Visual explanations prioritize functional properties at the expense of visual fidelity

Holly Huey, Xuanchen Lu, Caren M. Walker, and Judith E. Fan

Department of Psychology

University of California San Diego

Author Note

Holly Huey https://orcid.org/0000-0002-6522-6962

Judith E. Fan https://orcid.org/0000-0002-0097-3254

We have no known conflict of interest to disclose.

Word count: 6982

All experimental materials, study preregistrations, data, and analysis code are publicly available in our GitHub repository: https://github.com/cogtoolslab/causaldraw_public2021.

Correspondence concerning this article should be addressed to Judith E. Fan, Department of Psychology, University of California, San Diego, La Jolla, CA 92093. E-mail: jefan@ucsd.edu

Abstract

Visual explanations play an integral role in communicating mechanistic knowledge about how things work. What do people think distinguishes such pictures from those that are intended to convey how things look? To explore this question, we used a drawing paradigm to elicit both visual explanations and depictions of novel machine-like objects, then conducted a detailed analysis of the semantic information conveyed in each drawing. We found that visual explanations placed greater emphasis on parts of the machines that move or interact to produce an effect, while visual depictions emphasized parts that were visually salient, even if they were static. Moreover, we found that these differences in visual emphasis impacted what information naive viewers could extract from these drawings: explanations made it easier to infer which action was needed to operate the machine, but more difficult to identify which machine it represented. Taken together, our findings suggest that people spontaneously prioritize functional information when producing visual explanations but that this strategy may be double-edged, facilitating inferences about physical mechanism at the expense of preserving visual fidelity.

Keywords: natural pedagogy, causal learning, explanation, visual production

Visual explanations prioritize functional properties at the expense of visual fidelity

From infants exploring the objects in their immediate environment to scientists exploring the frontiers of our solar system, humans are driven to understand how things work and use that knowledge to generate desired outcomes. However, acquiring such mechanistic knowledge from firsthand experience can often be costly in time and effort (Lagnado & Sloman, 2004; Steyvers et al., 2003) and thus the majority of our knowledge about the world depends on its faithful transmission from one generation to another (Boyd et al., 2011; Csibra & Gergely, 2009). This knowledge transmission has long been supported by mechanistic explanations, which help to expose causal relationships latent in otherwise fleeting and complex information (Keil & Lockhart, 2021).

What characterizes good mechanistic explanations, and how do they relate to the phenomena they are intended to explain? Prominent theoretical perspectives highlight several hallmark features (Bechtel, 2011; Wimsatt, 1976), noting that effective mechanistic explanations decompose a causal system into its interacting parts and specify the causal relationships between those parts in the context of a particular function. For example, a bicycle functions by transferring power from the movement of the pedals to the drive wheel via the roller chain between the two wheels, propelling the entire bicycle forward. Such an explanation can be distinguished from a merely descriptive report (e.g., “a bicycle has two wheels, pedals, and a chain), which does not specify the causal relationship between the interacting parts (Corriveau & Kurkul, 2014), and from a teleological explanation (e.g., “a bicycle is for riding from one location to another), which does not decompose the causal system into any constituent parts nor specify how they interact (Kelemen & Rosset, 2009). In addition to playing a key role in scientific theories (Bechtel, 2009), there is growing evidence that mechanistic explanations are also privileged in people’s intuitive understanding of artifacts and biological entities (Chuey et al., 2020; Lockhart et al., 2019). Nevertheless, our understanding of what intuitions people have about what information to prioritize when producing mechanistic explanations themselves is less well developed. Initial insights may be gleaned from prior work investigating the content of explanations that people

produce while studying a physical system, which has documented the inclusion of abstract principles (Chi & VanLehn, 1991) and the notion that some explanations may prioritize outward appearance while others emphasize internal properties (Walker et al., 2014; Walker et al., 2017). However, these analyses have generally lacked the resolution to tease apart different hypotheses concerning how people weigh these different kinds of information when constructing a coherent explanation.

While the majority of prior studies investigating explanation behavior have focused on verbal explanations (Chi & VanLehn, 1991; Legare & Lombrozo, 2014; Lombrozo, 2016; Walker et al., 2014; Walker et al., 2017), explanatory *visualizations* may be especially useful for probing the cognitive processes engaged during the communication of mechanistic knowledge (Hegarty, 2011; Mayer, 1999; Scaife & Rogers, 1996; Tversky, 2005). Visualizations of mechanistic phenomena play an important role across scientific domains, including in the biological (Callaway, 2016) and physical sciences (Lipşa et al., 2012). They naturally exploit shape-based and spatial cues to expose both the relevant part-based and relational abstractions that underlie mechanistic understanding (Forbus et al., 2011; Hegarty & Just, 1993; Hegarty et al., 2003; Tversky, 2001), as well as how these abstractions map back onto physical parts of the target system (Bobek & Tversky, 2016; Fan, 2015; Gobert & Clement, 1999; Newcombe, 2013). Moreover, there is ample evidence that visualizations can facilitate learning and inference by comparison with text alone (Glenberg & Langston, 1992; Hegarty & Just, 1993; Larkin & Simon, 1987; Mayer, 1989) by leveraging a small set of relational symbols, such as lines and arrows (Heiser & Tversky, 2006; Tversky, 2005; Tversky et al., 2002; Tversky et al., 2000). However, previous studies that have elicited visual explanations of mechanistic phenomena have not included the detailed analyses of their content that would be required to understand what distinguishes visual explanations in people's minds from other types of visualizations. In particular, while prior work has found that visualizations prompted by functional descriptions of a physical system contain more arrows than those cued by structural ones (Heiser & Tversky, 2006), it remains unclear whether these symbols were simply added to an otherwise ordinary

illustration, or whether they formed part of a distinct type of visualization emphasizing information in a substantially different way.

