How do the semantic properties of visual explanations guide causal inference?

Holly Huey
Department of Psychology
University of California, San Diego
hhuey@ucsd.edu

Caren M. Walker
Department of Psychology
University of California, San Diego
carenwalker@ucsd.edu

Judith E. Fan
Department of Psychology
University of California, San Diego
jefan@ucsd.edu

Abstract

What visualization strategies do people use to communicate abstract knowledge to others? We developed a drawing paradigm to elicit visual explanations about novel machines and obtained detailed annotations of the semantic information conveyed in each drawing. We found that these visual explanations contained: (1) greater emphasis on causally relevant parts of the machine, (2) less emphasis on structural features that were visually salient but causally irrelevant, and (3) more symbols, relative to baseline drawings intended only to communicate the machines' appearance. However, this overall pattern of emphasis did not necessarily improve naive viewers' ability to infer how to operate the machines, nor their ability to identify them, suggesting a potential mismatch between what people believe a visual explanation contains and what may be most useful. Taken together, our findings advance our understanding of how communicative goals constrain visual communication of abstract knowledge across behavioral contexts.

Keywords: visual communication; explanation; causal reasoning; object identification; sketch interpretation

Introduction

From infants exploring the objects in their environments to scientists exploring the frontiers of our solar system, we seek to explain our observations and use that knowledge to generate desired outcomes. Although acquiring such knowledge firsthand can often be costly in time and effort (Lagnado & Sloman, 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003), our ability to transmit and build upon knowledge previously learned by others is a fundamental aspect of human cognition (Boyd, Richerson, & Henrich, 2011). This propensity for sharing what we know has enabled us to accumulate rich knowledge about the structure of our world.

Explanations provide an important means of conveying causal knowledge. Prior work has established that people tend to prefer explanations that are simple, abstract, and broad (Lombrozo, 2006, 2016) and that generating explanations can also yield benefits for one's own learning (Chi, De Leeuw, Chiu, & LaVancher, 1994; Fonseca & Chi, 2011). In particular, learners who engage in explanation are more likely to privilege causal information over visual similarity when judging which objects share latent properties (Walker, Lombrozo, Legare, & Gopnik, 2014) and to selectively remember causally relevant information (Legare & Lombrozo, 2014) than learners who do not explain. While studies investigating the consequences of self-explanation have shed light on the specific cognitive processes accompanying explanation-seeking behavior, they leave open key questions about how people produce explanations that effectively transmit causal knowledge to others (Csibra & Gergely, 2009).

First, what distinguishes the content and organization of explanations from that of merely descriptive reports? To date, few studies have analyzed what specific information people include in their explanations (Williams & Lombrozo, 2010), often relying instead on holistic classifications (Legare & Lombrozo, 2014; Walker et al., 2014). However, a more detailed characterization of causal explanations is critical for advancing our understanding of how people transform their direct experience with the world into compressed representations that explain how things work, cast at an appropriate level of abstraction. A promising strategy for addressing this gap may be to exploit the rich information contained in visual explanations (Bobek & Tversky, 2016; Hegarty, Carpenter, & Just, 1991). Because visualizations share visual-spatial features in common with the physical objects that they depict (Tversky, 2015), this approach enables us to identify the relationships between people's causal knowledge and the specific information they opt to include in their explanations.

Second, what properties of explanations are critical for supporting the successful transmission of causal knowledge to others? While much prior work has focused on measuring *judgments* of the quality of explanations (Lombrozo, 2016), there has been less work examining how explanations guide downstream learning *behavior*. Given that there is sometimes a mismatch between what people think could be useful in pedagogical contexts and what is actually useful for supporting learning (Bonawitz et al., 2011), it is important to validate the apparent quality of an explanation against how well it actually supports the social transmission of knowledge.

