierrez, 2006). For example, if you wanted to figure out if something is complicated, what would you want to know about it? However, since formulating questions can be difficult for young children, our current studies instead ask children to evaluate the utility of the information provided by experimenters. This method reduces the cognitive load of generating questions and also allows us to closely control the relevance of the information we present.

In Study 1, we present participants with information about properties one can learn through acting upon an object and ask whether a given property is indicative of the object's mechanistic complexity. Some information is relevant to causal mechanisms (e.g., how hard an object is to fix) while other information is related to other object properties but largely irrelevant to causal mechanisms (e.g., how hard an object's name is to spell). Knowing what kinds of information are useful for diagnosing complexity may be a precondition for effective and selective inquiry about complex objects. We predicted that older children and adults would distinguish relevant from irrelevant information but younger children would struggle to do so.

Study 1

Method

Participants Our final sample of children included thirty 5- and 6-year-olds (M_{age} = 73.13 months, *SD* = 6.77), referred to as "younger children," and thirty 7-, 8- and 9-year-olds (*M* = 101.50, *SD* = 10.78), referred to as "older children," tested in museums (*n* = 50) and our lab (*n* = 10). An additional 13 children were excluded for final scale check failure (*n* = 11; 10 were in the younger age group) or other difficulties with comprehension or attention (*n* = 2). Our final adult sample included thirty participants (M_{age} = 38.27 years, *SD* = 12.50) who were recruited and tested online via Amazon's Mechanical Turk and Qualtrics. For brevity's sake, only the methods for children will be described; methods for adults were similar except for minor changes to allow for online administration. Six additional adults were excluded for short study durations (*n* = 3) or comprehension failures (*n* = 3).

Scale introduction Participants were shown five blue circles, oriented horizontally, whose coloring ranged from almost empty ("really bad," the leftmost circle, labeled with a 1) to completely full ("really good," the rightmost circle, labeled with a 5). The experimenter moved from left to right and pointed to the appropriate circle while explaining that the circles indicated information that was "really bad," "bad," "just okay," "good," or "really good" for helping them understand something (see Mills, Danovitch, Rowles, & Campbell, 2017). As a comprehension check, participants were asked to point to the "really bad," "just okay," and "really good" circles. The experimenter explained that "some objects are complicated in terms of how they work"; we defined "complicated" as meaning that "it's really hard to learn how [they] work." We included this definition primarily for the benefit of children who may not know the term, but for the sake of consistency, all children and adults heard it.[1] To find out "the best way to learn if an object is complicated," the experimenter asked different people for answers. The participants' task was to use the circles to indicate whether each answer (i.e., piece of information) "helps us decide" an object's level of complexity.

Scale rating task Participants separately rated six different items, which are shown in Table 1. Three items were intended as relevant[2] to mechanistic complexity (henceforth labeled "relevant"), warranting high ratings, while the other three items ("irrelevant") were intended as irrelevant, warranting low ratings. The items were read aloud in a new randomized order for each participant using Qualtrics, with relevant and irrelevant items interspersed. As each item was introduced, the experimenter stated, "one person said that the best way to tell if an object is complicated is to learn if it's hard to [*item*; e.g., build, to make it or put it together]. Does knowing if an object is hard to [build] help us decide if it's complicated in terms of how it works?" Participants then pointed to the circle scale to rate each item.

Table 1: Relevant and irrelevant items for each study.

Relevant items	Irrelevant items
Study 1	
Use (as it should be used)	Destroy (to break it)
Fix (if it's broken)	Spell its name (using letters)
Build (make it, or put it together)	Cover (w/ a sheet to keep it clean)
Study 2	
Use (as it should be used)	Put a piece of tape on it
Fix (if it's broken)	See it at night
Build (make it, or put it together)	Sing a song close to it

Forced-choice task Next, participants completed three trials of forced-choice judgments. Items from the scale rating task were discussed again. Each trial paired a relevant item with an irrelevant item. The participants' task was to decide which item in each pair would be most helpful for figuring out an object's complexity (e.g., finding out how hard an object is to build vs. how hard an object is to spell). The pairs' presentation order and which item per pair was mentioned first was randomized for each participant. To ensure that a given participant encountered all items exactly once during this task, three versions of the task were created (e.g., "build" was paired with "cover" in version 1, "spell" in version 2, etc.). Version assignment was counter-balanced with respect to age and gender for children and randomized for adults. For each trial, participants were given a 1 or a 0 for choosing the relevant or irrelevant item as helpful, respectively; total scores could range from 0 to 3.