The current studies aim to overcome key limitations of prior work by conducting a thorough investigation of what information people prioritize when generating mechanistic explanations, and leveraging the distinctive properties of *visual* explanations to gain insight into how explanatory abstractions are grounded in our direct experience with mechanical systems. We elicited these visual explanations using an open-ended drawing task, following prior work (Heiser & Tversky, 2006). In Experiment 1, we measure how much people emphasize information about visual appearance or physical mechanisms when producing explanatory drawings of novel mechanical objects, as opposed to depictive illustrations. We used novel objects to probe people's intuitions about how to create informative explanations when generalizing to a specific mechanical system they were not already familiar with, while still being able to rely on prior knowledge about the types of physical mechanisms in play. In Experiment 2, we measure how well naive viewers can map such information back to the corresponding source object. Together, data from these two experiments help to distinguish two potential hypotheses concerning how people generate visual explanations. Under the *cumulative* hypothesis, people first produce a complete depiction of an object's parts, after which they augment this representation with symbols that convey how these parts interact. Under the *dissociable* hypothesis, people intending to communicate mechanistic knowledge refrain from drawing all the parts of the object, instead emphasizing the most relevant ones and how they interact, rather than preserving information about the object's overall appearance. Overall, our results were more consistent with the latter dissociable hypothesis: explanatory drawings emphasized different parts from depictions and more effectively communicated mechanistically relevant information to naive viewers, while less effectively conveying information about an object's visual appearance. Together, these findings suggest that people engaging in visual explanation spontaneously prioritize functional information (i.e., how parts move and interact) at the expense of visual fidelity (i.e., what parts look like and where they are), and thus that visual explanations are distinct from other kinds of

illustrations not only in terms of what information they contain, but also what they omit. As such, the balance of structural and functional information that characterizes visual explanations does not necessarily make it more useful in all contexts; instead, they may be better thought of as a tool for communicating knowledge at a specific level of abstraction. Moreover, these findings generalize prior work on verbal explanations to the visual modality, lending support to the notion that similar cognitive mechanisms may support explanatory behavior across communication modalities (Legare & Lombrozo, 2014; Walker et al., 2014).

Experiment 1A: Production of visual explanations and depictions

Our first goal was to identify the semantic properties that characterize visual explanations of mechanistic 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; 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 about the machines. 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). Additionally, 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 et al., 2012; Fan et al., 2018). In this experiment, we presented participants with simple machines composed of gears, levers, and pulleys. These parts were chosen since they were likely familiar to participants and are the basic components of more complex compound mechanical systems (Prater, 1994). By using these simple machines, our aim was to gain a purer measure of how people translate their high-level goals of either depicting how a machine functioned or what a machine looked like, without need for expertise in a domain or

otherwise extensive familiarity with our task.

Method

Participants

50 participants (29 male; mean age = 39.1 years) were recruited from Amazon Mechanical Turk for the visual production experiment. Two additional participants were recruited, but their data were not included in the study for not meeting our predefined exclusion criteria (e.g., the drawings consisted of scribbles or were otherwise uninterpretable). In this and all subsequent experiments, participants provided informed consent in accordance with the UC San Diego IRB.

Stimuli

We designed 6 novel machines composed from simple mechanical parts (i.e., gears, levers, pulleys). There were two machines employing each type of part. Half of the mechanical parts in each machine were *causal*, meaning they could be used to produce a desired effect (i.e., turn on a light bulb attached to each machine); the other half of mechanical parts were *non-causal*. To match how visually salient they were, the causal and non-causal parts within each machine were always of the same type (e.g., gear), and were approximately matched in size and number (Fig. 2, left). For each machine, we produced a video demonstration of it in which a demonstrator's hand was shown to interact with both the causal and non-causal mechanical parts twice each, in a counterbalanced order, to show that the causal part reliably turned on the light, whereas the non-causal part did not. The order of manipulation was counterbalanced across all machines for a total of 12 video demonstrations. Each video was 30 seconds, and the duration of time in which the researcher manipulated each causal and non-causal part was controlled for through post-production video editing. We also conducted a separate validation study to ensure that participants could generally determine how the machines could be operated to activate the light bulb based on these video demonstrations (see Supplementary Materials).

Procedure

We presented a naive group of participants with a series of 6 videos (one of each machine). After each video finished playing, participants were cued to produce one of two kinds of drawings: on *explanation* trials, they were prompted to produce visual explanations intended to help a naive viewer learn how the machine could be operated to activate a light bulb; on baseline *depiction* trials, they were prompted to produce visual depictions intended to help a naive viewer identify the machine by its appearance (Fig. 1). All participants produced three visual explanations and three visual depictions, in a randomized sequence, such that they drew one of each type of drawing for each type of machine. 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). While drawing on a digital canvas may be more effortful for some participants than drawing on paper, our approach is motivated by prior work that has successfully used digital drawing interfaces to reliably measure variation in drawing production (Bainbridge et al., 2019; Fan et al., 2020; Fan et al., 2018; R. X. Hawkins et al., 2019). Each stroke was rendered in real time on the participant's screen as they drew and could not be deleted once drawn, approximating key aspects of drawing with an ink pen on paper. We reasoned that while it was possible that preventing participants from deleting individual strokes might lead to drawings that sometimes contained extraneous details or strokes produced accidentally, there was no reason to believe that this aspect of the drawing interface would impact one more condition than the other. Participants were not limited in amount of time that they could spend drawing in each trial. At the beginning of each session, participants also completed two practice trials to familiarize themselves with the drawing interface.

Results & Discussion

The resulting dataset contained 300 drawings from 50 unique participants: 150 visual explanations and 150 depictions (Fig. 2). Insofar as participants are predicted to include more information in visual explanations in accordance with the *cumulative hypothesis*, we predicted that visual explanations would contain more visual detail and take more time to produce, relative

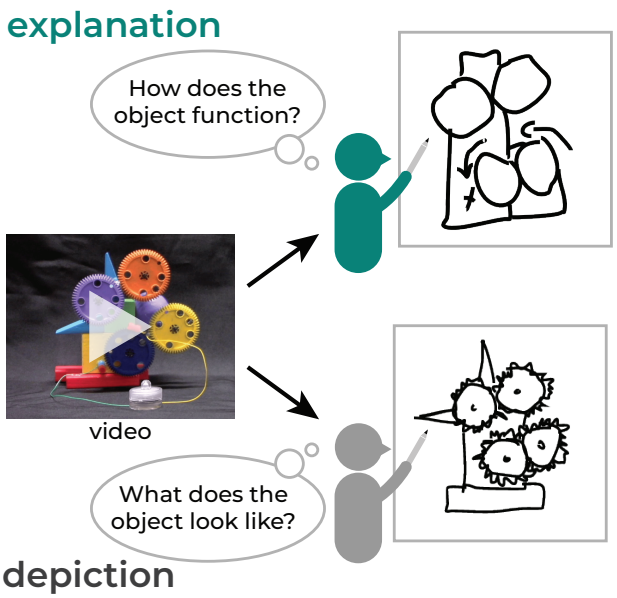

Figure 1

Study 1: Visual Production Task. *On each trial, participants viewed a 30-second video demonstrating how to operate a machine to turn on a light bulb. On half of the trials, after the video finished playing, participants were then prompted to produce an explanatory drawing. On the other half of the trials, they were prompted to produce a depictive drawing.*

to visual depictions. On the other hand, if participants invest a similar amount of effort in both conditions but differ in their semantic content as predicted by the *dissociable hypothesis*, we predicted that the two types of drawings would not substantially differ in how detailed they were nor how much time they took to be produced. To distinguish these possibilities, we analyzed the number of strokes and total drawing time using a linear mixed-effects model predicting the number of strokes from condition and included random intercepts for the type of machine (e.g., gear, lever, pulley) and individual participant.