Guided by these two overarching questions, the current paper investigates: (1) what information people choose to include in visual explanations of novel mechanical systems; (2) how this information differs from that contained in visual depictions of the same systems; and (3) the behavioral consequences these visual explanations have on naive observers relying only on these explanations to learn about these systems. Our work builds on initial insights gained from recent studies using drawing paradigms to elicit visual explanations of simple mechanical systems (Bobek & Tversky, 2016; Heiser & Tversky, 2006). We explore the hypothesis that producing effective visual explanations of causal phenomena relies on combining information about structure (i.e., what kinds of

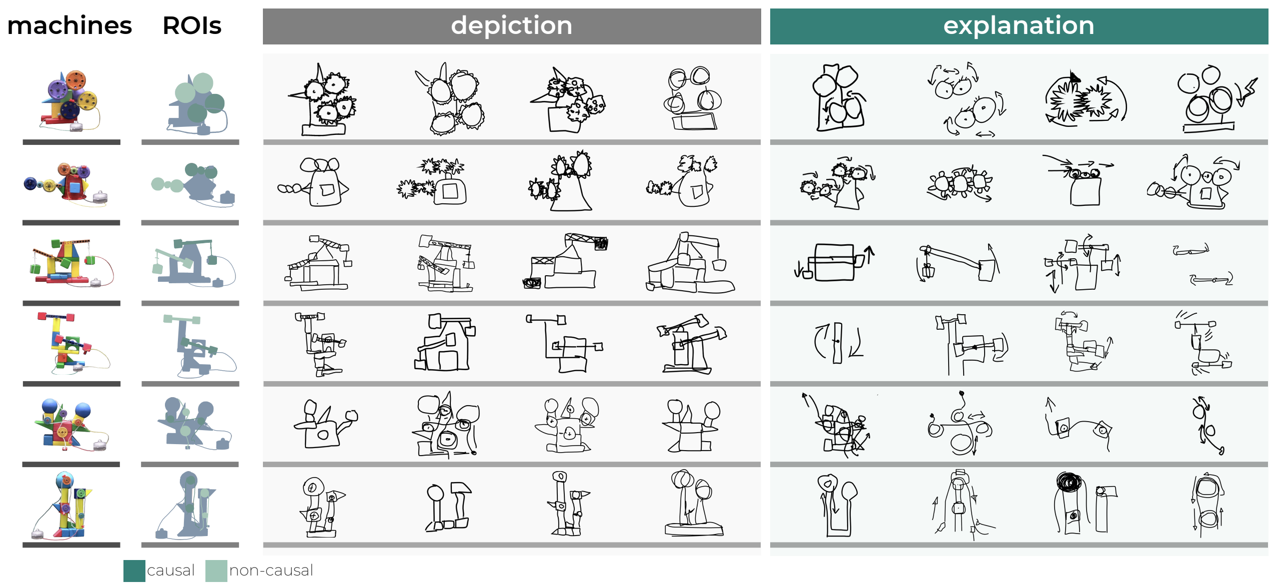

Figure 1: *Left:* Each machine consisted of multiple functional and structural elements. Each region-of-interest (ROI) image indicates the location of causal and non-causal mechanical elements for illustration purposes. *Right:* Example depictive and explanatory drawings.

entities there are) with information about function (i.e., how these entities interact). Concretely, we predicted that effective visual explanations tend to highlight causally relevant information for the function of objects, while preserving enough structural information to establish how viewers should map that information back to the target system.

To evaluate this prediction, we developed a novel drawing paradigm in which participants observed how a machine could be used to activate a light bulb, after which they drew a visual explanation intended to help a naive viewer understand how the machine worked. To identify the specific semantic properties that distinguish visual explanations, participants also drew depictions to help a naive viewer identify the machine by its appearance, thus establishing a baseline for comparison of drawings generated in the absence of a communicative goal to convey causal knowledge. We then used crowdsourcing to obtain detailed annotations of the semantic information conveyed in each drawing (i.e., how each drawn stroke corresponded to parts of the machine). Finally, we presented these drawings to naive viewers and measured how quickly and accurately they could be used to either identify the machine or infer how to operate it. By systematically measuring the semantic properties that characterize visual explanations, as well as the downstream behaviors they support, these studies advance our understanding of the cognitive constraints on the visual communication of causal knowledge.

Experiment 1: What information is prioritized in visual explanations of causal knowledge?

