What counts? Visual and verbal cues interact to influence what is considered a countable thing

Dana L. Chesney · Rochel Gelman

Abstract Many famous paintings illustrate variations in what we here dub "referential depth." For example, paintings often include not only portrayals of uniquely referenced items, but also reflections of those items in mirrors or other polished surfaces. If a painting includes both a dancer and that dancer's reflection in a mirror, are there one or two dancers in the painting? Although there are two images of a dancer, both images reference the exact same dancer. Consequently, counting both may seem to violate the constraint against double counting (Gelman & Gallistel, 1978). This illustrates that determining which things "count" in a given context may not be straightforward. Here we used counting tasks paired with illustrations that manipulated referential depth to investigate the conceptual, perceptual, and language variables that may influence whether a "thing" is a "countable thing." Across four experiments, 316 participants counted items in displays that included both foreground items and items placed inside *mirrors*, *picture frames*, and *windows*. Referential depth and frame boundaries both influenced counting: For one thing, participants were more likely to count items contained by windows than by picture frames or mirrors. Moreover, items in mirrors were rarely counted unless they were interpreted as reflections of items "off screen." Also, the items contained inside windows were sometimes (~10 % of trials) excluded from the counts, when counting them would require crossing frame boundaries. We concluded that conceptual and perceptual contexts both influence people's decisions about the physical boundaries of the to-be-counted set and which items within these boundaries are countable.

Keywords Counting · Enumeration · Containment · Framing · Sets · Referential depth

D. L. Chesney (✉)
Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210, USA
e-mail: dlchesney@gmail.com

R. Gelman
Department of Psychology, Rutgers University-New Brunswick, 152 Frelinghuysen Road, Piscataway 08854, NJ, USA

Introduction

Gelman and Gallistel's (1978) counting principles state that any real or imagined thing that can be thought of as an individual—even an imaginary playmate—may be collected into a countable set. However, the permissiveness of this definition of a countable thing produces a serious "framing problem." If anything *can* be counted, then determining which things *should* be counted must be a critical first step in the counting process. This determination may not always be straightforward; a person's intuitions about "what counts" may not align with what another might call the "correct" set. Consider Fig. 1. Schoolchildren are often asked how many squares appear in grids like this (Craine, 1994). Educators often consider the "correct" answer to be 14, counting all of the 1 × 1, 2 × 2, and 3 × 3 squares. Yet many feel the intuitive answer is nine, counting only the 1 × 1 squares (see Witkin, Moore, Goodenough, & Cox, 1977). Nevertheless, although different individuals show some variation when judging "what counts," people share strong intuitions about what sets may be counted. For example, if an individual counts one of the 1 × 1 squares, there is the strong intuition that all nine 1 × 1 squares should be counted. How people determine what is countable in a given context invokes a fundamental question about human thought. Here we investigated this aspect of human cognition by manipulating perceptual and conceptual variables within a counting framework, to determine how they impact what things people judge should be counted.
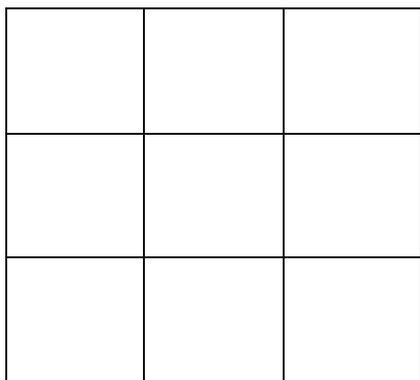
Fig. 1 There are 14 squares in this grid: nine 1 × 1 squares, four 2 × 2 squares, and one 3 × 3 square

Introducing "Referential depth" as a conceptual feature that may impact what counts

Artists provide us with many rich examples that illustrate the problems involved in deciding whether a "thing" is a countable thing. For instance, many paintings not only portray objects, but also objects' reflections. It is also not uncommon for artworks to include a painting of a painting. This leads us to ponder whether these conceptually complex variations in representation impact how people determine what things "count," and thus which things are countable. Here we coin the phrase "referential depth" to discuss these variations in representation.

Referential depth essentially refers to how "real" an item is within a given context. Items that are considered real within a given context would be at Level 1, with "less real" items being at greater levels of referential depth. The three paintings by Edgar Degas in Fig. 2 illustrate different levels of referential depth. If you, the reader, assign yourself and the world around you to Level 1, the images in Fig. 2 would be at Level 2, and the items in the "paintings" within the central image would be at Level 3. However, we can shift our frame of reference to consider the relative levels of "realness" within the *painted world*, and assign the "real" items in the paintings to Level 1.

In the painting on the left, both the woman in the foreground and the buildings we can see through the window behind her are at the same level of referential depth; they are both "real" within the painted world, and thus both can be assigned to Level 1. In the central Degas painting, the items in the "paintings" on the wall are at a greater level of referential depth than the women in the foreground. The women are real within the painted world and can be assigned to Level 1. In comparison, the images in the picture frames on the wall are paintings of paintings; they reference paintings within the painted world rather than real items. Thus, these images would be at Level 2. The rightmost Degas painting shows three women on the right and two "reflections" of women in the mirror on the left. If we assign the three images of women on the right to Level 1, then the two images of women in the mirror would be assigned to Level 1' (' following the mathematical notation for a derivative): They are *projections* of real women at Level 1, but are not the women themselves. Interestingly, one reflection is of a woman "off-screen," and thus indicates the presence of a fourth unique woman in the painted world. The other reflection is of one of the three women on the right, and thus does not reference a unique individual. We can thus further specify that this image is at Level 1'* (* following the linguistic notation for a disallowed item): It references an individual already represented by a Level 1 image. Thus, although the question of how many women are in the Degas painting on the left is quite straightforward—one woman— the question of how many women are in the painting on the right is not. Answers of three (the "real" Level 1 dancers on the right), four (the real dancers, and the dancer implied by her reflection to exist "off screen"), or five (all of the images of dancers) are plausible.

