BIROn - Birkbeck Institutional Research Online

Brody, G. and Revencu, B. and Csibra, Gergely (2022) Images of objects are interpreted as symbols: a case study of automatic size measurement. Journal of Experimental Psychology: General , ISSN 0096-3445.

Journal of Experimental Psychology: General

Images of Objects Are Interpreted as Symbols: A Case Study of Automatic Size Measurement

Gabor Brody, Barbu Revencu, and Gergely Csibra

Online First Publication, November 10, 2022. https://dx.doi.org/10.1037/xge0001318

CITATION

Brody, G., Revencu, B., & Csibra, G. (2022, November 10). Images of Objects Are Interpreted as Symbols: A Case Study of Automatic Size Measurement. *Journal of Experimental Psychology: General*. Advance online publication. https://dx.doi.org/10.1037/xge0001318

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001318

Images of Objects Are Interpreted as Symbols: A Case Study of Automatic Size Measurement

Gabor Brody[1], Barbu Revencu[2], and Gergely Csibra[2, 3]
[1] Department of Cognitive, Linguistic and Psychological Sciences, Brown University
[2] Cognitive Development Center, Department of Cognitive Science, Central European University
[3] Department of Psychological Sciences, Birkbeck, University of London

Are photographs of objects presented on a screen in an experimental context treated as the objects themselves or are they interpreted as symbols standing for objects? We addressed this question by investigating the size Stroop effect—the finding that people take longer to judge the relative size of two pictures when the real-world size of the depicted objects is incongruent with their display size. In Experiment 1, we replicated the size Stroop effect with new stimuli pairs (e.g., a zebra and a watermelon). In Experiment 2, we replaced the large objects in Experiment 1 with small toy objects that usually stand for them (e.g., a toy zebra), and found that the Stroop effect was driven by what the toys stood for, not by the toys themselves. In Experiment 3, we showed that the association between an image of a toy and the object the toy typically stands for is not automatic: when toys were pitted against the objects they typically represent (e.g., a toy zebra vs. a zebra), images of toys were interpreted as representations of small objects, unlike in Experiment 2. We argue that participants interpret images as discourse-bound symbols and automatically compute what the images stand for in the discourse context of the experimental situation.

Keywords: external symbols, size Stroop effect, methodology, communication, pragmatics

Supplemental materials: https://doi.org/10.1037/xge0001318.supp

In communication, humans often use *external representations*: physical scenes created and manipulated to convey information about entities that are relevant to the current communicative situation (Clark, 2016; Ittelson, 1996; Walton, 1990). Such external representations can be encountered in a variety of communicative media, ranging from pretend play and puppet shows to drawings, diagrams, and animations. The constituent parts of these representations are *symbols* in the broad sense: visual objects such as props, marks on paper, or pixel constellations, that temporarily or conventionally stand for entities other than themselves (DeLoache, 2004).

In this article, we focus on two properties of symbols. First, unlike ordinary objects, symbols require a referent that they can stand for. That is, after all, what a symbol is for: to carry information about some other entity. This implies that one needs to interpret external symbols to figure out what they currently stand for. Second, the interpretation assigned to symbols is, in general, achieved only with respect to the communicative context in which they are used. Consider a blue and red proportional bar-graph without legend and labels—one cannot tell what the graph stands for by only looking at it. The interpretation of the graph is only possible in response to a perceived context, like an explanation, a legend, or a previously established expectation about its communicative content. This implies that the nature of the referent that a given symbol stands for cannot always be determined based on the perceptible properties of the symbol. While there is little doubt that ad-hoc, arbitrary symbols (e.g., blue and red bars) require more than just visual information for interpretation, it is often assumed that iconic symbols, such as images of objects, have a more direct link to the objects they stand for by virtue of their iconicity. However, even iconic symbols do not always elicit an appropriate mapping in a context-independent way. While an image of a green olive may be interpreted by default as standing for a green olive, the very same image can stand for olives in general (in a grocery store), for olive oil (on a bottle), or for an olive tree plantation (on a map).

Such symbolic communication beyond natural language is ubiquitous, effortless, and quick. For instance, movies, animations, or

---

Barbu Revencu https://orcid.org/0000-0001-8701-9123

Gabor Brody and Barbu Revencu contributed equally and are listed alphabetically.

This project was supported by the European Research Council (ERC 742231). This study was not preregistered. The design and analyses closely follow previous work using the same paradigm as in the current article. All data have been made publicly available at the Open Science Framework (OSF) repository and can be accessed at https://osf.io/q2yzc/. The data and ideas in this article have not been disseminated anywhere before submission of the article. We have no known conflict of interest to disclose. We thank Athulya Aravind, Rachel Dudley, Roman Feiman, Barbara Pomiechowska, and Dan Sperber for comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Barbu Revencu, Cognitive Development Center, Department of Cognitive Science, Central European University, Quellenstraße 51, 1100 Wien, Austria. Email: revencu_barbu@phd.ceu.edu

video games are often fast-paced and still understood "instantaneously," with rich meaning attributed to them beyond what is visually encoded (e.g., Hochberg, 1986). While these pieces of art are created and consumed as communicative media, understanding their content rarely seems to require reasoning over particular interlocutors (e.g., movie directors, game designers).

In short, humans often find themselves in communicative situations in which they have to (a) figure out what the symbols in front of them stand for; and (b) use these assignments to interpret the messages that the symbols help convey. Nonetheless, many experiments in cognitive psychology present participants with pictures or animations on a screen, glossing over the possibility that participants might interpret these stimuli as symbols that are part of a communicative context (but see, e.g., Politzer, 2004; Snow & Culham, 2021). However, the very fact that participants are ostensively shown something (by the experimenters) may prompt them into interpreting it as an instance of communication (Sperber & Wilson, 1995). In addition, screens themselves are widely used as representational devices outside the psychology lab, which may be sufficient to trigger a communicative interpretation of the situation (DeLoache, 2004; Ittelson, 1996; Troseth & DeLoache, 1998; but see Millikan, 2017). If symbolic interpretations are triggered upon encountering experimental stimuli, this would carry both methodological and theoretical implications. Methodologically, it might urge researchers to consider and control for unintended effects of communicative inferences that their stimuli might induce in participants. Theoretically, it would provide evidence for the view that external symbols gain a communicative interpretation rapidly and automatically.

Are experimental stimuli presented on a screen interpreted as communicative symbols? The case study we will focus on in this article is the familiar-size Stroop effect reported first by Konkle and Oliva (2012). Konkle and Oliva (2012) had participants judge which of two images was displayed smaller or larger on the screen and found that participants slowed down and made more errors on trials in which the size difference direction between the two images was opposite to the real-world size difference direction of the depicted objects. For instance, participants responded slower when they were presented with a large image of a palm leaf and a small image of an elephant compared with a display of a small image of a palm leaf and a large image of an elephant. The fact that elephants are larger in the world than palm leaves interfered with judgments of image sizes, even though participants did not need to interpret the image contents for the task. This suggests that the process of encoding image contents is automatic. However, this conclusion leaves open several possibilities regarding the nature of this encoding process.