We found that participants used a similar number of strokes (explanation: 20.33; depiction: 18.9; $b = 1.44$, $t = 1.04$, $p = 0.301$; Fig. 1B, *left*) and amount of time drawing in both conditions (explanation: 59300ms; depiction: 57689ms; $b = 1144.75$, $t = 0.359$, $p = 0.72$), suggesting that participants had invested a similar degree of effort when producing both types of

drawings. However, while these results provide preliminary evidence against the *cumulative* hypothesis, they also indicate that such simple effort-based measures are insufficient to capture differences in the *semantic information* conveyed by each type of drawing.

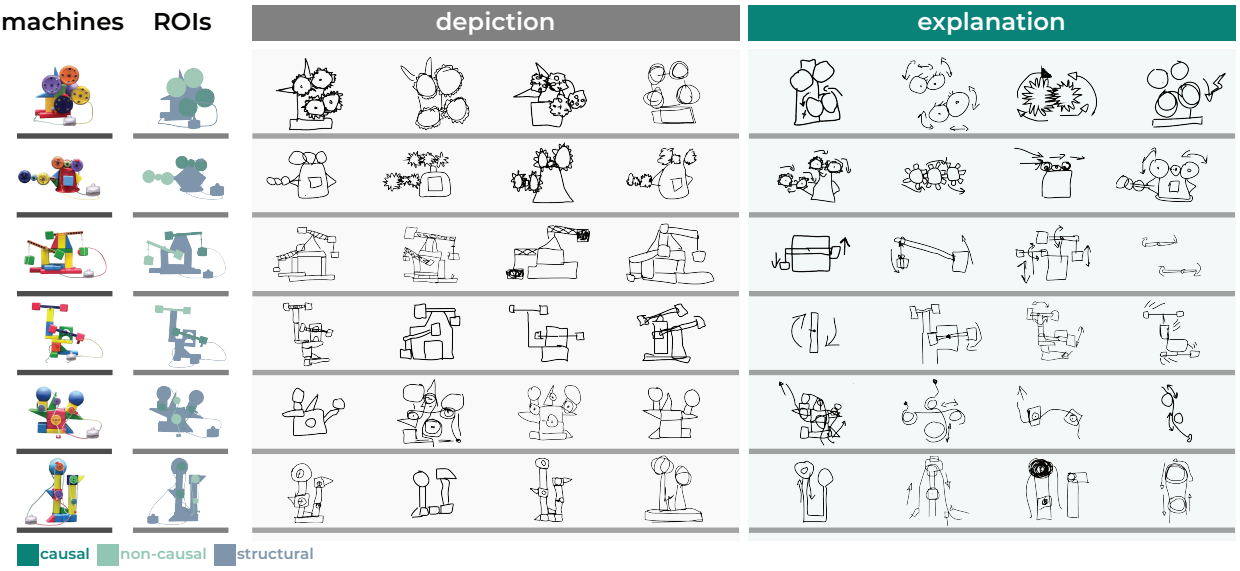

Figure 2

Study 1: Visual Production Dataset. Left: *Each machine consisted of multiple mechanical and structural parts. Each region-of-interest (ROI) image indicates the location of both causally relevant and non-causally relevant mechanical parts.* ***Right:*** *Example depictive and explanatory drawings.*

Experiment 1B: Characterizing semantic content in visual explanations and depictions

To go beyond these effort-based measures, we next crowdsourced annotations from a separate group of naive participant in order to systematically characterize the *semantic information* contained in these drawings. We used these annotations in two ways: first, to understand which parts of the machine participants in Experiment 1A had thought relevant to include in their drawing; and second, to quantify the degree to which each drawing faithfully preserved the relative size and location of each part. One possibility is that visual explanations focus on causally relevant parts, but still faithfully preserve their visual properties. Alternatively, they may distort their visual properties, for example, by making these causally relevant parts more visually salient in their drawing. To distinguish these possibilities, we leveraged techniques from computer vision to precisely measure the differences in the apparent size and location of each drawn part and its actual size and location in the target machine.

Method

Participants

252 participants (210 male; M_{age} = 38.9 years) were recruited from Amazon Mechanical Turk to provide semantic annotations of the drawings produced in Experiment 1A. We excluded data from 28 additional participants, who did not meet our preregistered inclusion criteria (i.e., low accuracy on attention-check trials, response time <5s).

Task Procedure

Annotators were presented with a set of 10 drawings that were randomly sampled from those drawn in the visual production experiment, as well as reference color photographs of the original machines. In these photographs, each part was color-coded and assigned a unique label and numerical identifier (e.g., ‘Gear 2’). Annotators were asked to tag each pen stroke in the drawing based on which part they thought it represented. If a stroke depicted a symbol (e.g., arrow, motion line) rather than a physical part of the machine, annotators were asked to additionally label which part(s) the symbol referred to. If a stroke’s meaning was not clear, they

could select an “I don’t know” option instead. Annotators also completed one attention-check trial that used a drawing from the Experiment 1A dataset that was particularly straightforward to parse and had been manually segmented by the authors. If annotators made 3 or more errors when labeling strokes in the attention-check drawing, all data from that session were excluded from subsequent analysis.

Preprocessing annotation data

For each stroke in every drawing, we obtained labels from at least three annotators indicating which part of the machine it corresponded to (e.g., “gear”, “lever”, “structural”). Each of these labels were then further grouped into higher-level semantic categories: *causal* strokes representing mechanical parts that were causally related to turning on the light bulb, *non-causal* strokes representing mechanical parts that were not causally related to turning on the light bulb, *structural* strokes representing structural parts, and *symbolic* strokes, including arrows and other marks indicating motion and interactions between parts.

We found that 64.9% of strokes received the same label by all three annotators, and 95.0% of strokes received the same label by at least two of the three annotators. 5.0% of strokes did not reach a majority consensus and received more annotations to resolve this conflict. Moreover, within visual explanations, 55.5% of strokes received the same label by all three annotators, and 93.2% of strokes received the same label by at least two of the three annotators. Within depictions, 75.0% of strokes in depictions received the same label by all three annotators, and 96.9% of strokes received the same label by at least two of the three annotators. In subsequent analyses, we collapsed across annotators and assigned the modal label to strokes which had been given the same label by at least two annotators. For the remaining strokes that did not receive a modal label, we randomly sampled an annotation from the set of annotations that had been assigned to it. We also excluded 5 drawings from subsequent analyses that were deemed to be entirely uninterpretable.