Final scale comprehension check A final check tested whether participants had a general ability to distinguish relevant from irrelevant information and use our scale properly. Participants were told about two characters who were eating apples, with presentation order randomized. One character said they decided to eat an apple because apples are their favorite food (relevant), and the other said they decided to eat an apple because their friend likes wearing sneakers (irrelevant). For each character, participants used the circle scale to rate whether the response helped them learn why the character decided to eat an apple. Only participants who assigned a higher rating to the relevant than irrelevant response were included in the final sample.

Results

Scale ratings We predicted that the older age groups would assign higher overall ratings to relevant than irrelevant items but had no item-specific predictions. We generally found similar ratings for all items of a given category for each age group (see Figure 1); in no case would our main conclusions be altered by removing a single higher- or lower-scoring item. Thus, our main analyses compared the mean ratings of irrelevant vs. relevant items, which were computed for each type by averaging the ratings given to its three items.

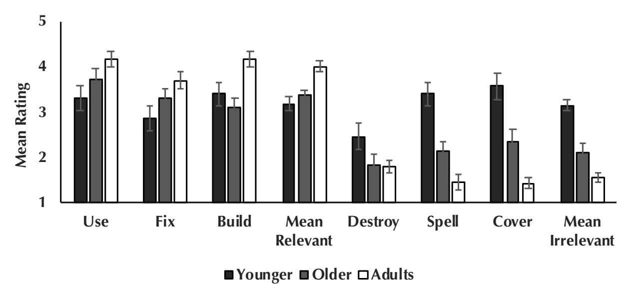

Figure 1: Mean ratings for Study 1 items and item types. Error bars indicate ± 1 SEM, for this and all other figures.

A mixed-design analysis of variance (ANOVA) was conducted for the children's data, with child age as the between-subjects factor, item type as the within-subjects factor, and rating as the dependent measure. There were

[1] Our working definition emphasized mechanism ("*how* it works") and was used because piloting found that some children do not know what "complicated" means. How children would have scored if "complicated" had not been defined (see Kominsky et al., 2018), or if just our definition (without mentioning "complicated") had been given, is an open question, but we believed providing both the term and a definition was superior to providing only one of them.

[2] The difficulty in using an object is *less* diagnostic of mechanistic complexity than the difficulty in fixing or building it (particularly for complex technologies, which are often designed with ease of use in mind) but is still relevant to mechanistic complexity (see Kominsky et al., 2018), and certainly more relevant than the information conveyed in the irrelevant items.

significant main effects for item type, $F(1,58) = 17.12$, $p < .001$, $\eta^2_p = .23$, and age, $F(1,58) = 8.88$, $p = .004$, $\eta^2_p = .13$, which were qualified by a significant interaction between item type and age, $F(1,58) = 14.90$, $p < .001$, $\eta^2_p = .20$.

Due to our *a priori* analysis plan and the significant interaction, planned paired-samples *t* tests were conducted to compare the mean ratings given to relevant vs. irrelevant items for each age group separately, with Bonferroni corrections resulting in an α of $p < .025$. The mean ratings for relevant items minus the mean ratings for irrelevant items (i.e., the difference scores) are shown in Figure 2. In Study 1, younger children assigned similar ratings to relevant and irrelevant items, $M_{difference} = .04$, $SD = 1.08$, $t(29) = .27$, $p = .82$, Cohen's $d = .04$. However, the difference between the two item types was significant for older children, who assigned higher ratings to relevant items, $t(29) = 5.08$, $p < .001$, $d = .93$. (Adults also assigned higher ratings to relevant items, $t(29) = 14.37$, $p < .001$, $d = 2.62$.)

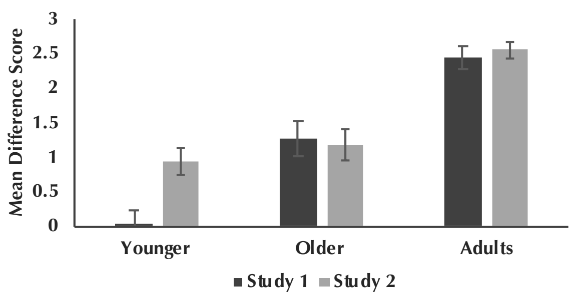

Figure 2: Mean difference score (relevant rating minus irrelevant rating) for each age group and study.