Here we consider two accounts that might underlie the familiar-size Stroop effect. The first possibility is that the interfering size measurement is an outcome of the perceptual processes that identify the category or the features of the objects depicted by the images. If this is the case, automatic size computation reflects previous experience of encountering such features and objects and/or computations internal to the visual system that use featural information as cues for object size. There is work suggesting that object features, rather than object categories, may drive the effect on the automatic size measurements which give rise to the size Stroop effect. Long and Konkle (2017) ran the familiar-size Stroop task using distorted images of objects (called *texforms*), which preserved only midlevel featural information (e.g., curvature). Even though the basic kinds to which these objects belonged were no longer recognizable, a Stroop effect was still present, implying that the midlevel features carry sufficient information about the size of objects. While these findings show that accurate basic-level recognition is not necessary for the Stroop effect to occur, they do not completely rule out a category-based explanation. It remains possible that participants inadvertently attempted to categorize the images at the basic level based on the midlevel features they were presented with. Even if these categorization attempts did not correctly identify the basic level kind of the underlying objects, the kinds identified by the participants could still have been in the right ballpark in terms of size (e.g., they could have been more likely to guess "building" or "statue" when showed a texform of a vending machine than when shown a texform of a perfume bottle). Thus, based on the available evidence, neither object-features nor object-category can be conclusively ruled out as the causal factor. For the current purposes though, these options are equivalent because both assume that the Stroop effect is driven by the mismatch between what is perceived on the screen and the perceptually similar individuals in the world (Konkle & Oliva, 2012; Long & Konkle, 2017). We will group these explanations (categories and features) under a single general account, which we will refer to as the *object recognition account*.

The second option is that the familiar-size Stroop effect arises because of the communicative inferences that are derived about the images. Under this account, we can construe each trial as a mini-discourse consisting of a question (e.g., "Which one is larger on the screen?") with two possible answers (the two images/response buttons). If participants interpret the images displayed on the screen as symbols, they will inadvertently encode what entities these images might be conveying information about. On this account, the incongruency comes from a mismatch between the relative sizes of the on-screen images and the relative sizes of the interpreted referents (e.g., a car image conventionally refers to a car). This account implies that the real-world size of the on-screen object matters less than the real-world size of the referent that the on-screen object currently stands for (e.g., an image of a toy car may activate the CAR-concept just as well as an image of an actual car). Moreover, the interpretation that is assigned to the images should be a function not only of the image features but also of the context in which the image is embedded (e.g., it might be influenced by other symbols that are present on the screen). Under this account, there is no direct, one-to-one mapping between pictures of objects and corresponding representations of size. For example, an image of a toy car will be taken to sometimes stand for a car, sometimes for a toy, depending on what else accompanies the image. We refer to this account as the *symbol interpretation account*.

We designed three experiments to evaluate the relative likelihood of the hypotheses outlined above. Experiment 1 is a replication of the familiar-size Stroop effect (Konkle & Oliva, 2012, Experiment 1a) with new stimuli. Experiment 2 introduces miniature objects, such as toys, which are ideally suited to tease the two hypotheses apart because they are small in the real world but also typically used to stand for entities that are large in the real world (e.g., a car). If the familiar-size Stroop effect is driven by object recognition, a small image of a toy car next to a large image of a couch should not slow down image size judgments as toy cars are much smaller than real couches. However, if the toy car (and consequently the image of the toy car) is taken to stand for an actual car, we should observe the opposite pattern: participants' judgments will be slower when a small image of a toy car is presented next to a large image of a couch, even though this configuration preserves the real-world size difference between the two objects depicted on the screen. Note that the symbol interpretation account does not predict that an image of a toy would necessarily be interpreted as standing for its nontoy counterpart. Instead, it predicts that some pictures—due to what they stand for in the context—give rise to a contrast between the real-world size of the depicted object and the perceived object size if participants interpret the depicted object to stand for a different referent. This prediction does not apply, a priori, to toys. Indeed, an image of a toy car can be thought of as ambiguous between a toy-car-interpretation and a car-interpretation. If participants opt for the toy interpretations when seeing images of toys, Experiment 2 will not be able to adjudicate between the recognition and the symbol interpretation accounts. However, if they opt for the toy referents interpretation, Experiment 2 could be construed as evidence against the object recognition account, as this would show a dissociation between the object and the size measurement.

Finally, Experiment 3 asks whether the relationship between an object and its attributed size measurement would be modulated by the identity of the second object. We investigate this question by comparing images of toy objects to images of the larger objects that the same toys typically represent (e.g., a toy car to a car). Should the very same image (e.g., a toy car) be interpreted as a large object in Experiment 2 but as a small toy in Experiment 3 (due to the explicit within-category contrast), this would suggest that participants interpret the two pictures presented on a screen in an integrated rather than piecemeal fashion. This pattern would provide evidence in favor of the symbol interpretation account, as external symbols should be interpreted as constituent parts of the scene they are embedded in. On the other hand, the object recognition account—in our current formulation—takes perception to output size measurements of individual object images, for which the identity of the second object on the screen should be irrelevant. We return to this issue in the General Discussion, where we discuss several ways in which the object recognition account could be expanded to accommodate such contextual effects in light of the data we present.

Experiment 1: Replication

Method

Experiment 1 was closely modeled on Experiment 1a of Konkle and Oliva (2012). Participants were presented with displays consisting of two different-sized images of real-world objects. The participants' task was to judge which of the two images was larger or smaller on the screen. In some of the trials, the larger image depicted the object that was larger in the real world; in other trials, the larger image depicted the object that was smaller in the real world (Figure 1, top row).

Transparency and Openness

We report how we determined our sample size, data exclusions (if any), manipulations, and measures in the study. The design of the study and the analyses were not preregistered, as all experiments closely followed the design and analyses of Experiment 1a in Konkle and Oliva (2012). All stimuli, anonymized data, analysis code, and research materials are available on the online Open Science Framework (OSF) repository of the project, accessible at https://osf.io/q2yzc/. Data for all reported experiments were analyzed using R, Version 4.1.2 (R Core Team, 2021), and the packages *ggplot*, Version 3.3.5 (Wickham, 2016) and *effectsize*, Version .6.0 (Ben-Shachar et al., 2020).

Participants

The sample consisted of 50 English-speaking participants (age: 19–70 years old, $M_{\text{age}} = 29.9$ years, $SD_{\text{age}} = 12.0$ years) recruited via the Testable Minds platform from all over the world. The sample size was chosen based on a pilot with 12 subjects to detect an effect of trial type with 99.9% power at significance level $\alpha = .05$ (pilot Cohen's $d = .734$). All participants gave informed consent before completing the experiment.