Spatial error analysis

To evaluate how accurately the drawings preserved information about the location and size of each part, we used the following procedure. First, to compute the size and location of drawn parts, we grouped all strokes within a drawing that were tagged with the same semantic label, then determined the coordinates of the rectangular bounding box containing those parts (Fig. 3B). For example, if a drawing contained strokes representing four different gears and some structural parts, then this step would yield five bounding boxes, one for each gear, and the fifth containing all structural parts. Strokes representing symbols and/or the light bulb were excluded from analysis. Next, to compute the size and location of target parts, we color-coded each part of the still images of the machines in Adobe Photoshop and grouped all the pixels of the same color. We then calculated the coordinates of the individual bounding boxes for each part. Because the goal of our analysis was to measure how accurately drawings preserved *relative* size and location information, we aligned each drawing to its target machine before computing size and location errors. Specifically, we defined the bounding box containing the entire drawing and the bounding box of the target machine containing the entire machine in the still image, then applied the translation and scaling transformations needed to align these two bounding boxes.

To calculate raw location error for a given part, we computed the Euclidean distance between the centroid of the bounding box for each drawn part and the centroid of the bounding box for the target part. The raw location error for the drawing as a whole was computed by taking the mean of these distances across all parts that appeared in the drawing. We then divided this raw location error by the length of the diagonal of the machine's bounding box to derive a normalized measure of location error, enabling more straightforward aggregation of location error estimates between machines of different sizes. Here, a value of zero indicates that the centroid of a part was drawn exactly in the same location as the centroid of the target part. Additionally, to calculate the raw size error for a given part, we calculated the absolute value of the difference in area between the bounding box of the drawn part and the bounding box for the target part. We then normalized this raw size error by dividing it by the area of the target part, making it easier to aggregate size

error estimates between parts of different sizes. Under this procedure, a value of zero indicates that the size of the drawn part exactly matched the size of the target part. And any deviation in size between drawn and target parts increased normalized size error, regardless of whether the drawn part was larger than or smaller than the target part. The normalized size error for a drawing as a whole was computed by taking the mean across all parts that appeared in the drawing.

Results & Discussion

Insofar as visual explanations place a greater emphasis on functional information than depictions do in accordance with the *dissociable hypothesis*, we hypothesized that visual explanations would contain: (1) more strokes representing causally relevant parts than non-causally relevant parts and (2) more strokes devoted to conveying movement and interactions between parts, such arrows and other symbols, rather than to representing the physical parts themselves. To evaluate the first hypothesis, we constructed a linear mixed-effects model predicting the number of strokes labeled as "causal" from condition and included random intercepts for individual drawing and individual participant. To evaluate the second hypothesis, we constructed a linear mixed-effects model predicting the number of strokes labeled as "symbol" from condition and included random intercepts for the type of machine (e.g., gear, lever, pulley) and individual participants.

Consistent with the first hypothesis, we found that among strokes representing a mechanical part (i.e., gear, lever, or pulley), a greater proportion were devoted to representing causal parts in visual explanations than in depictions (explanation: 58.0%, depiction: 42.0%, b = 0.382, z = 3.44, $p = 5.9e-4$; Fig. 3A). Consistent with the second hypothesis, a higher proportion of strokes in visual explanations were classified as symbols than in depictions (explanation: 24.8%, depiction: 1.0%, b = 2.48, t = 1.39, $p = 1.67e-1$) and a lower proportion of strokes represented physical parts, including both causal and non-causal parts (explanation: 25.0%, depiction: 45.8%, $b = -2.77$, $t = -4.86$, $p = 1.31e-5$). These results suggest that the goal of communicating mechanistic knowledge leads people to produce drawings that place greater emphasis on causally relevant components and how they move, and less emphasis on static

components, even if they are visually salient.

These results are consistent with findings from prior work (Heiser & Tversky, 2006) that has documented an association between drawings explaining how mechanical systems work and the inclusion of arrows. However, this earlier work could not tease apart the degree to which these arrows were simply added to otherwise ordinary depictive drawings ("cumulative" hypothesis), or whether the inclusion of these arrows was accompanied by a general increase in relative emphasis on causally relevant information by comparison with other visually salient, but non-causally relevant information ("dissociable" hypothesis). By collecting detailed semantic annotations of the elements represented in both kinds of drawings, our current findings go beyond prior work to provide direct support for the latter hypothesis.

Insofar as visual explanations exaggerate the appearance of important parts of each machine, we hypothesized that they would not preserve information about their relative sizes and locations as accurately as visual depictions do. Specifically, we predicted that: (1) visual explanations might exaggerate the size of causally relevant parts to make them more salient to the viewer and (2) visual explanations might not preserve information about the relative locations of parts, insofar as such information is deemed less relevant for communicating about causal interactions between parts. To evaluate this hypothesis, we fit a linear mixed-effects model predicting the size and location error from condition, including random intercepts of individual machine and participant. We found that mechanical parts were consistently drawn larger in visual explanations than in depictions (explanation: 72.4px, depiction: 60.8px, $b = 8.41$, $t = 1.97$, $p = 4.96e-2$), in addition to being drawn somewhat further from their actual locations, relative to other parts of the machine (explanation: 75.1px, depiction: 62.3px, $b = 11.6$, $t = 3.15$, $p = 0.18e-2$; Fig. 3C). These findings are consistent with the notion that when explaining how a machine functions, people distort the appearance of functionally relevant parts to make them more salient and discount the importance of preserving exact spatial relationships. Taken together, Experiments 1A and 1B provide evidence that having the goal of communicating mechanistic knowledge systematically affects the kind of information people prioritize when