Forced choice task One-sample *t* tests compared each age group's mean forced-choice score to the chance score of 1.5. As shown in Figure 3, younger children scored marginally better than chance in Study 1, $t(29) = 1.92$, $p = .06$, $d = .35$. Older children, $t(29) = 5.33$, $p < .001$, $d = .97$, and adults, $t(29) = 44.00$, $p < .001$, $d = 8.03$, scored significantly better than chance.

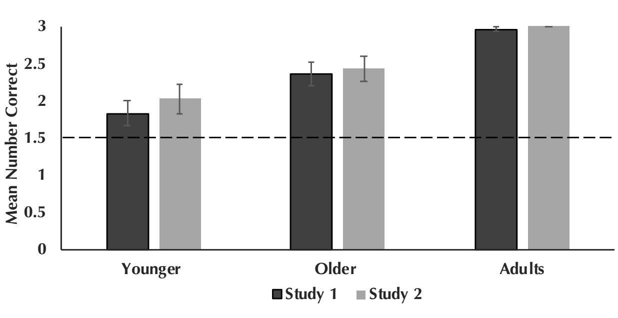

Figure 3: Mean number of correct forced-choice answers for each age group and study. Scores could range from 0 to 3.

Discussion

We found that older children and adults robustly distinguished relevant and irrelevant information regarding complexity, while younger children did not do so. Younger children succeeded in assigning high ratings to relevant information but, unlike the older age groups, they failed to assign low ratings to irrelevant information, and therefore rated both item types similarly. Thus, younger children were persuaded by irrelevant items that older children dismissed as unhelpful. Younger children's scores were better in the forced-choice task but were still only marginally above chance. The high number of final scale check failures, which is one limitation of our study, indicates that many children struggle with broader relevance judgments. However, since data from such children were excluded, the poor scores of included children are not easily explainable by difficulties with relevance judgments in general but rather difficulties with relevance judgments regarding complexity specifically.

In Study 1, the irrelevant items were deemed quite unhelpful by the older age groups but still bore some relation to other object properties, including those that may be deemed relevant to complexity. For instance, the difficulty in covering an object with a sheet is influenced by the property of size, which may be viewed as correlated with complexity. The way in which the irrelevant Study 1 actions are performed is somewhat dependent on the identities of their target objects (e.g., an object's size affects how one covers it). Thus, in Study 2, we created a new set of more irrelevant items. We chose actions that are performed in essentially the same way regardless of the identities of their target objects. Perhaps younger children will only assign lower ratings to irrelevant items when the contrasts are extreme. If, however, younger children have entrenched difficulties identifying irrelevant information about complexity, or other surface features of the study script are problematic for this age range (e.g., the high verbal load, which is one limitation of our study, or the use of the term "complicated," which children may struggle with even though we provided a definition), their performance should not improve in Study 2.

Study 2

Piloting

To test whether the irrelevant items intended for Study 2 were rated as less helpful and less relevant than the irrelevant items in Study 1, we ran pilots on two separate and new groups of adult participants. We doubled our sample size to 60 participants per study *a priori* because our key comparisons were between two sets of irrelevant and presumably low-scoring items; we expected floor effects to reduce statistical power. We briefly describe our piloting here.

In Pilot Study 1, our procedure was similar to that of Study 1 except that we added the Study 2 items and used a 100-point scale instead of a 5-point scale, to allow for a greater range in responses. A repeated-measures ANOVA with a Greenhouse-Geisser correction compared average scores of relevant, Study 1 irrelevant, and Study 2 irrelevant items. The effect of item type was significant, $F(1.77,104.43) = 823.61$, $p < .001$, $\eta^2_p = .93$. Follow-up planned comparisons (with Bonferroni corrections resulting in an α of $p < .025$) found that the Study 2 irrelevant items received lower ratings than the relevant items ($p < .001$) and, crucially, than the Study 1 irrelevant items ($p < .001$).