Stimuli

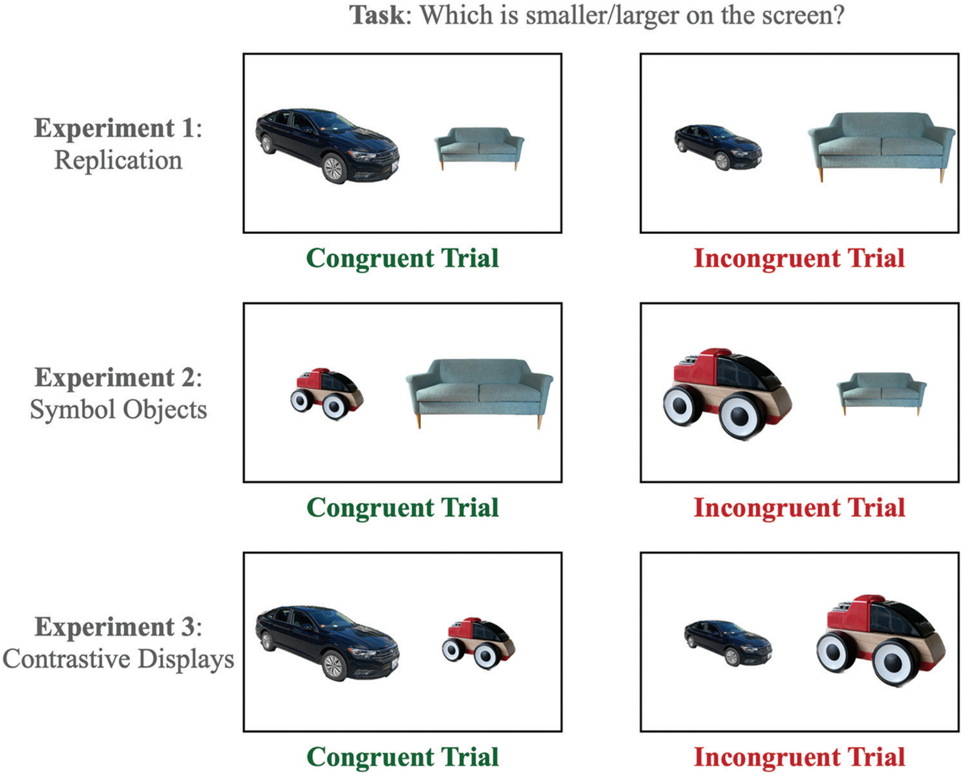
For Experiments 1–3, we gathered 36 triplets of objects, each of which contained a large object *X*, a small toy object *X*, and a different object *Y*, whose size lied in-between that of the real object *X* and that of the toy object *X*. In addition, for each of the three triplet pairs (*X*-*Y* in Experiment 1, toy *X*-*Y* in Experiment 2, and *X*-toy *X* in Experiment 3), we made sure that the absolute difference between the aspect ratios of the bounding boxes was at most .25 (based on Konkle & Oliva, 2012, Experiment 1a, with a slightly wider margin because the constraint had to hold across triplets rather than pairs). The triplet items were not matched in terms of filled areas (proportions of filled space in the corresponding bounding boxes), but we controlled for these differences statistically when analyzing the data (see online supplemental materials). The 108 object images were then resized to create two different-sized versions for each image. The large versions were resized such that the diagonal of the bounding box of the object was approximately 1,000 pixels; the small versions were created by scaling the large images down by a factor of .6. In congruent pairs, the size difference between the images was in the same direction as the real-world size difference of the depicted objects; in incongruent pairs, the size difference between the images was in the opposite direction to the real-world size difference of the depicted objects (see Figure 1).

Procedure

Participants were told that they would see two images on each trial and were instructed to press the 'F'-key on their keyboard to select the image on the left or the 'J'-key for the one on the right. They were also instructed that in one block of trials they would have to judge which of two images was *smaller* on the screen, while in another block of trials they would have to judge which of two images was *larger* on the screen. Participants underwent two short practice blocks (eight trials in total), in which they had to select which of two colored circles was smaller/larger on the

Figure 1

Schematic Design of Experiments 1–3



Note. Left: the size difference between the images is in the same direction as the real-world size difference of the depicted objects (congruent trials). Right: the size difference between the images is in the opposite direction to the real-world size difference of the depicted objects (incongruent trials). In Experiment 2 (middle row), congruency is defined based on the actual size of the objects depicted in the images (e.g., toy cars are smaller than couches). See the online article for the color version of this figure.

screen, to get familiarized with the task and with the two response keys. Once participants answered correctly to four consecutive practice trials in both blocks, the test phase began.

Each of the two test blocks (larger vs. smaller, order randomized across participants) consisted of 144 trials (36 pairs × 2 congruency [congruent vs. incongruent] × 2 sides [left vs. right of the screen]). Thus, the entire experiment consisted of 288 trials. Each trial started with a fixation cross for 700 ms, followed by the image comparison display. Correct responses were followed immediately by the next trial, while incorrect responses received error feedback ("Oops, this is incorrect! Remember, choose the one which is *smaller/larger* on the screen.") and by a 5-s interval before the next trial began. The order of trials was randomized for each participant. Once participants finished the first block, they were congratulated and told which task they would have to solve in the remaining block (*smaller* or *larger*, depending on the first block).

Results

Reaction Times

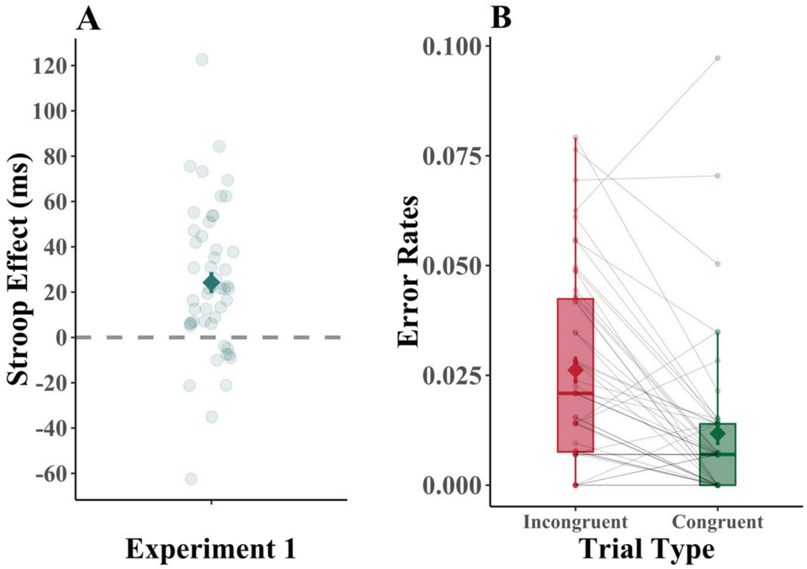
As in Konkle and Olivia (2012, Experiment 1), we excluded from the analysis incorrect trials and trials in which RTs were shorter than 200 ms or longer than 1,500 ms. This left us with 273.3 trials per participant, on average, out of a possible 288. To investigate RTs, we obtained a Stroop effect for each participant by subtracting the average RTs on congruent trials from the average RTs on incongruent trials. A positive Stroop score would mean that participants take longer to give an answer on incongruent trials; a negative Stroop score would mean that participants take more time to respond on congruent trials.

The overall effect of real-world size congruency replicated the original finding (Figure 2A): it took participants longer to make a visual size judgment on two images when the image sizes were incongruent with the real-life sizes of the objects depicted in the images ($M_{\text{congruent}} = 625.8$ ms, $M_{\text{incongruent}} = 650.0$ ms; $t(49) = 5.243$, $p < .001$, Cohen's $d = .742$, 95% CI [.429, 1.063]. To consult the results by task (*Which one is larger?* vs. *Which one is smaller?*), see the online supplemental materials.

Error Rates

Following Konkle and Oliva (2012), we also compared error rates across trial types within each condition (Figure 2B). While participants were, on average, 98% accurate, they were more likely to err in incongruent trials than in congruent trials, $t(49) = 5.308$, $p < .001$, Cohen's $d = .751$, 95% confidence interval, CI [.438, 1.073].