Our first goal was to identify the semantic properties that characterize visual explanations of causal knowledge. To accomplish this, we developed a web-based drawing platform in which participants were presented with a series of novel machines and asked to produce two kinds of drawings: on *explanation* trials, they were prompted to produce visual explanations to help a naive viewer learn how the machine functioned in order to operate it; on *depiction* trials, they were prompted to produce visual depictions to help a naive viewer identify the machine by its appearance. To identify the properties that are distinctive of visual explanations, we use depictions as a baseline for comparison, which were produced in the absence of any explicit goal to communicate causal information. We chose drawing in our visual production task because it is a basic visualization technique that requires minimal equipment (i.e., any stylus and surface), but is a versatile and accessible technique for communicating information in visual form (Sayim & Cavanagh, 2011). Moreover, people have a robust ability to interpret drawings, despite the fact that drawings produced by novices may omit many details and distort the size and proportion of represented objects (Eitz, Hays, & Alexa, 2012; Fan, Yamins, & Turk-Browne, 2018).

Method

Participants 52 participants (27 male; M_{age} = 39.1 years) recruited from Amazon Mechanical Turk completed the experiment. We excluded data from two participants, who did not meet our preregistered inclusion criteria for generating drawings that represented the target stimuli. In this and all subsequent studies, participants provided informed consent in accordance with our institution's IRB.

Stimuli In order to obtain a diverse dataset of drawings from participants, we constructed 6 novel machines consisting of three types of mechanical elements: gears, levers, and pulleys (Fig. 1, *left, machines*). These elements were chosen both because of their simplicity and likely familiarity to many participants, but also their pervasiveness in complex mechanical systems. Half of the elements in our stimuli could be manipulated to activate an attached light bulb (causal), whereas the other half could not (non-causal). Within each machine, these causal and non-causal elements were equated in size, number, and type (i.e., gear vs. lever) to ensure that they were matched in perceptual salience (Fig. 1, *left, ROIs*). During test trials, participants viewed 30s video demonstrations of how to operate the machines to activate the light bulb, in which they saw a hand manipulate the different mechanical elements. The sequence in which the causal and non-causal elements were manipulated in these videos was temporally counterbalanced following an ABBA or BAAB order, where A refers to the causal element, and B refers to the non-causal element. Within each machine type, they also varied across two levels of complexity (e.g., one gear machine contained 4 gears, while the other contained 6).

Visual production experiment All participants completed 6 trials, in which they were presented with each machine once in a randomized sequence. For half of these machines, participants produced a visual explanation; for the remaining half, they produced a depiction. The order in which they produced each type of drawing was randomized across participants. On each trial, participants watched a video demonstration of how to operate a machine, and were then prompted to either produce a visual explanation of how the object functioned or a depiction of what the object looked like (Fig. 2A). Following each demonstration, the video was removed from view and participants were asked to produce a drawing of the machine. The prompt remained on screen for the duration of each trial. Participants used their cursor to draw in black ink on a digital canvas embedded in their web browser (canvas = 500 x 500px; stroke width = 5px). Each stroke was rendered in real time on the participant's screen as they drew and could not be deleted once drawn. Participants also completed 2 practice trials prior to test trials to ensure that they were familiar with the drawing interface.

Semantic part annotations The resulting dataset contained 300 drawings from 50 unique participants: 150 visual explanations and 150 depictions (Fig. 1, *right*). To measure the semantic content that might distinguish visual explanations and depictions, we crowdsourced annotations of the drawings from a separate group of 140 participants (59 male; M_{age} = 38.8 years) from Amazon Mechanical Turk using an annotation paradigm adapted from a prior study (Mukherjee, Hawkins, & Fan, 2019). Annotators were presented with a set of 10 drawings randomly sampled from the visual production experiment. Each drawing was accompanied by a reference photograph of the machine it corresponded to, where each mechanical and structural element was numbered and color-coded to facilitate identification. For each stroke in a given drawing, annotators provided a label selected from a menu of machine-specific part labels (e.g., "gear 1", "background", "lever 2"). If annotators judged that a stroke did not represent a machine element but was instead some kind of symbol