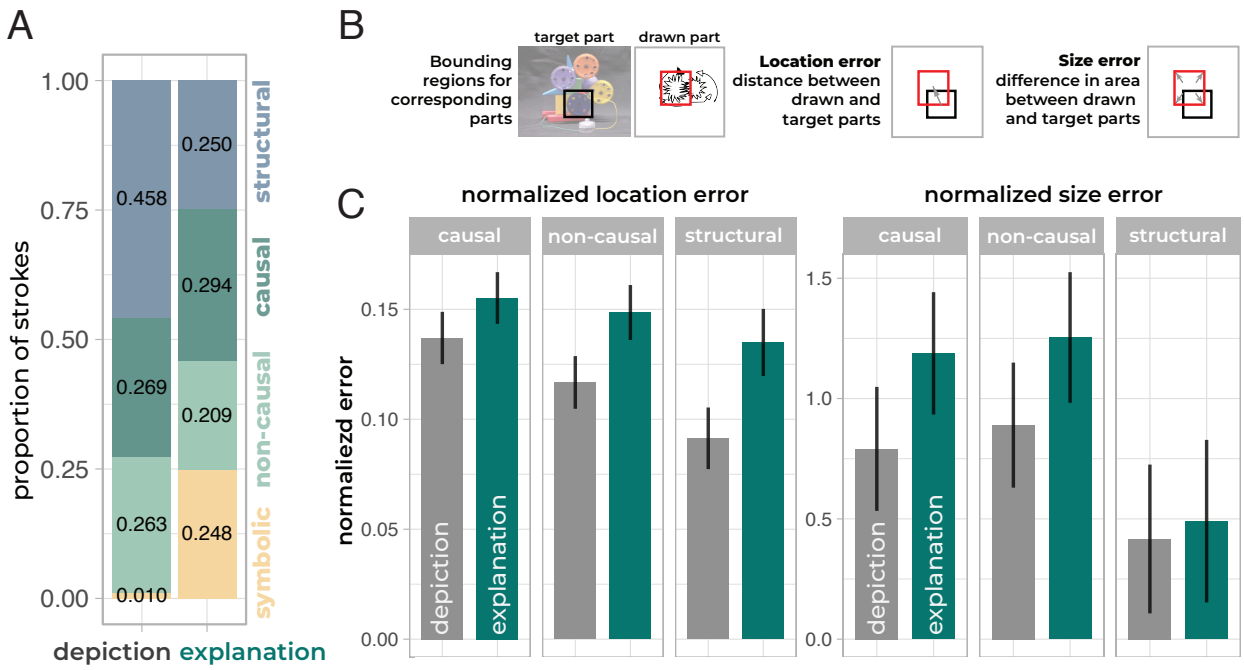

Figure 3

Study 1: Results. A: *Proportion of strokes conveying different semantic information: causal strokes representing mechanical parts that turned the light on; non-causal strokes representing mechanical parts that did not; structural strokes representing static parts; and symbolic strokes, including arrows and other marks indicating motion and interactions between parts.* ***B:*** *Accuracy of spatial information in drawings was estimated by defining bounding regions for corresponding parts in each drawing and video, then computing the difference in size and location between the drawn and target parts.* ***C:*** *Normalized location and size errors for different semantic part categories. Normalized location errors reflect relative differences between the target and drawn parts, rescaled by the size of the machine. When the normalized location error is zero, the relative locations of each drawn part exactly match the relative locations of each part of the target machine. Normalized size errors reflect relative differences between the target and drawn parts, rescaled by the size of the target part. When the normalized error for size is equal to zero, the relative sizes of each drawn part exactly match the relative sizes of each part of the target machine. Error bars represent 95% CIs.*

producing visual explanations.

Experiment 2A: Object identification

However, a critical test of how *useful* such communicative strategies are can be measured how well other people can interpret these drawings to achieve their own behavioral goals. In Experiment 2, we recruited three additional cohorts of naive participants to view the drawings made in the visual production experiment (Experiment 1A) and measured how well each drawing supported their ability to identify the original machine (Experiment 2A), to infer which part of the machine to intervene on to operate it (Experiment 2B), or to infer which action was needed to operate the machine to activate the light (Experiment 2C).

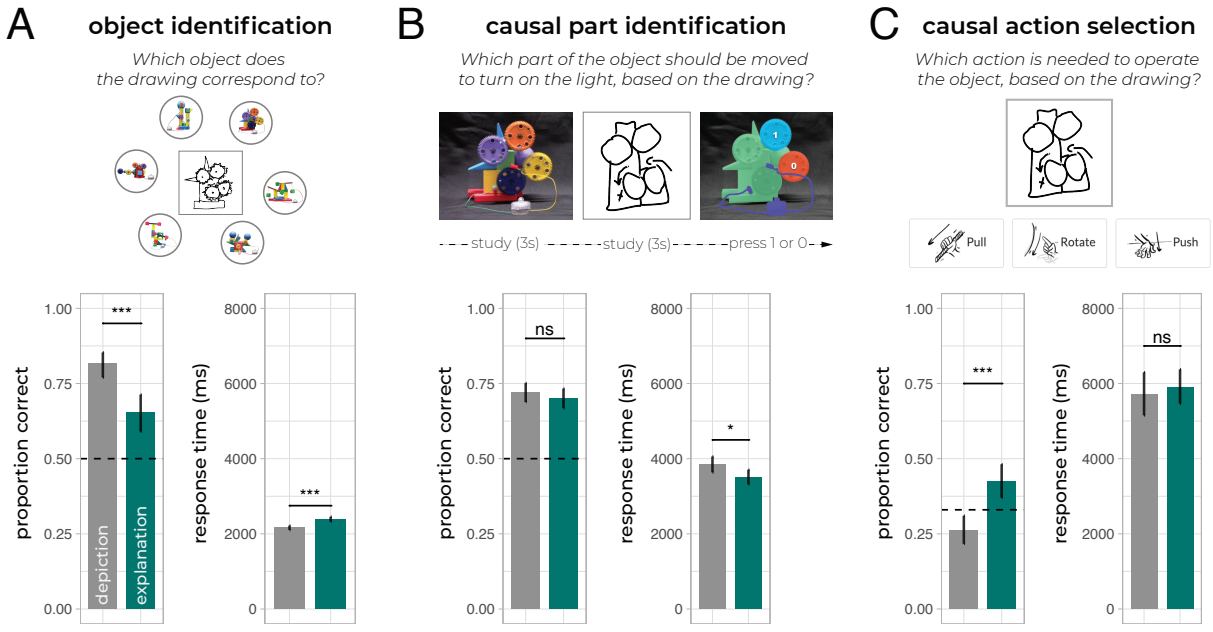

Figure 4

Study 2: Visual Inference Tasks and Results. A: *In Study 2A, participants identified the machine that matched each drawing.* ***B:*** *In Study 2B, participants identified which part of the machine they should intervene on to turned on the light bulb.* ***C:*** *In Study 2C, participants inferred which action they would need to perform to turn on the light. Error bars represent 95% CIs.*

In Experiment 2A, we hypothesized that the reduced emphasis on structural parts in visual

explanations, based on there being relatively fewer strokes devoted to representing them, would make it harder to match it to the original machine, relative to visual depictions. To test this hypothesis, we designed a visual search task to probe how quickly and accurately naive viewers could identify the machine that corresponded to each drawing.

Method

Participants

50 participants (24 male; M_{age} = 20.5 years) were recruited from the UC San Diego study pool. Two additional participants were recruited, but data from their sessions were excluded for technical problems (i.e., inability to click on images).

Task Procedure

Each participant was presented with all 300 drawings from Experiment 1A in a randomized sequence. At the beginning of each trial, participants moved their cursor to a crosshair displayed 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 (125px x 100px, radius = 250px), one of each machine (Fig. 4A). The angular distance between each photo was constant (i.e., 60 degrees) and their angular locations 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 & Discussion

To investigate how well these drawings support participants' ability to identify the machines, we fit a null model predicting identification accuracy that included random intercepts for different production participants. Although there were 6 machines, we defined chance-level performance at 50%, a theoretical upper bound reflecting our expectation that confusions would be most likely to arise between machines of the same type (e.g., gears).