Figure 2
Experiment 1 Results



Note. (A) Stroop effects. Transparent circles represent within-subject Stroop effects (incongruent—congruent RTs); opaque diamonds show group average Stroop effect ±1 *SEM*. (B) Error rates. Transparent circles and the lines connecting them represent individual error rates as a function of trial type; opaque diamonds depict group averages ±1 *SEM*; boxplots indicate the median and interquartile range. See the online article for the color version of this figure.

Discussion

Experiment 1 replicated the size Stroop effect reported by Konkle and Oliva (2012): (a) participants were slower to make a visual size judgment in incongruent trials, that is, when the size relation of two on-screen images did not align with the size relation of the depicted objects; and (2) participants were less accurate in incongruent than in congruent trials. As noted in the beginning of the article, however, the results are ambiguous as to the nature of the process that gives rise to this effect. Is it simply the case that the category and/or their perceptual features of the objects in the images are associated to a previous encoding of such features and objects? Or do participants compute what these images stand for? We addressed this question in Experiment 2.

Experiment 2: Symbol Objects

To find out whether the familiar-size Stroop effect is driven by object recognition or by symbol interpretation, we replaced the large objects in Experiment 1 with miniature versions of the same objects. When participants compare images of couches to images of toy cars, which of the two size differences will they take longer to judge? Because toy objects are small, the visual object recognition account predicts that participants will take longer if the toy/miniature objects are depicted as larger than the medium-sized objects from Experiment 1. By contrast, only the symbol interpretation account can account for the possibility that participants will find those trials easier in which the toy/miniature objects are depicted as larger than the same medium-sized objects.

Method

The methods were identical to Experiment 1 except that the large objects were replaced by small symbol objects that represent them (e.g., car → toy car). Thus, in Experiment 2, participants had to compare displays of medium-sized objects versus small objects which typically stand for large objects (Figure 1, middle row). We define congruency based on the actual size of the objects depicted. For instance, a large image of a toy car next to a small image of a couch is an *incongruent* trial, as toy cars are typically smaller than couches. This choice is, of course, arbitrary, because the purpose of this experiment was to find out which of the two trial types would be incongruent for participants, but it is important to keep in mind for interpreting the results.

The sample consisted of 50 participants (age: 18–67 years old, $M_{\text{age}} = 31.9$ years, $SD_{\text{age}} = 11.7$ years) recruited via the Testable Minds platform. The sample size was chosen based on a new pilot with 12 subjects to detect an effect of trial type with 99.9% power at significance level $\alpha = .05$ (pilot Cohen's $d = .74$). All participants gave informed consent before completing the experiment.

Results

Reaction Times

As in Experiment 1, we removed from the analysis incorrect trials and trials in which RTs were shorter than 200 ms or longer than 1,500 ms. Based on these criteria, participants provided, on average, 280.2 valid trials. We again obtained a Stroop effect for each participant by subtracting the average RTs on congruent trials from the average RTs on incongruent trials. Because we defined congruency based on the actual size of the objects depicted by the images, a positive Stroop score would indicate that slower responses were produced when miniature object images were large (e.g., large toy car vs. small couch); and a negative score would mean indicate slower responses when miniature object images were small (e.g., small toy car vs. large couch).

Unlike Experiment 1, we found an overall negative Stroop effect (Figure 3A): participants' RTs were longer on *congruent* trials, where the visual size of the images on the screen *matched* the sizes of the objects depicted on-screen ($M_{\text{congruent}}$ = 568.5 ms, $M_{\text{incongruent}}$ = 561.1 ms; $t(49) = -2.462$, $p = .017$, Cohen's d = .348, 95% CI [.062, .639]).

Error Rates

Consistent with the RTs results, participants were more likely to make an error on congruent compared with incongruent trials (Figure 3B), $t(49) = -4.282$, $p < .001$, Cohen's d = .606, 95% CI [.304, .914].

Bimodality

The smaller Stroop effect in Experiment 2, in conjunction with the observation that one-third of the participants seem to have exhibited the opposite effect, could be driven by an underlying bimodal distribution. The bimodality would arise because some participants would take toy size measurements whereas others would take the size measurements of the large objects represented by the toy. This, however, does not seem to be the case. First, this smaller effect size compared with Experiment 1 is only apparent in the RTs but not in the error rates. If error rates and RTs are the consequence of the same incongruence, then these effect size differences should go hand in hand, but they do not. Second, the statistical analysis for multimodality on RT difference scores does not reject the null hypothesis (Hartigan's dip test for bimodality, $D = .036$, $p = .944$).

Experiments 1 and 2: Contrast

RTs Across Experiments 1 and 2

To compare the results of the first two experiments, we aggregated the two data sets and analyzed RTs as a function of trial type (congruent vs. incongruent, *within*-subjects) and experiment (1 vs. 2, *between*-subjects; see Figure 4). A 2 × 2 mixed analysis of variance (ANOVA) revealed a main effect of experiment, $F(1, 98) = 8.872$, $p = .004$, $\eta_p^2 = .083$; a main effect of trial type, $F(1, 98) = 9.326$, $p = .003$, $\eta_p^2 = .087$; and an Experiment × Trial Type interaction, $F(1, 98) = 32.926$, $p < .001$, $\eta_p^2 = .251$.

Item-Based Comparison Across Experiments 1 and 2

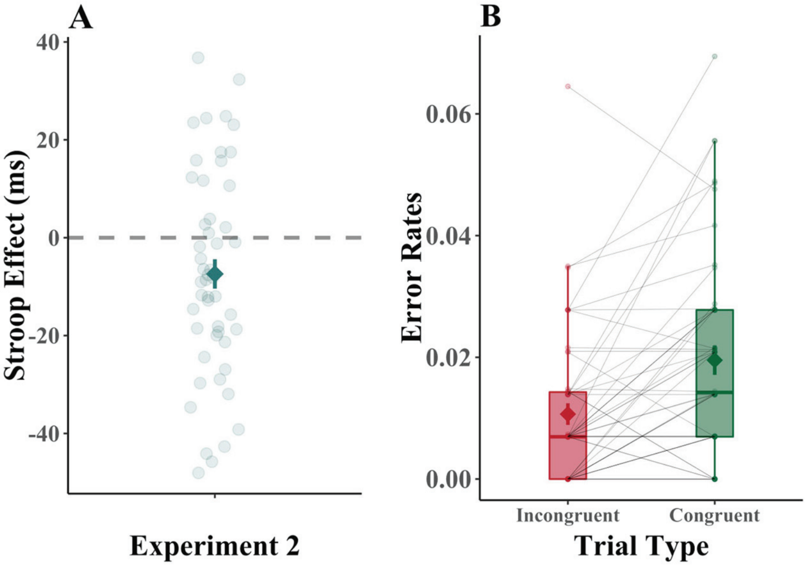
As an additional exploratory measure, we grouped the data by image pair (e.g., *train-fountain*, *bear-drumset*), and calculated the item-wise correlation of Stroop effects across the two conditions (see Figure 4). We found a strong negative correlation ($r(34) = -.65$, $p < .001$). This indicates that the Stroop effects we obtained tended to be driven by the same pairs across conditions. If, for instance, we observed a processing advantage for congruent *truck-hairdryer* trials in Experiment 1, we found a similar but opposite effect in Experiment 2 despite the many differences between trucks and toy trucks (the correlation is negative because we defined congruency at the level of the depicted objects).