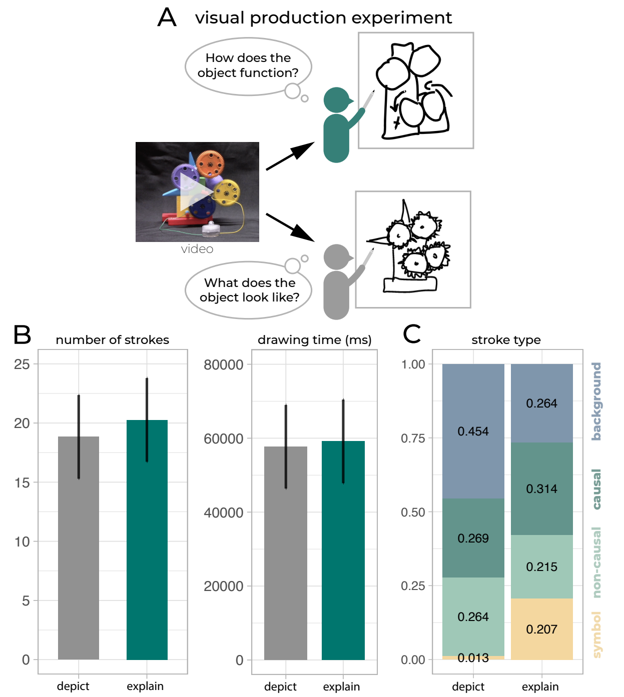

Figure 2: Visual production experiment. (A) Participants watched a video demonstration and were prompted to produce a visual explanation or depiction of the machine. (B) Number of strokes and amount of time spent drawing in each condition. Error bars represent 95% CIs. (C) Semantic annotations of how participants allocated strokes to different elements.

(e.g., arrow, motion line), they were asked to additionally label which element(s) the symbol referred to. We excluded data from 28 participants, who did not meet our preregistered inclusion criteria (i.e., low accuracy on attention-check trials, response time <5s). We ensured that each drawing in our dataset was annotated by at least 3 participants.

Results

Participants used a similar number of strokes (explanation: 20.3; depiction: 18.8; $b = 1.44$, $t = 1.04$, $p = 0.301$; Fig. 2B, *left*) and amount of time drawing in both conditions (explanation: 59136ms; depiction: 57689ms; $b = 1447$, $t = 0.453$, $p = 0.651$; Fig. 2B, *right*), suggesting that they devoted a similar degree of effort when producing visual explanations and depictions. We also analyzed the inter-rater consistency of annotators in order to determine how often annotators agreed on what each stroke in a drawing represented. We found that 93.2% of strokes in visual explanations received the same label by at least two of the three annotators, and 96.9% of strokes in depictions received the same label by at least two of the three annotators. Moreover, 55.5% of strokes in visual explanations and 75.02% of strokes in depictions received the same label by all three annotators. In subsequent analyses, we collapsed over interannotator variation and assigned the modal label to strokes to which at least two annotators had given the same label. Strokes were then assigned to different stroke types: "causal" for strokes that represented mechanical elements that were causally relevant for operating the machine to activate the light bulb, "non-causal" for strokes that represented mechanical elements that were manipulated in the video demonstration but that did not activate the light bulb, "background" for strokes that represented static elements of the machine that were not manipulated, and "symbol" for strokes that represented symbols (e.g., arrow, motion line) lines that appeared to depict latent relations. Leveraging these semantic part annotations, results revealed that visual explanations contained a higher proportion of strokes depicting causally relevant information (e.g., gears that activated the light bulb) than non-causally relevant elements (e.g., gears that did not activate the light bulb), relative to depictions (explanation: 59%, depiction: 50%, b = 0.417, z = 3.93, p = 8.56$e-05$; Fig. 2C). Furthermore, visual explanations contained a higher proportion of symbols, (e.g., arrows, motion lines; explanation: 20.7%, depiction: 1.3%, $b = 9.91$, $t = 4.62$, $p = 1e-05$). These drawings also contained fewer strokes depicting static background elements, relative to depictions (explanation: 26.4%, depiction: 45.4%, $b = -7.58$, $t = -5.49$, $p = 1.16e-07$). These results suggest that the goal of communicating causal knowledge systematically shifts drawings toward more abstract, functional information, even at the expense of fidelity to other visually salient features of the object.