In Pilot Study 2, we tested whether participants viewed the Study 2 irrelevant items as less diagnostic of object properties than the other items. We asked participants to imagine that they did not know the identity of an object; would finding out a given piece of information help one figure out other properties of the object? Participants rated the helpfulness of each item on a 100-point scale. A repeated-measures ANOVA with a Greenhouse-Geisser correction compared average scores for the three item types. The effect of item type was significant, $F(1.64, 96.50) = 22.15$, $p < .001$, $\eta^2_p = .27$. Follow-up planned comparisons (with Bonferroni corrections resulting in an α of $p < .025$) found that the Study 2 irrelevant items received lower ratings than the relevant items ($p < .001$) and, crucially, than the Study 1 irrelevant items ($p = .003$).

Our piloting found that the Study 2 irrelevant items were deemed less helpful, and also less related to other object properties, than the Study 1 irrelevant items, and were thus suitable for subsequent testing.

Method

Participants and procedure Our final sample of children included thirty 5- and 6-year-olds ($M_{age} = 73.20$ months, $SD = 6.65$) and thirty 7-, 8- and 9-year-olds ($M = 103.47$, $SD = 11.01$), tested in museums ($n = 55$) and our lab ($n = 5$). An additional 21 participants were excluded due to final scale check failure ($n = 15$; 13 were in the younger age group) or other comprehension, attention, or administration difficulties ($n = 6$). Our final adult sample included 30 participants ($M_{age} = 35.63$ years, $SD = 10.80$), recruited as described previously. Eleven additional participants were excluded for short study durations ($n = 6$) or comprehension failures ($n = 5$). The procedure was identical to Study 1 except that we used a new set of irrelevant items, as described previously and displayed in Table 1 and Figure 4.

Results

Scale ratings As shown in Figure 4, we generally found similar scores for all items of a given type for each age group. A two-way mixed-design ANOVA for the children's data found a significant main effect for item type, $F(1,58) = 50.12$, $p < .001$, $\eta^2_p = .46$, with lower ratings for irrelevant items, and a significant main effect for age, $F(1,58) = 7.93$, $p = .007$, $\eta^2_p = .12$, with the older children assigning lower ratings than the younger children. Unlike in Study 1, the interaction between item type and age was not significant, $F(1,58) = .61$, $p = .44$, $\eta^2_p = .01$.

Despite the lack of a significant interaction, we conducted planned paired-samples t tests comparing mean ratings for relevant vs. irrelevant items for each age group separately (recall that the younger age group assigned similar ratings to the two item types in Study 1), with Bonferroni corrections resulting in an α of $p < .025$. Mean ratings for relevant items minus the mean ratings for irrelevant items are shown in Figure 2. The younger children assigned significantly higher ratings to relevant items, $t(29) = 4.84$, $p < .001$, $d = .88$, as did older children, $t(29) = 5.18$, $p < .001$, $d = .94$. (Adults also did so, $t(29) = 22.77$, $p < .001$, $d = 4.13$.)

Forced choice As shown in Figure 3, both younger children, $t(29) = 2.66$, $p = .013$, $d = .49$, and older children, $t(29) = 5.70$, $p < .001$, $d = 1.04$, scored significantly better than chance. All adults received perfect scores of 3.0.

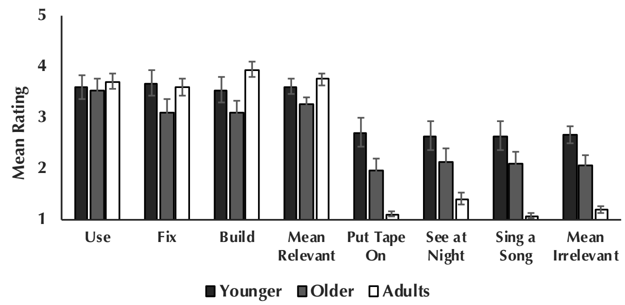

Figure 4: Mean ratings for Study 2 items and item types.

General Discussion

Study 2 replicated the success of the older age groups seen in Study 1; again, older children and adults gave higher ratings to relevant information and also preferred such information in the forced-choice trials. Given the greater relevance contrasts in Study 2, it may seem surprising that older children did not more strongly differentiate the two item types in Study 2, or come closer to converging with adults' scores, relative to Study 1. This pattern may be attributable to floor and ceiling effects in children's use of the scale (e.g., a reluctance to use scale endpoints). Our findings suggest that, starting around age 7, children know the kinds of information that do and do not imply mechanistic complexity. Such knowledge likely reflects greater experience with complex systems and a broader understanding of causal relations. It can be utilized to efficiently investigate complex entities, which are important targets for mechanistic learning.