Discussion

Experiment 2 produced a different pattern of results from that of Experiment 1. Participants were slower and more error-prone on

Figure 3

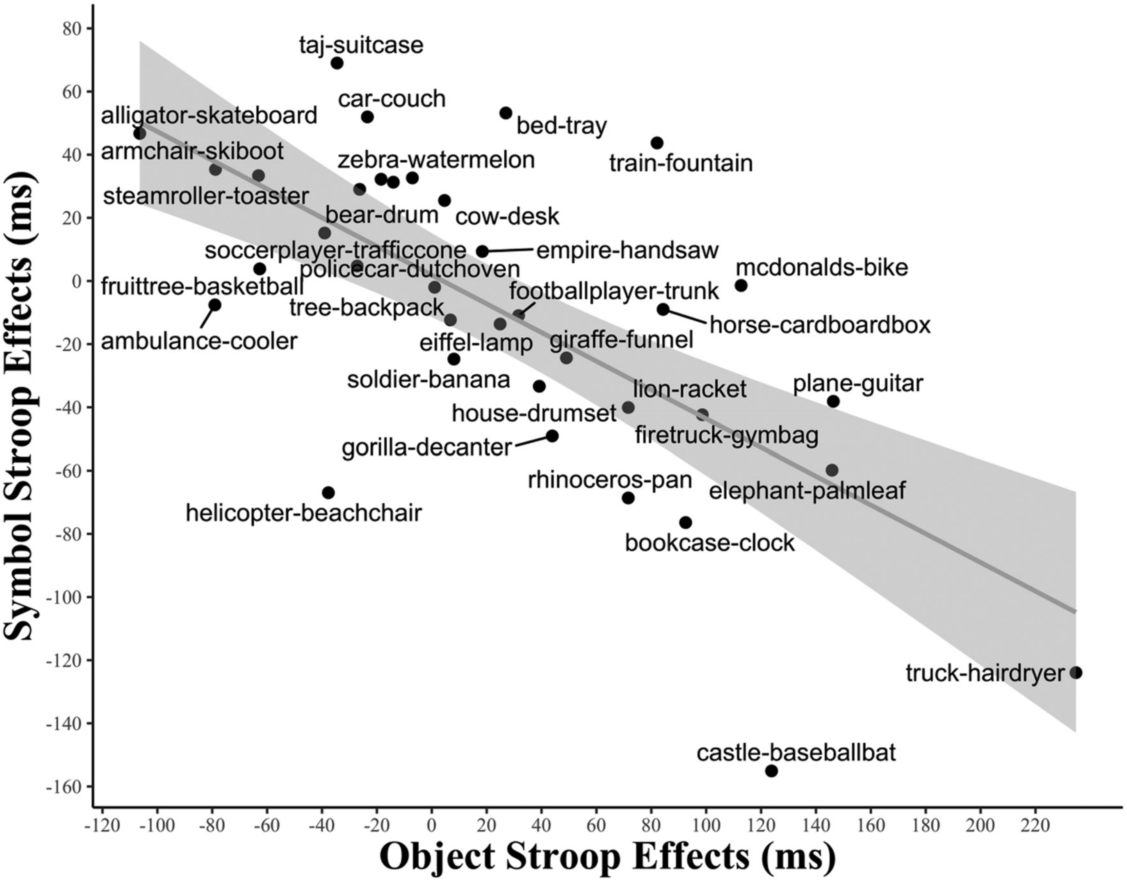
Experiment 2 Results



Note. (A) Stroop effects. (B) Error rates. See the online article for the color version of this figure.

Figure 4

Correlation Between Stroop Effects in Experiments 1 and 2 Across Items (x-Axis: Experiment 1; y-Axis: Experiment 2)



Note. The grey-shaded area represents the 95% confidence interval around the regression line.

congruent trials even though the difference in size between the two on-screen images went in the same direction as the real-world difference. Moreover, the Stroop effects by stimuli pair in the two conditions were strongly correlated, indicating that size judgments were similarly slowed down when the large object (e.g., a zebra) was depicted in a small image, irrespective of whether it was directly represented by a member of the kind (an image of a real zebra) or indirectly by an object that is often used to refer to it (an image of a toy zebra). Taken together, the results of Experiments 1 and 2 suggest that the Stroop size effect is not driven primarily by object recognition (e.g., toy zebra → small or toy zebra features → small) but by the inferred referent of the image (toy zebra → zebra → large).

Compared with Experiment 1, more participants in Experiment 2 exhibited a positive Stroop effect in RTs (but not in error rates). It is possible that the participants exhibiting a positive Stroop effect more likely interpreted the toy objects as toys (as opposed to as the objects they were toys of) than the majority did. If that were the case, this would provide further evidence for the symbol interpretation account. Under this account, communicative inferences on the visual input are responsible for the size measurements underlying the Stroop effect. Due to the ambiguity of toy images, some participants could simply have had different assumptions about what they are used to communicate about. On the object recognition account, this pattern of results would require auxiliary hypotheses. How could the very same visual input drive different effects across trials and across participants? Are toy objects visually bistable between toys and nontoy objects? Or do people have different visual systems to such an extreme degree? Neither of these explanations seems very promising. In short, while it is not clear whether the true distribution of the Stroop effect in Experiment 2 was indeed bimodal, if it were, it would support the symbol interpretation account more than the alternative.

Two concerns remain, however. First, it is possible that the features attended to in the visual processing of the images are orthogonal to the toy-real distinction. If this were the case, images of toys and images of real things would end up in identical outputs (e.g., both a zebra image and a toy zebra image output “zebra”). That is, participants in Experiment 2 might have, in some sense, mistaken the toys for the objects the toys stood for. Second, it is possible to slightly modify the object recognition account to accommodate the results even if participants did not mistake the toys for real objects. If the size Stroop effect is driven by object categories, one can postulate that participants always retrieve the conceptual content conventionally associated with the object in the image (e.g., a *toy zebra* always activates *zebra*). If, on the other hand, the size Stroop effect is driven by midlevel features, one could argue that toy-objects share the relevant midlevel visual features with real-object counterparts. Both these modifications double down on the core idea of the *object-recognition account*, namely that the main input to the size measurement of an image are its perceptible features. The *symbol interpretation accoun*t, by contrast, assumes that participants assign an interpretation to the symbols presented to them in relation to a discourse context. On this account, the context (e.g., the other image on the screen) can shift the interpretation of the images depicting the toy objects. We tested this prediction in Experiment 3, while also controlling for the possibility that participants in Experiment 2 mistook the toys for the objects the toys represented.

Experiment 3: Contrastive Displays

In Experiment 3, we tested whether participants are sensitive to the context in which an image is embedded when judging its relative size. We did this by pairing the 36 large-object images from Experiment 1 with their corresponding miniature versions from Experiment 2. If participants inflexibly assign a zebra interpretation to both a toy zebra image and a zebra image because that is the commonly associated conceptual content of both images, the Stroop effect should disappear. If, on the other hand, participants are sensitive to the communicative context, they should consider both images when assigning them an interpretation. As participants were faced with a direct contrast between large objects and their miniature versions, under the symbol interpretation account we expected them to go for a different interpretation for the toy object images than they did in Experiment 2: They should now stand for the corresponding concepts of toys, rather than for the concepts that the toys themselves usually stand for, because the paired images already stand for those concepts. This contrastive interpretation of the images predicts a size Stroop effect and a processing advantage for congruent trials.