Experiment 2: How well do visual explanations support downstream behaviors?

In Experiment 1, we found that having the goal of communicating causal knowledge impacts what information people prioritize. However, a critical test of how *useful* such communicative strategies are is to evaluate how well other people can interpret these drawings to achieve their own behavioral goals. In our next set of experiments, we recruited two additional cohorts of naive participants to view the drawings made in the visual production experiment and measured how well each drawing supported their ability to identify the original machine (*Experiment 2A*) or to infer how to operate the machine to activate the light (*Experiment 2B*).

Experiment 2A: Using drawings to identify objects

In Experiment 2A, our goal was to test the hypothesis that the reduced emphasis on structural elements (i.e., 'background') in visual explanations would make it harder to match it to the original machine, relative to visual depictions. If so, we predicted that naive viewers would be slower and more error prone in an identification task with visual explanations than with depictions.

Participants 52 participants (24 male; M_{age} = 20.5 years) were recruited from the study participant pool at our institution. Data from two sessions were excluded for technical problems (i.e., inability to click a stimulus).

Stimuli & Procedure Each participant was presented with all 300 drawings generated in the visual production experiment in a randomized sequence. At the beginning of each trial, participants moved their cursor to a crosshair displayed

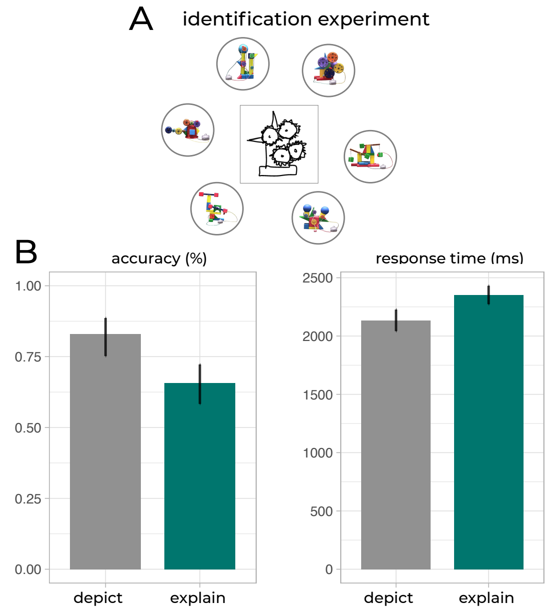

Figure 3: Identification experiment. (A) Example task display. Participants identified which machine the target drawing represented. (B) Accuracy and response time. Error bars represent 95% CIs.

at the center of an empty display. When ready, participants clicked this crosshair to reveal a single drawing (175 x 175px) at that location, surrounded by a circular array of six color photographs (125 x 100px, radius = 250px), one of each machine. The angular distance between each photo was constant (i.e., 60°), and the locations of the machine photos were randomized between trials. Participants were instructed to click on the machine that the drawing corresponded to as quickly and accurately as possible. At the beginning of the session, participants completed 6 practice trials where they were cued with *photos* of each machine (instead of drawings), and had to click on the matching photo in the array.

Results

To investigate how well the drawings supported participants' ability to identify the machines above chance, we fit a null model predicting identification accuracy, including random intercepts for different production participants. Although there were 6 machines, we defined chance-level performance to be at 50%, a theoretical upper bound reflecting our expectation that confusions would likely be between machines of the same type (e.g., gears). We found that participants were reliably above chance performance when cued with both visual explanations and depictions (explanation: $b = 0.578$, $z = 4.27$, p = 1.99$e-05$; depiction: b = 1.27, z = 10, $p<2e-16$; Fig. 3B, *left*). Next, to evaluate our primary hypothesis concerning differences between conditions, we fit a linear mixed-effects model predicting response time from condition, as well as additional predictors controlling for the number of each type of stroke within a drawing (i.e., causal, non-causal, background, symbol), the interaction between condition and the number of each type of stroke, and random intercepts for individual drawings and participant. Consistent with our prediction, participants were slower to respond (correct trials only: explanation: 2351ms; depiction: 2132ms; $b = 9.79e-02$, $t = 3.091$, $p = 0.002$; Fig. 3B, *right*) when cued with a visual explanation than with a depiction. Using a logistic-regression model sharing the same structure to predict accuracy, we found that participants were less accurate when cued with a visual explanation, relative to a depiction (explanation: 66%; depiction: 83%; $b = -0.737$, $z = -2.32$, $p = 0.0201$; Fig. 3B, *left*). These results show that the differences in semantic information contained in these drawings have distinct behavioral consequences: visual explanations, which prioritized functional information at the expense of other visually salient information, were less informative to naive viewers about the identity of the target machine than depictions were.