To evaluate our hypothesis that participants would be slower when presented with visual explanations relative to when they were presented depictions, we fit a linear mixed-effects model predicting response time from condition and random intercepts for individual drawings and participants. Additionally, to evaluate our hypothesis that participants would be less accurate when viewing visual explanations rather than depictions, we fit a mixed-effects logistic regression model to predict individual trial outcomes, with the same random effects structure as our response-time model above.

We found that participants were reliably above chance performance for both types of drawings (explanation: $b = 0.561$, $z = 3.94$, $p = 8.16e-5$; depiction: $b = 1.28$, $z = 10.1$, $p = 2e-16$; Fig. 4A). We found that participants were slower to respond (correct trials only: explanation: 2387ms, 95% CI: [2321ms, 2455ms]; depiction: 2161ms, 95% CI: [2103ms, 2220ms]; $b = 9.96e-2$, $t = 5.90$, $p = 1.43e-8$; Fig. 4A) and were less accurate when cued with a visual explanation than with a depiction (explanation: 65.4%, 95% CI: [59.0%, 71.0%]; depiction: 81.5%, 95% CI: [77.0%, 85.0%]; $b = -0.847$, $z = -5.033$, $p = 4.84e-7$; Fig. 4A, *left*). These results suggest that our manipulation of communicative goals in Experiment 1A measurably impacted how well viewers could extract relevant information from each type of drawing, such that depictive drawings were more informative about the identity of the target machine.

Experiment 2B: Causal part identification

How well do visual explanations support naive viewers' ability to identify which part of the machines to intervene on to produce desired goals? In Experiment 2B, we hypothesized that greater emphasis on functional parts (i.e., additional strokes, drawn larger), especially those that were causally relevant, would make it easier for learners to infer which component to intervene on to activate the light bulb. To test this hypothesis, we designed another visual search task that probed how quickly and accurately naive viewers could locate the causally relevant part when provided with a drawing of the machine.

Method

Participants

297 participants (100 male; M_{age} = 28.4 years) were recruited from Prolific (N=99) and the study participant pool at UCSD (N=198). 8 additional participants were recruited but data from their sessions were excluded, for technical problems with displaying the experimental stimuli (e.g., the videos did not load). We used a larger sample size in Experiment 2B to collect approximately the same number of observations per drawing as we had collected in Experiment 2A.

Task Procedure

Participants were presented with a randomly sampled set of 6 drawings from Experiment 1A, one of each machine, in a randomized sequence. On every trial, participants were presented with three images laid out in a horizontal array, appearing in succession: first, a color photograph of one of the machines appeared on the left; second, after a 3-second delay, a drawing of it appeared in the middle; and third, after another 3-second delay, another photograph of the same machine appeared on the right, this time with one causal part and one non-causal part highlighted in different colors (Fig. 4B). The rationale for sequencing the presentation of these three images in this manner was to provide an approximation to the scenario wherein a person first encounters a novel device, then consults a diagram to make sense of how the device works, before turning back to the device to interact with it. Participants were instructed to press a key (i.e., either 0 or 1) to indicate which of the highlighted parts they would intervene on to turn on the light, 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 task interface.

Results & Discussion

As in Experiment 2A, we fit a null model predicting identification accuracy that included random intercepts for different production participants to evaluate the degree to which participants performed above chance. To evaluate whether participants would be faster in identifying the

causal part when presented with a visual explanation rather than a depiction, we constructed a linear mixed-effects model to predict response time from condition and random intercepts for individual drawings and participants. Additionally, to evaluate our hypothesis that participants would be more accurate for explanations than depictions, we fit a mixed-effects logistic regression model to predict response time from condition and random intercepts for participants. Unlike in Experiment 2A, where each participant saw all drawings produced in Experiment 1A, participants in Experiment 2B were only presented with 6 drawings per session, one of each machine. Given the smaller number of measurements obtained for each drawing in Experiment 2B, we did not have sufficient data to include random intercepts for individual drawings in our statistical models.

We found that both types of drawings supported above-chance performance (explanation: $b = 0.849$, $t = 10.53$, $p = 2e-16$; depiction: $b = 0.919$, $z = 13.04$, $p = 2e-16$; Fig. 4B), suggesting that both types of drawings carried meaningful signal about the identity of the causally relevant parts. However, we found that participants were no more or less accurate when cued with a visual explanation than with a depiction (explanation: 70.16%, 95% CI: [67.0%, 73.0%]; depiction: 72.1%, 95% CI: [69.0%, 75.0%]; $b = -9.48e-2$, $z = -0.842$, $p = 0.4$; Fig. 4B, *left*). Nevertheless, we did find a small response-time advantage for visual explanations, such that participants were slightly faster to make their response when presented with an explanatory drawing rather than a depictive one (explanation: 3508ms, 95% CI: [3319ms, 3708ms]; depiction: 3840ms, 95% CI: [3635ms, 4057ms]; $b = -9.059e-2$, $t = -2.42$, $p = 1.57e-2$; Fig. 4B, *right*.) Taken together, these results do not provide unequivocal evidence that the greater visual emphasis on causal parts in explanatory drawings improved others' ability to more accurately identify these parts *in situ*. Indeed, such null effects raise the possibility that participants' judgments were not informed by the drawing at all, although the shorter response times for visual explanations relative to depictions suggest at least some effect of the drawing on how individuals produced their judgments. Overall, these findings instead suggest that there may be more to the construction of an effective visual explanation than displaying the most functionally important entities more prominently.

Experiment 2C: Causal action selection

While the prior experiment evaluated how well visual explanations supported naive viewers' ability to identify *where* to intervene on the machines, here we evaluated how well these drawings could support participants' ability to infer *how* to intervene on the machines. In other words, how well do visual explanations support naive viewers' ability to infer which action is needed to successfully operate the machines? Similar to Experiment 2B, we hypothesized in Experiment 2C that greater emphasis on functional parts, especially those that were causally relevant, would make it easier to infer which action was necessary to intervene on the machines to activate the light bulb. To test this hypothesis, we developed a task probing how quickly and accurately naive viewers could identify the appropriate action to perform when provided with a drawing of each machine.

Method

Participants

267 participants (75 male; M_{age} = 21.3 years) were recruited from the UC San Diego study pool. Three additional participants were recruited, but data from their sessions were excluded for technical problems (i.e., videos did not load).