Method

Experiment 3 was identical to Experiments 1 and 2, except for the stimuli pairs, which now consisted in pairs of objects and their corresponding miniature versions selected from Experiment 1 and 2, respectively (Figure 1, bottom row).

The sample consisted of 50 participants (age: 20–60 years old, M_{age} = 31.9 years, SD_{age} = 10.4 years) recruited via the Testable Minds platform. The sample size was chosen based on a pilot with 12 subjects to detect an effect of trial type with 99.9% power at significance level $\alpha = .05$ (pilot Cohen's $d = .86$). All participants gave informed consent before completing the experiment.

Results

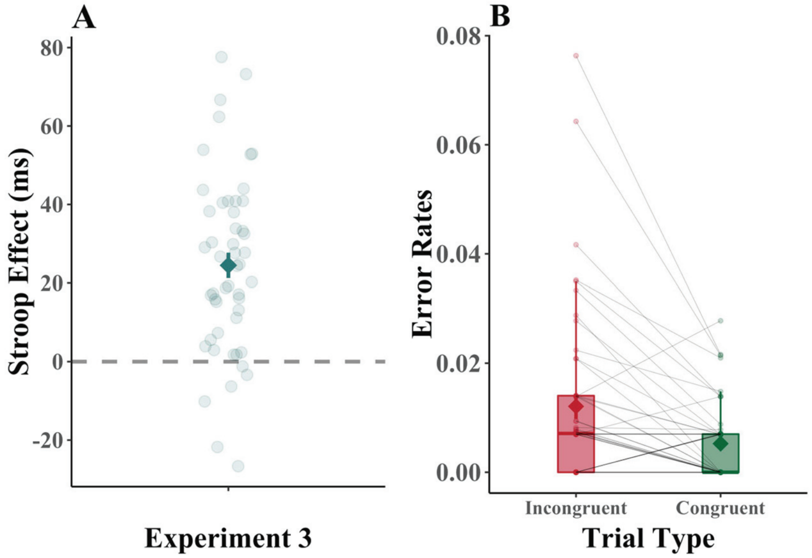
Reaction Times

As in Experiments 1 and 2, we excluded incorrect trials and trials that lasted shorter than 200 ms or longer than 1,500 ms. Each participant provided, on average, 276.9 valid trials out of a maximum of 288.

We found an advantage for congruent trials (Figure 5A). RTs on incongruent trials were longer than those on congruent trials ($M_{\text{congruent}}$ = 592.3 ms, $M_{\text{incongruent}}$ = 616.8 ms; $t(50) = 7.575$, $p <$ .001, Cohen's $d = 1.003$, 95% CI [.726, 1.431]). The Stroop effect replicates the original Konkle and Oliva (2012) finding, as well as its replication in Experiment 1 above.

Figure 5

Experiment 3 Results



Note. (A) Stroop effects. (B) Error rates. See the online article for the color version of this figure.

Error Rates

As in Experiments 1 and 2, participants were highly accurate in solving the task (on average, 99% correct responses), but they were more likely to make a mistake on incongruent trials than on congruent trials, $t(50) = 3.745$, $p < .001$, Cohen's $d = .53$, 95% CI [.233, .832], as in Experiment 1 (Figure 5B).

Discussion

Experiment 3 rules out two potential explanations for the findings of Experiment 2. First, the results of Experiment 2 were not due to participants' mistaking the toys for the objects they typically represent. Had they done so, they would have assigned the same interpretation to both images, which would have led to a null effect in Experiment 3. This was not what we observed. Participants took longer to reach a decision and made more mistakes on incongruent trials, that is, when an object was displayed as smaller than its miniature counterpart. Second, and more importantly, we have shown that the interpretation people give to images is not inflexible but changes according to the context. The results provided evidence that participants made a contrastive inference, as shown by longer RTs and more mistakes on the incongruent trials. By placing the toys next to the objects they usually represent, we elicited a different interpretation in Experiment 3 (e.g., a toy zebra stands for a toy) from the interpretation in Experiment 2 (e.g., a toy zebra stands for a zebra). These results rule out the modified object recognition accounts. Participants could not have assigned a conceptual description to an image is dictated solely by their conventional use nor could they have derived a size measurement from the visual features in the image: neither of these accounts are compatible with a shift in interpretation due to a change in context. Taken together, Experiments 2 and 3 suggest that people interpret images presented to them as symbols embedded in a communicative discourse. While the visual properties of a symbol might constrain the domain of entities and concepts it can represent, participants flexibly use the context to converge on an interpretation when multiple reasonable candidates are present (as in the case of toys).

General Discussion

We conducted a set of studies to explore how participants represent visual stimuli encountered in experimental contexts. Do their encoding of images of objects reflect the real-life entities that are depicted or do they interpret these images as symbols? And if the latter, how they decide what a picture stands for? Our case study for investigating these two questions was the familiar-size Stroop effect (Konkle & Oliva, 2012), which allowed us to make progress on both issues with simple manipulations.

In Experiments 1 and 2, we asked whether visual stimuli are recognized as the real-world objects they depict or if they are represented as symbols—objects that stand for something else. Having successfully replicated the original size Stroop effect in Experiment 1, we changed the stimuli in Experiment 2 such that, in every test pair, the larger one of the two objects was swapped with a toy version of the same object category. If automatic size measurements are based on the size of real-world entities, toy objects should be construed as small objects relative to the objects they were paired with. On the other hand, if pictures of toy objects carry the same symbolic content as images of their nontoy counterparts, participants should still treat them as larger than an actually bigger object. We found that the size Stroop effect followed the symbolic content rather than the real-life size of the depicted objects: participants represented pictures of *toy* zebras on par with *real* zebras when it came to automatic calculations of object size. This fits neatly only with the proposal that participants think of object presentations as symbolic.

Having found evidence for symbolic encoding in Experiments 1 and 2, we turned to our second question: what processes are responsible for creating the symbolic connection between picture and content? On one view, symbols, just like real objects, are recognized based on their perceptible properties. Communication using external symbols are often mediated by iconic convention. Pictures of zebras conventionally symbolize zebras and not, for instance, horses or houses. In fact, one would arguably call any picture a zebra-picture if it could be recognized by others as representing a zebra, even if that picture is cartoonish or barely resembles entities belonging to the subgenus *Hippotigris*. If you know how zebras are conventionally depicted, you might automatically encode any zebra-depiction, toy or otherwise, as standing for a zebra. On the alternative view, the connection between object symbols and their content is not just a recognition process but an interpretive one; whereby, participants decide what an object stands for in relation to some context. A stick-figure at a crosswalk signals that it is "okay to walk," but on a door at the airport it signals the location of the restroom. Without interpreting the context (in this case, the physical environment), there is no recognition-process that could reliably output the appropriate content. This view suggests that there is no context-independent way of identifying whether the image of a toy zebra is standing for a toy zebra, a real zebra, or maybe something else. Our results in Experiment 3 were consistent with the interpretation account: participants encoded a toy object as *smaller* than its nontoy counterpart. When looking at the size encoding of toy-objects, this is the exact opposite of what we found in Experiment 2, where the contrast between object categories drove the effect rather than ontological status (toy vs. real).