Experiment 2B: Using drawing to plan interventions

In Experiment 2B, our goal was to test the hypothesis that greater emphasis on functional elements, especially those that were causally relevant, would make it easier to infer which component to intervene on to activate the light bulb. If this allocation of visual information supports successful causal intervention, we predicted that naive viewers would be faster and more accurate in an inference task with visual explanations than with visual depictions.

Participants 633 participants (210 male; $M_{age} = 28.4$ years) were recruited from Prolific (N=99) and the study participant pool at our institution (N=526).[1] 8 data sessions were excluded for technical problems with displaying the experimental stimuli (e.g., videos did not load).

Stimuli & Procedure Participants were presented with a set of 6 drawings drawn by participants in the visual production experiment, one of each machine, in a randomized sequence. On every trial, participants saw a horizontal array of three images, which appeared in succession: *first*, a natural photograph of one machine appeared on the left; *second*, after a 3s delay, a drawing from the visual production experiment appeared in the middle; *third*, after another 3s delay, a semantically segmented photograph of the same machine appeared on the right, where one causal element and one non-causal element were enumerated and highlighted in different colors (Fig. 4A). Participants were instructed to press a key (i.e., 0 or 1) to indicate which of the highlighted elements they would intervene on to activate the light bulb, and to do so as quickly and accurately as possible. At the beginning of the session, participants completed a series of practice trials in which they were familiarized with the timing and keyboard interface.

Results

To investigate how well these drawings supported causal inference about how to operate the machines above chance,

[1]We increased our sample size to acquire more responses to each sketch. In the causal intervention experiment, participants responded to 6 sketches and there were on average 12.5 responses per sketch; in the identification experiment, participants responded to all 300 sketches and there were on average 38.7 responses per sketch.

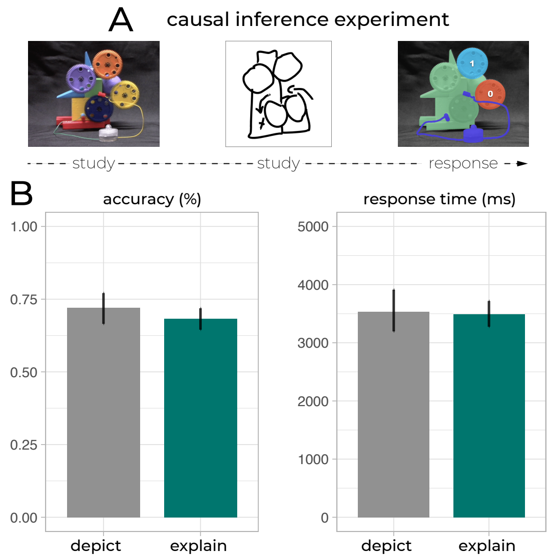

Figure 4: Causal inference task. (A) Example task display. Participants indicated which part of the machine they should intervene on to activate the light based on their interpretation of the target drawing. (B) Accuracy and response time. Error bars represent 95% CIs.