Unlike in Study 1, younger children succeeded at privileging relevant information in Study 2; they assigned lower ratings to irrelevant items and also avoided such items in the forced-choice trials. Their performance indicates that young children can dismiss completely irrelevant information, specifically regarding actions that are not diagnostic of meaningful object properties, as unhelpful for determining complexity. However, their difficulty in Study 1 indicates that young children are easily misled by irrelevant information that bears some relation to object properties but is not actually diagnostic of complexity.

Do 5- and 6-year-olds simply not value information about causal mechanisms? This conclusion is too extreme, particularly given an extensive body of literature indicating that young children do value such information, and given their high ratings of information relevant to causal mechanisms in current studies. Rather, their difficulty may stem from a failure to conclude that something is *unrelated* to causal mechanisms. Perhaps young children are too reluctant to view true but irrelevant information as unhelpful. Given the difficulties of learning about complex causal mechanisms, which arise through numerous and diverse properties that are often mysterious even to adults, young children may grasp at any potential entries to acquiring such information, even those that are dead-ends.

Does this mean that children do not know which real-world objects are complicated? Not necessarily; children frequently do well at identifying complex objects as such (Kominsky et al., 2018), but this phenomenon may be because relevant information about objects' functions and mechanisms is often selected for pedagogical demonstration by adults and then quickly learned by children (e.g., Butler & Markman, 2014). Young children may falter when learning new object features (which may be causally irrelevant) and integrating them into broader complexity judgments, particularly when such learning is not closely guided by adults. A key challenge for young children appears to be knowing when *not* to care about a new piece of information. Only by limiting informational search to relevant factors can children and adults effectively make sense of a complicated world.

Acknowledgments

We thank Yale Peabody Museum of Natural History, Living Laboratory ® at Stepping Stones Museum for Children, and our participants. This research was supported by NSF Grant DRL 1561143, awarded to Frank C. Keil.

References

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*, 828-845.

Alvarez, A. L., & Booth, A. E. (2015). Preschoolers prefer to learn causal information. *Frontiers in Psychology*, *6*.

Arbesman, S. (2016). *Overcomplicated: Technology at the Limits of Comprehension*. New York,NY: Penguin Books.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, *64*, 711-728.

Bauer, P. J., & Larkina, M. (2017). Realizing relevance: The influence of domain-specific information on generation of new knowledge through integration in 4-to 8-year-old children. *Child Development*, *88*, 247-262.

Bloom, P. (1998). Theories of artifact categorization. *Cognition*, *66*, 87-93.

Butler, L. P., & Markman, E. M. (2014). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*, *130*, 116-127.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*, 341-349.

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*, 494-500.

Fourez, G. (1997). Scientific and technological literacy as a social practice. *Social Studies of Science*, *27*, 903–936

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.

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

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science*, *17*(6), 455-459.

Hernik, M., & Csibra, G. (2009). Functional understanding facilitates learning about tools in human children. *Current Opinion in Neurobiology, 19*, 34-38.

Johnston, A., & Keil, F.C. (under revision). Learning the relevance of relevance and the trouble with truth: Evaluating explanatory relevance across childhood.

Keil, F. C., Greif, M. L., & Kerner, R. S. (2007). A world apart: How concepts of the constructed world are different in representation and development. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation*. New York: Oxford University Press.

Kominsky, J. F., Zamm, A. P., & Keil, F. C. (2018). Knowing when help is needed: A developing sense of causal complexity. *Cognitive Psychology*, *42,* 491-523.

Legare, C. H., Sobel, D. M., & Callanan, M. (2017). Causal learning is collaborative: Examining explanation and exploration in social contexts. *Psychonomic Bulletin & Review*, *24*, 1548-1554.

Mills, C. M., Danovitch, J. H., Rowles, S. P., & Campbell, I. L. (2017). Children's success at detecting circular explanations and their interest in future learning. *Psychonomic Bulletin & Review*, *24*, 1465-1477.

Muentener, P., & Schulz, L. (2014). Toddlers infer unobserved causes for spontaneous events. *Frontiers in Psychology*, *5*.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521-562.

Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*, 467-482.

Trouche, E., Chuey, A., Lockhart, K. L., & Keil, F. C. Why teach how things work? Tracking the evolution of children's intuitions about complexity. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Wilson, R. A., & Keil, F. (1998). The shadows and shallows of explanation. *Minds and Machines*, *8*, 137-159.