Just as Konkle and Oliva (2012), we found that the processes that automatically generated the irrelevant-size measurements of the depicted objects are rapid, spontaneous, and lacking control—properties that have been argued to be necessary, if not sufficient, signatures of visual processing (Hafri & Firestone, 2021; Scholl & Gao, 2013). Nevertheless, we have reasons to suspect that visual processes are not able to explain the contextual effects in our findings or to account for the size Stroop effect. Why does automatic size measurement of toy objects depend on the other object presented on the screen (Experiments 2 vs. 3)?

One could defend the visual origin of this effect by arguing that visual features of images of toy objects are mapped both to toy and nontoy versions of the same object category, such that upon seeing an image of a toy-zebra, the visual system would either output a *toy-zebra* (and its size) or a nontoy *zebra* (and its size). However, this mapping alone would not predict *when* participants should encode a toy object in one or the other way. As such, it fails to provide an explanation for the *systematic* context-dependence in our findings. If one of the two types of content is chosen at random, we should have found no Stroop effect in Experiment 2. In half of the trials, the toy objects would have been perceived as the smaller object, and in the other half, they would have been perceived as the bigger object, resulting in no size difference on average. We would have also observed a smaller size effect in Experiment 3: the toy objects would have been perceived as small only in half of the trials. However, the size Stroop effect was even larger in Experiment 3 than in Experiment 1, ruling this option out.

A second visual account of our findings could be that toy objects—counterintuitively—look larger than the medium objects they were presented next to in Experiment 2. By this account, the toy-objects might have created a visual illusion: they shared enough features with their large nontoy variants that the size measurement they activated was closer to these objects than to their de facto (small) sizes. If we add a further assumption that this illusion was only partial, we might also be able to explain how Experiment 3 worked: perhaps the toys looked larger than the medium objects (in Experiment 2) but smaller than the large objects (in Experiment 3). We dispense with this account because it is highly stipulative. Intuitively, toy objects do not look large; they look like toys. In addition, there is suggestive empirical evidence against it. If toy objects looked size-wise similar to the large objects, Experiment 3 should have produced the smallest effect size of all. We found the opposite.

A further way to incorporate contextual dependence into visual processes is to assume that, when the nontoy objects are presented next to the toy objects (as in Experiment 3), it makes toy objects *look* more toy-like. This could be similar to well-known perceptual contrast effects, such as the modulation of color perception by the brightness of the background (simultaneous contrast effect, e.g., Kinney, 1965). However, such a contrast effect should be more than just the relation of toy and nontoy features neighboring each other, as that would predict a straightforward contrast effect between toys and nontoys in Experiment 2 as well, where we did not find one. To generate the appropriate contrast effect, vision should apply a rule along the lines of "for any toy-object *x*, if and only if there is another object *y* that represents the same category as *x* but is not a toy, encode *x* as a toy." A rule of this sort would be radically different from the visual contrast effects that have been discussed in the literature for at least two reasons. First, create such contrast effects, the relevant constraints would have to encode symbolic properties like "*toy version*" and "*representing the same category as*"—properties that paradigmatically fall outside perceptual processes. Second, this potential constraint would be a post hoc stipulation, which does not follow from any general account of perceptual processing, and as such has little to no predictive power.

Yet, another way of attempting to root the contextual contrast effect in visual processes is to suggest that it reflects a more general process of optimal visual inference under uncertainty (e.g., Weiss et al., 2002). This option would concede that the process that creates object descriptions is interpretative in some sense, but would still assume that this interpretation is created within the confines of the visual system. Depending on the visual context, it might be more optimal to encode some object as a *toy zebra* rather than *zebra*. However, why would it be optimal to encode pictures of toys as *toys* in Experiment 3, but as *nontoys* in Experiment 2? To make this work, one would have to posit that the likelihood that two neighboring objects that in principle belong to the same category are actually from the same category is low. Could it be that the presence of a zebra decreases the likelihood of encountering another zebra and increase the relative likelihood of encountering a toy one? If anything, the opposite seems more plausible: if there is a zebra around, the probability of encountering another (nontoy) zebra increases.

To generate the correct predictions, we need a different type of rational process, one that asks, "Why am I presented with these pictures?" instead of trying to figure out "What is most likely to be out there?." If the stimulus is understood as part of a communicative act, there is good reason to interpret the contrast between two items as a matter of identifying the communicative message and not as a matter of identifying object categories. One might assume that being presented with side-by-side images of two objects that conventionally stand for the same category is not the outcome of random sampling of tokens that happen to belong to the same object category but a deliberate contrast. And if the difference between the images is made on purpose, then this distinction should play a role in how one interprets what they stand for. This is a type of rational inference, but not one that the visual system could be straightforwardly responsible for.

This is not to say that vision plays no role in the above inferential processes and in the size Stroop effect in general. Every account of the effect must presuppose some visual processes, as there could be no size Stroop effect without a visual representation of the stimuli. Our findings indicate that the interfering size measurement may not originate directly from these perceptual processes, but instead from a communicative interpretation of their outputs. A good analogy to the role of vision in our studies is the role of the auditory systems in understanding the meaning of a spoken sentence: in both cases, perceptual processes have to create an encoding of the input that is amenable for an interpretation by other processes, but they themselves do not provide the interpretation. For this inference to be perceptual in nature, the visual system would need to accommodate a vocabulary and corresponding theory of communication that references notions like *communicative act*, *addressee*, *message*, or *on purpose*.

As a reviewer pointed out to us, our arguments against the visual origin of the size Stroop effect rely on inference to the best explanation, not on a definitive falsification of purely perceptual accounts. Indeed, it is possible to come up with a version of the object recognition account that can explain both symbolic encoding and context dependence. This can be achieved either by explicitly denying some of our premises (e.g., it may be that the presence of a zebra decreases the likelihood of encountering another zebra and/or increases the likelihood of encountering a toy one) or by enriching the visual system with processes that yet to be discovered. However, even if these amendments are made to the object recognition account, future work should still assess the relative contributions of recognition processes and interpretative processes in generating the size Stroop effect. For instance, one way to further probe this question, suggested by the same reviewer, would be to assess whether the context dependence we observed in Experiments 2 and 3 would still hold with texform versions of the same images, which hinder basic-level recognition while preserving midlevel information. If toy objects retain context-dependent encoding in the absence of explicit recognition, this might support a more pronounced role of visual processing.

However, if our analysis is correct and the size Stroop effect turns out to stem from communicative interpretation and if it reflects fast and automatic processing, then it follows that, just like visual processes, communicative interpretation of visual stimuli can also be fast and automatic. In fact, such fast automatic processes are commonly discussed in psycholinguistic research. One well-studied example is the phenomenon of scalar implicatures. Upon encountering a sentence such as "I ate some of the cookies," people quickly and automatically infer that the person uttering this sentence did not eat *all* of the cookies. While the precise nature of such inferences is debated (Fox, 2007; Franke, 2009; Horn, 1972), all accounts that have been put forth agree that humans automatically compute and consider some alternative states of affairs (e.g., why the speaker did not say "I ate *all* of the cookies") and interpret the utterance accordingly. Strikingly, listeners even struggle to stop entertaining scalar implicatures when they do not need to (Fox, 2014; Magri, 2009), not unlike the size Stroop effect. In both cases, the relevant processes seem not to be accessed nor used based on task relevance.