we fit a null model predicting identification accuracy that included random intercepts for different production participants. We found that both visual explanations and depictions supported intervention-task performance above chance (explanation: $b = 0.834$, $z = 10.36$, $p < 2e-16$; depiction: $b = 0.918$, $z = 13.32$, $p < 2e-16$; Fig. 4B, *left*). Next, to evaluate our primary hypothesis concerning differences between conditions, we used a linear model to predict response time from condition, as well as additional predictors controlling for the number of each type of stroke within a drawing (i.e., causal, non-causal, background, symbol) and the interaction between condition and the number of each type of stroke. We found that participants took a similar amount of time to make their response (correct trials only: explanation: 3490ms; depiction: 3536ms, $b = -2.15e-02$, $t = -0.312$, $p = 0.755$; Fig. 4B, *right*). We then used the same linear mixed-effects model to predict accuracy as in our identification experiment. Strikingly, we found that participants were actually less accurate when cued with a visual explanation than with a depiction (explanation: 68%; depiction: 72%, $b = -0.564$, $z = -2.826$, $p = 0.005$; Fig. 4B, *left*), suggesting that greater emphasis on causal elements and the usage of symbols in explanatory drawings did not necessarily help naive viewers locate the causally relevant element, at least relative to depictions.

We next sought to explore this surprising result further and better understand the factors that explain how well a given drawing supported the causal inference task, regardless of condition. To do this, we decomposed each drawing into its semantic components (i.e., causal, non-causal, background, symbol) and measured the association between the type of information emphasized in each drawing and how well it supported accurate responding in the inference task. Specifically, we represented each drawing by a 4-element feature vector, where each element in the vector represents the proportion of strokes corresponding to each semantic category.

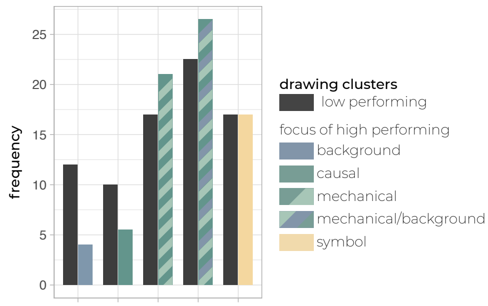

Figure 5: Drawing clusters supporting high and low success on the causal inference experiment. Performance was split by median accuracy (75% accuracy across conditions).

We then applied *k*-means clustering over all feature vectors to group drawings into five clusters that shared similar semantic properties: (1) background-focused; (2) causal-focused; (3) mechanical-focused (i.e., causal and non-causal functional elements); (4) mechanical/background-focused (i.e., included functional and structural elements); (5) symbol-focused. For each of these clusters, we computed how often each drawing in that cluster supported more-accurate responding (i.e., above-median accuracy) vs. less-accurate responding (i.e., below-median accuracy; Fig. 5). These analyses tentatively indicate that drawings containing both functional and structural information (i.e., "causal-focused" and "mechanical-focused") were more common and higher-performing than drawings that emphasized primarily structural information (i.e., "background-focused") or functional information (i.e., "causal-focused", "symbol-focused").

Discussion

How do we share our knowledge with others? Visualizations predate records of written language by thousands of years (Clottes, 2008) and are ubiquitous across cultures (Gombrich, 1989). Yet, despite the ancient and pervasive role of visual representations throughout human history, relatively little research has formally investigated the mechanisms that enable humans to communicate abstract knowledge to others in the form of visual explanations. In this paper, we developed a novel drawing paradigm to investigate how people prioritize information when producing visual explanations of simple mechanical systems. We found that visual explanations: (1) placed greater emphasis on the parts of machines that were most causally relevant for their operation, (2) placed less emphasis on visually salient, but structural features, and (3) were more likely to include symbols (e.g., arrows, motion lines) than depictions. These findings replicate those in the literature on verbal explanations, in which explaining tends to reduce attention to perceptually salient but irrelevant features (Legare & Lombrozo, 2014) and draw upon individuals' prior beliefs about causality (Lombrozo, 2006). This offers preliminary support that the cognitive mechanisms leveraged in explanation may apply to both verbal and visual domains. However, unlike prior studies that have relied on broad and holistic classifications of different types of explanations, the current approach allowed us to obtain detailed information about how every element of visual explanations corresponded to structural and functional elements of the target system.