Task Procedure

Participants were presented with a random set of 6 drawings from Experiment 1A, one of each machine, in randomized sequence. On each trial, participants were presented with a single drawing, under which there were 3 buttons labeled "Pull", "Push", "Rotate" and "I don't know" (Fig. 4C). Participants were instructed to click the button that corresponded to the action needed to operate the machine, based on their interpretation of the drawing, and were told to prioritize accuracy. At the beginning on of the session, participants completed a series of practice trials in which they were familiarized with the task interface.

Results & Discussion

To evaluate the degree to which participants performed the task above chance, we fit a null model predicting accurate responses that was identical in structure to that used in Experiment 2A and 2B. Next, to evaluate differences in how quickly participants could identify the correct action, we fit their responses using the same type of statistical model as in Experiment 2B. Additionally, to evaluate differences in how accurately participants could identify the correct action, we fit their responses using the same type of statistical model as in Experiment 2A and 2B.

We found that participants more accurately identified the correct action when cued with a visual explanation (chance = 33%; explanation: 42.5%, 95% CI: [37.0%, 48.0%]; depiction: 26.09%, 95% CI: [22.0%, 31.0%]; $b = 0.738$, $z = 4.34$, $p = 1.42e-5$; Fig. 4C, *left*). Between conditions, they took a similar amount of time to make their response (correct trials only, explanation: 5903ms, 95% CI: [5464ms, 6376ms]; depiction: 5696ms, 95% CI: [5152ms, 6298ms]; $b = 3.56e-2$, $t = 0.555$, $p = 0.579$; Fig. 4C, *right*), suggesting that the greater accuracy was unlikely to be due to a speed-accuracy tradeoff. Taken together with the results of Study 2B, these findings suggest that explanatory drawings better supported naive viewers' ability to figure out which action was needed to interact with the machine, even if they did not help them identify which part of the machine to interact with. More broadly, these results show that the visual differences between visual explanations and depictions that we measured in Experiment 1 lead to specific and dissociable consequences on the kind of information people can easily extract from them (e.g., object identity about what the object looks like vs. procedural knowledge about what type of action to use to successfully interact with the object).

General Discussion

Explanatory visualizations are a crucial tool for conveying mechanistic knowledge, and thus play a key role in many different scientific fields, including biology, physics, and engineering (Callaway, 2016; Chi et al., 1994; Heiser & Tversky, 2006; Lipşa et al., 2012). Nevertheless, there has been a longstanding gap in our understanding of what ordinary people think is relevant when trying to explain how something works, as well as how these visual explanations guide

people towards appropriate inferences. Towards closing this gap, here we investigated what information people prioritize when drawing visual explanations of simple mechanical objects (Experiments 1A & 1B). In addition, we measured how well these explanations enabled other people to learn about these objects based on these drawings (Experiments 2A, 2B, & 2C). We found that people spontaneously emphasized functionally important parts of these objects when producing an explanation, using more strokes to draw these parts and making them appear larger than when they only aimed to produce a visually accurate drawing of the object. They also selectively included abstract symbols in their explanations, including arrows and motion lines, suggesting that they believe that providing an explanation means going beyond drawing physical components of the same object. While these explanatory drawings more effectively communicated which action was needed to interact with the object than depictive drawings, this enhancement was accompanied by a loss in diagnostic information about the object's visual appearance. Taken together, our findings suggest that ordinary people can behave in systematic ways when asked to produce a visual explanation, prioritizing information about function (i.e., how parts move and interact) over information about structure (i.e., what parts look like and where they are). This work replicates and extends prior work on visual explanations (Heiser & Tversky, 2006) by showing how they are distinct from other kinds of illustrations not only in terms of what they include (e.g., arrows), but also what they omit (e.g., non-causally relevant parts).

Our findings contribute to a growing body of work characterizing how people evaluate and produce explanatory language (Chi et al., 1994; Legare & Lombrozo, 2014; Lombrozo, 2016; Walker et al., 2014; Walker et al., 2017). In this prior work, individuals who are prompted to produce verbal explanations of causal mechanisms also prioritize functional properties over perceptual features that are salient, but not causally relevant (Legare & Lombrozo, 2014; Walker et al., 2014). This pattern of results is broadly consistent with the current study, even though we elicited drawing-based explanations rather than verbal ones. However, our study goes beyond this prior work by further examining how the balance of structural and functional information in visual explanations guide inferences made by downstream learners. We found that visual

explanations outperformed visual depictions for supporting some inferences but not others, suggesting that explanations are not necessarily superior to depictions in all settings, but rather a specific tool for conveying knowledge cast at a particular level of abstraction. Moreover, by generalizing prior findings derived from verbal explanations to the visual modality, our work lends support to the notion that similar cognitive mechanisms may account for key aspects of explanatory behavior, regardless of whether these explanations are expressed using words or pictures. Taken together with other recent work extending principles originally developed to account for linguistic phenomena to the visual domain, this body of findings offers converging evidence for a substantial degree of domain generality concerning the mechanisms governing natural communication (Bergen et al., 2016; Fan et al., 2020; Frank & Goodman, 2012).

Our findings also have potential connections to theories of how goals influence how attention is allocated to different elements of a visual scene. In particular, the ability to convey the most goal-relevant information in a drawing may depend not only on what the person producing the drawing is attending to, but also what they expect *someone else* to attend to upon being shown the drawing. Recent work provides some support for the contribution of the former when the goal is to encode the entire visual scene, with visually salient objects being more likely to be included in a drawing (Bainbridge et al., 2019; Harel et al., 2006; Henderson & Hayes, 2017). To what degree does the way that different goals impact how visual attention is deployed across a scene (Chun et al., 2011; Yantis et al., 2000) also determine what information a person is most likely to draw? And how could such influences be differentiated from those providing the basis for adopting the perspective of one's communication partner and thus appropriately emphasizing the information that should be most salient to them (R. D. Hawkins et al., 2021), even if it is not what is most salient to oneself? Future studies could investigate the first question by measuring what an individual attends to in a visual scene under different communicative goals, for example by analyzing patterns of eye movements, and relating these measures to which objects they end up including in their drawing. To investigate the second question, future experiments could systematically vary the visual salience of some objects independently of their communicative

relevance, which would provide key measurements that could be used to develop and test quantitative theories of how these different factors jointly predict what and how people communicate information in drawings.