So far, we have centered the discussion around the processes that might generate the irrelevant size measurement of the depicted objects without asking why the Stroop interference occurs in the first place. While there is no shortage of cognitive theories on explaining why the simultaneous representation of two measurements could interfere with each other (see MacLeod, 1991, for a review), we think that a communicative account of the effect opens the door to a novel type of explanation building on an insight common to many pragmatic theories: communicative acts are expected to be efficient (Chierchia et al., 2012; Gazdar, 1979; Grice, 1975; Sperber & Wilson, 1995).

We propose that participants consider every stimulus in the experimental situation as conveying some communicative message. We assume that participants analyze these messages with the expectation of communicative efficiency: they expect that whatever the message is, it should be transmitted in the most straightforward, least confusing way possible. Computing communicative efficiency outside laboratory environments might be harder to evaluate, but in a classic size Stroop study, there are only two degrees of freedom to consider: (a) the actual size of the pictures; and (b) the symbolic content of the pictures with their associated size measurements. Equipped with these two variables, we can explain why congruent trials are more efficient than incongruent ones. In congruent trials, the addressee converges on the same message regardless of which of the two variables they base their interpretation on: actual size and symbolic size comparisons point in the same direction. On the other hand, on incongruent trials, the two variables always point in opposite directions for reasons that are not specified. Suppose the relevant message in a trial is "the picture on the left is larger." There are only two possible ways that this could be transmitted, either with the symbolic content also being larger on the left (congruent) or it being smaller on the left (incongruent). Assuming that the two options are equally costly, why should a communicator ever opt for the incongruent variant? Thus, the reason for the size-Stroop effect might be that incongruent trials violate this expectation of communicative efficiency (and perhaps prompt participants to search for further communicative content). While highly speculative at this point, this account has the advantage of giving a principled pragmatic account to the size Stroop effect, and possibly to other Stroop effects as well.

We are not suggesting that all Stroop effects could be explained in a communicative framework. What we are arguing for is that separating mechanisms that reason over efficient communication from other processes might turn out to be theoretically fruitful and, in some cases, methodologically necessary. Using pictures or other visual stimuli in experimentation is ubiquitous, from vision science to social psychology. The present findings have methodological implications for such studies. We found that even when it comes to rapid automatic decisions, participants do engage in a communicative interpretation of the stimuli. Therefore, in any experiment in which participants encounter visual stimuli, their behavior might reflect participants' *interpretation* of those stimuli as external, communicative symbols rather than their mere recognition of the entities that are depicted on the screen. Consider a simple animation involving two geometric shapes moving in a contingent way on the screen (Heider & Simmel, 1944). People interpret such an animation in agentive terms and parse the on-screen interaction as a chasing event. This interpretation has been proposed to be due to the self-propelled motion exhibited by the geometric shapes (e.g., Scholl & Tremoulet, 2000). However, if viewers treat the animation as a representation and its constitutive parts as symbols to be interpreted, finding that, say, self-propelled motion is a cue to agency is ambiguous between purely a perceptual interpretation (people perceive agency when confronted with self-propelled motion) and a communicative interpretation (self-propelled motion is an efficient way of conveying agency). While both interpretations imply a strong link between self-propulsion and agency ascription, teasing apart the relative contribution of these candidate processes requires careful experimental controls.

Conclusion

We provided evidence that participants encode pictures of objects as having symbolic and context-dependent content, indicating that the familiar-size Stroop effect is driven by communicative inferences rather than just visual recognition. We have argued that, when presented with images on a screen, humans do not simply encode their features or category but automatically try to figure out what the visual objects in front of them currently stand for. Moreover, this interpretive process, which depends on perception but does not originate in perception, exhibits signature properties of vision: it happens quickly, automatically, and without direct relevance to the task at hand. Placing this task in a communication framework opens the door for exploring Stroop phenomena as arising from perceived communicative inefficiency—a speculative proposal in need of further empirical investigation.

Context

While many cognitive psychologists would agree that pictures are representations of objects and scenes, they hardly consider the possibility that this fact contributes to their studies. Many times this is not relevant, but sometimes may be. While studying how infants understand screen-based depictions of events (Revencu & Csibra, 2021), we realized that this aspect of the experimental situation is underappreciated and underresearched even beyond developmental research. The familiar-size Stroop effect (Konkle & Oliva, 2012) offered us a case study to address this question in adult participants.

References

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), Article 2815. https://doi.org/10.21105/joss.02815

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2297–2331). Mouton de Gruyter.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347. https://doi.org/10.1037/rev0000026

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, *8*(2), 66–70. https://doi.org/10.1016/j.tics.2003.12.004

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). Palgrave Macmillan. https://doi.org/10.1057/9780230210752_4

Fox, D. (2014). Cancelling the Maxim of Quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics*, *7*, Article 5. https://doi.org/10.3765/sp.7.5

Franke, M. (2009). *Signal to act: Game theory in pragmatics*. Institute for Logic, Language and Computation. https://hdl.handle.net/11245/1.313416

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. Academic Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Volume 3. Speech acts* (pp. 41–58). Academic Press.

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, *25*(6), 475–492. https://doi.org/10.1016/j.tics.2021.01.006

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259. https://doi.org/10.2307/1416950

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Volume 1. Sensory processes and perception* (pp. 22.1–22.64). Wiley.

Horn, L. R. (1972). *On the semantic properties of logical operators in English.* (Doctoral dissertation). University of California, Los Angeles. https://linguistics.ucla.edu/images/stories/Horn.1972.pdf

Ittelson, W. H. (1996). Visual perception of markings. *Psychonomic Bulletin & Review*, *3*(2), 171–187. https://doi.org/10.3758/BF03212416

Kinney, J. A. S. (1965). Effect of exposure time on induced color. *Journal of the Optical Society of America*, *55*(6), 731–736. https://doi.org/10.1364/JOSA.55.000731

Konkle, T., & Oliva, A. (2012). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 561–569. https://doi.org/10.1037/a0028294

Long, B., & Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition*, *168*, 234–242. https://doi.org/10.1016/j.cognition.2017.06.025

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*(3), 245–297. https://doi.org/10.1007/s11050-009-9042-x

Millikan, R. G. (2017). *Beyond concepts: Unicepts, language, and natural information*. Oxford University Press. https://doi.org/10.1093/oso/9780198717195.001.0001

Politzer, G. (2004). Reasoning, judgement and pragmatics. In I. N. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 94–115). Palgrave. https://doi.org/10.1057/9780230524125_5

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revencu, B., & Csibra, G. (2021). For 19-month-olds, what happens on-screen stays on-screen. *Open Mind: Discoveries in Cognitive Science*, *5*, 71–90. https://doi.org/10.1162/opmi_a_00043

Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford & V. A. Kuhlmeier (Eds.), *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 197–230). MIT Press. https://doi.org/10.7551/mitpress/9780262019279.003.0009

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Snow, J. C., & Culham, J. C. (2021). The treachery of images: How realism influences brain and behavior. *Trends in Cognitive Sciences*, *25*(6), 506–519. https://doi.org/10.1016/j.tics.2021.02.008

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell Publishing.

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development*, *69*(4), 950–965. https://doi.org/10.1111/j.1467-8624.1998.tb06153.x

Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Harvard University Press.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604. https://doi.org/10.1038/nn0602-858

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Received March 14, 2022
Revision received September 12, 2022
Accepted September 21, 2022 ■