Another critical contribution of our work is providing a quantitative account of the behavioral consequences of the semantic content included in visual explanations. The current findings build on extensive research about visualization that has demonstrated how drawing may increase learning about mechanical systems relative to written explanations (Alesandrini & Rigney, 1981; Bobek & Tversky, 2016; Lesgold, Levin, Shimron, & Guttmann, 1975), increase attention to the physical properties of biological systems (Dirnberger, McCullagh, & Howick, 2005; Leslie, 1980), and provide sketchers the opportunity to identify gaps in their knowledge (Van Meter, Aleksic, Schwartz, & Garner, 2006). While this prior work demonstrates that drawing may increase learning in certain contexts, visualization researchers have yet to identify the specific properties of those drawings that best support that learning. In the current work, we measure the utility of the drawings produced to convey causal knowledge by testing how well naive viewers could *use* these drawings to either identify the original machines or infer how to operate them. We found that while visual explanations and depictions were functionally distinct in their ability to support accurate identification, some of the communicative strategies that participants used to produce explanatory drawings did not always translate to better understanding about causal intervention. These results thus shed light on how the production of visual explanations guides subsequent interpretation by naive viewers, enabling us to map the correspondence between production and evaluation of explanation in the visual domain.

One major question raised by our results concerns how a sketcher's specific explanatory goals constrains their visual production behavior. In our drawing paradigm, participants were prompted to produce visual explanations to represent how the machines functioned. However, sketchers may have interpreted this prompt differently: Representing how a machine "functions" might be interpreted as a prompt to explain the causal mechanism (e.g., how the gears work), or to explain the procedure that will successfully activate the light (e.g., which gear to turn). This ambiguity may explain why visual explanations did not outperform depictions on the intervention task. Future work will aim to clarify the differences in semantic content between visual explanations of mechanistic versus procedural causal knowledge.

In sum, our paper reveals novel insights about how communicative goals constrain the production of visual explanations and what features make them most effective across different behavioral contexts. Ultimately, insights from such studies may lead to the creation of improved visual communication tools, as well as a deeper understanding of how we encode and explain abstract knowledge in visual form.

Acknowledgments

Many thanks to the members of the Cognitive Tools Lab and Early Learning & Cognition Lab at UC San Diego for helpful discussion. This work was supported by an NSF CAREER Award #2047191 to J.E.F.

All code and materials available at:
`https://github.com/cogtoolslab/causaldraw_cogsci2021`

References

Alesandrini, K. L., & Rigney, J. W. (1981). Pictorial presentation and review strategies in science learning. *Journal of Research in Science Teaching*, *18*(5), 465–474.

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications*, *1*(1), 27.

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.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(Supplement 2), 10918–10925.

Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Clottes, J. (2008). *Cave art*. London: Phaidon.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.

Dirnberger, J. M., McCullagh, S., & Howick, T. (2005). Wiriting & drawing in the naturalist's journal. *The Science Teacher*, *72*(1), 38.

Eitz, M., Hays, J., & Alexa, M. (2012). How do humans sketch objects? *ACM Transactions on Graphics*, *31*(4), 1–10.

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive Science*, *42*(8), 2670–2698.

Fonseca, B. A., & Chi, M. T. (2011). Instruction based on self-explanation. *Handbook of Research on Learning and Instruction*, 296–321.

Gombrich, E. (1989). *Art and illusion: A study in the psychology of pictorial representation.* Princeton, NJ: Princeton University Press.

Hegarty, M., Carpenter, P. A., & Just, M. A. (1991). 23 diagrams in the comprehension of scientific texts. *Synthesis*, *3*, 293–336.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, *30*(3), 581–592.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 856.

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology*, *126*, 198–212.

Lesgold, A. M., Levin, J. R., Shimron, J., & Guttmann, J. (1975). Pictures and young children's learning from oral prose. *Journal of Educational Psychology*, *67*(5), 636.

Leslie, C. W. (1980). Nature drawing: A tool for learning. *Dubuque, IA: Kendall Hunt.*.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470.

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

Mukherjee, K., Hawkins, R. X., & Fan, J. W. (2019). Communicating semantic part information in drawings. In *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 2413–2419).

Sayim, B., & Cavanagh, P. (2011). What line drawings reveal about the visual brain. *Frontiers in Human Neuroscience*, *5*, 118.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*(3), 453–489.

Tversky, B. (2015). The cognitive design of tools of thought. *Review of Philosophy and Psychology*, *6*(1), 99–116.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*(2), 142–166.

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

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*(5), 776–806.