Our experimental approach also enables follow-up studies that probe how different kinds of communicative goals may subtly impact the kind of information people believe to be important to include in their explanations. In our study, participants were cued to produce drawings explaining how the machines functioned to produce the desired effect. However, participants may have interpreted these instructions to mean that they should either: (a) explain the specific mechanisms that cause the desired effect for this machine (i.e, how *these* gears turn the light on) or (b) explain the general principles governing the class of mechanisms used by the machine (i.e., how gears work in general). A participant approaching the task with the latter interpretation may be expected to produce drawings that departed more substantially from the visual appearance of the machine than a participant equipped with the former interpretation. Such drawings may be less effective for helping a naive viewer understand any specific machine, but potentially more effective for helping them generalize to a wide variety of machines employing similar physical mechanisms. Future studies could test these predictions directly, shedding light on how the tradeoff between functional and structural information may be modulated by how general a visual explanation is intended to be.

Another key direction for future work is to examine how expertise influences visual explanation behavior. The participants in our studies were unlikely to have received specific training in how to design effective visual explanations, and thus it may not be surprising that the explanations they produced did not outperform depictions in supporting identification of causally relevant parts. One potential explanation for this finding is that, by frequently omitting other (non-causal) mechanical parts and structural parts, these explanations failed to provide enough contextual information to help viewers situate the causally relevant part relative to the rest of the object. Future work could test this hypothesis by prompting drawers to take the perspective of a naive viewer (Shafto et al., 2014), to examine whether they would be more likely to include

enough additional structural information to produce more informative visual explanations. Such evaluations may help to clarify the role of perspective taking and pedagogical expertise in the production of explanations that are effective for different audiences.

Overall, this work contributes to our understanding of how visual explanations communicate mechanistic knowledge. In the long run, these studies may lead to both more unified theories of how visual perception, causal reasoning, and social cognition interact to support explanatory behavior, as well as improvements in how visualizations are designed to communicate scientific knowledge in educational and research contexts.

Author contributions

H. Huey, C. M. Walker, and J. E. Fan developed the study concept and design. Testing and data collection were performed by H. Huey and J. E. Fan. H. Huey performed the data analysis of Experiment 1A, 1B, 2A, 2B, and 2C under the supervision of J. E. Fan, and X. Lu performed the spatial analyses of Experiment 1B under the supervision of J. E. Fan. H. Huey drafted the manuscript, and C. M. Walker, and J. E. Fan provided critical revisions. All authors approved the final version of the manuscript.

Acknowledgements

Thanks to the members of the Cognitive Tools Lab and Early Learning & Cognition Lab at University of California, San Diego for helpful discussion. This work was supported by an NSF CAREER Award #2047191 to J.E.F. A subset of these findings were presented as part of the Proceedings of the 43rd Annual Meeting of the Cognitive Science Society.

References

Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature communications*, *10*(1), 1–13.

Bechtel, W. (2009). Constructing a philosophy of science of cognitive science. *Topics in Cognitive Science*, *1*(3), 548–569.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of science*, *78*(4), 533–557.

Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, *9*.

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

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.

Callaway, E. (2016). The visualizations transforming biology. *Nature*, *535*(7610), 187–188.

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

Chi, M. T., & VanLehn, K. A. (1991). The content of physics self-explanations. *The Journal of the Learning Sciences*, *1*(1), 69–105.

Chuey, A., Lockhart, K., Sheskin, M., & Keil, F. (2020). Children and adults selectively generalize mechanistic knowledge. *Cognition*, *199*, 104231.

Chun, M. M., Golomb, J. D., Turk-Browne, N. B., et al. (2011). A taxonomy of external and internal attention. *Annual review of psychology*, *62*(1), 73–101.

Corriveau, K. H., & Kurkul, K. E. (2014). "why does rain fall?": Children prefer to learn from an informant who uses noncircular explanations. *Child development*, *85*(5), 1827–1835.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci*, *13*(4), 148–153.

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

Fan, J. E. (2015). Drawing to learn: How producing graphical representations enhances scientific thinking. *Translational Issues in Psychological Science*, *1*(2), 170.

Fan, J. E., Hawkins, R. D., Wu, M., & Goodman, N. D. (2020). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, *3*(1), 86–101.

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.

Fiorella, L., & Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review*, *30*(3), 1115–1137.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). Cogsketch: Sketch understanding for cognitive science research and for education. *Topics in Cognitive Science*, *3*(4), 648–666.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of memory and language*, *31*(2), 129–151.

Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *36*(1), 39–53.

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in neural information processing systems*, *19*.

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive science*, *45*(3), e12926.

Hawkins, R. X., Sano, M., Goodman, N. D., & Fan, J. W. (2019). Disentangling contributions of visual information and interaction history in the formation of graphical conventions. *CogSci*, 415–421.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in cognitive science*, *3*(3), 446–474.

Hegarty, M., & Just, M.-A. (1993). Constructing mental models of machines from text and diagrams. *Journal of memory and language*, *32*(6), 717–742.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and instruction*, *21*(4), 209–249.

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

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature human behaviour*, *1*(10), 743–747.

Keil, F. C., & Lockhart, K. L. (2021). Beyond cause: The development of clockwork cognition. *Current Directions in Psychological Science*, *30*(2), 167–173.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143.

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

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, *11*(1), 65–100.

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

Lipşa, D. R., Laramee, R. S., Cox, S. J., Roberts, J. C., Walker, R., Borkin, M. A., & Pfister, H. (2012). Visualization for the physical sciences. *Computer graphics forum*, *31*(8), 2317–2347.

Lockhart, K. L., Chuey, A., Kerr, S., & Keil, F. C. (2019). The privileged status of knowing mechanistic information: An early epistemic bias. *Child development*, *90*(5), 1772–1788.

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

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of educational psychology*, *81*(2), 240.

Mayer, R. E. (1999). Multimedia aids to problem-solving transfer. *International Journal of Educational Research*, *31*(7), 611–623.

Newcombe, N. S. (2013). Seeing relationships: Using spatial thinking to teach science, mathematics, and social studies. *American Educator*, *37*(1), 26.

Prater, E. L. (1994). *Basic machines*. Echo Point Books & Media.

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

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International journal of human-computer studies*, *45*(2), 185–213.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cogn. Psychol*, *71*, 55–89.

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. (2001). Spatial schemas in depictions. *Spatial schemas and abstract thought*, *79*, 111.

Tversky, B. (2005). Prolegomenon to scientific visualizations. In *Visualization in science education* (pp. 29–42). Springer.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International journal of human-computer studies*, *57*(4), 247–262.

Tversky, B., Zacks, J., Lee, P., & Heiser, J. (2000). Lines, blobs, crosses and arrows: Diagrammatic communication with schematic figures. *International conference on theory and application of diagrams*, 221–230.

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

Walker, C. M., Lombrozo, T., Williams, J. J., Rafferty, A. N., & Gopnik, A. (2017). Explaining constrains causal learning in childhood. *Child development*, *88*(1), 229–246.

Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. In *Consciousness and the brain* (pp. 205–267). Springer.

Yantis, S., et al. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Attention and performance*, *18*(Chapter 3), 73–103.