Similar variations in referential depth are invoked by the scenes in Fig. 3. These scenes all have the same general composition: two items in the foreground and a smaller, matched pair of items above them inside a quasi-rectangular visual frame that can be identified as a *window*, *mirror*, or *picture frame*. These frames' identities place the items they contain at different levels of referential depth. The foreground items in

Fig. 2 Three images from paintings by Edgar Degas, showing different levels of referential depth: on the left, *Dancer at the Photographer's Studio* (1874); in the center, *Mary Cassatt and Her Sister at the Louvre* (1879); on the right, *The Dancing Class* (1870)
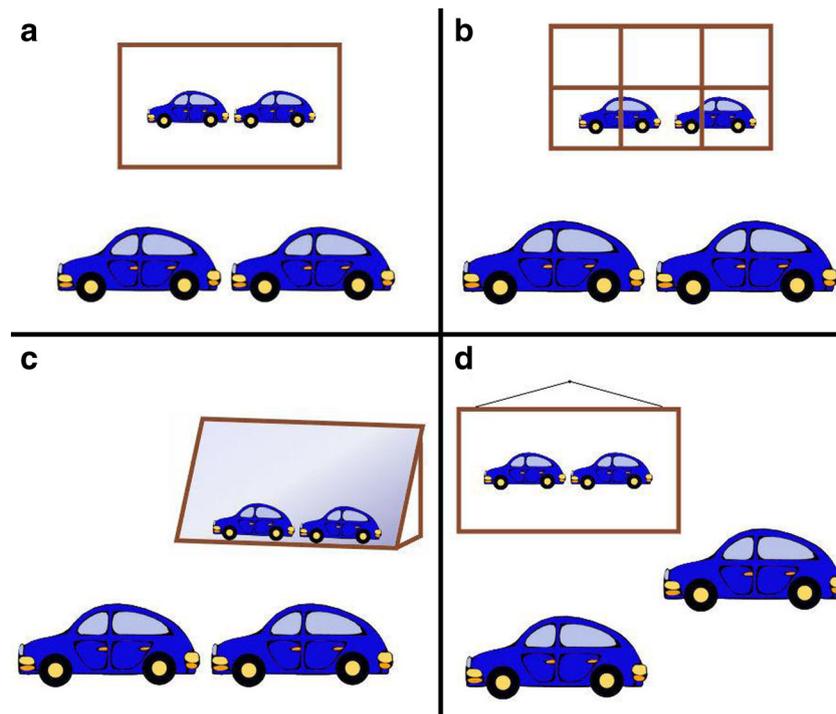
Fig. 3 Scenes presenting cars both below and inside various quasi-rectangular visual frames. In panel a, the frame is drawn as a simple rectangle, but the other panels show it drawn to resemble (b) a window, (c) a mirror, and (d) a picture frame

all four scenes and the items in the window in scene *b* can be assigned to Level 1. Following from this, the items in the picture frame in scene *d* would be at Level 2. The items in the mirror in scene *c* would be at Level 1', or alternatively at Level 1'* if they were considered reflections of the Level 1 items portrayed in the foreground. The items in the rectangle in scene *a* can be placed at any of these levels, depending on whether the observer interprets the rectangle to be a window, picture frame, or mirror.

It is possible that an item's level of referential depth may impact people's intuitions about whether an item is countable. There is evidence in the counting literature that people share intuitions about what they should count. For example, Shipley and Shepperson (1990) demonstrated that young children have a "whole object bias": When presented with sets made up of several different kinds of toy animals, young children were much more successful when asked to enumerate individual items than when asked to enumerate item kinds. People also share intuitions regarding when and if "reconstructed" items should be counted. When adults viewed arrays containing both whole objects—for instance, three forks—and objects broken into parts—for instance, one fork broken into three pieces—they were more likely to combine broken parts in their counts when asked "Can you count the forks?" (example answer: "four," three whole forks + one reconstructed fork) than when asked "Can you count these *things*" (example answer: "six," three whole forks + three fork pieces) (Shipley & Shepperson, 1990; Wagner & Carey, 2003). Although one

can count any item that one can individuate from the world— such as a reconstituted fork—this individuation is not sufficient for inclusion in a countable set (Giralt & Bloom, 2000; Markman, 1979). Importantly, however, adults in these studies showed little other variation in the kinds of counts they produced. Adults did not, for example, count the tines of the forks, even when asked to count "things." Convergence on such a limited range of countable sets would not occur unless people share strong biases about what items should be counted.

Here we investigated whether and how referential depth impacts counting behavior. We predicted that items at greater levels of referential depth would be counted less frequently, particularly when contrasted with items that were more "real": Items at Level 1 thus should be counted more often than items at Level 1' or 2. We further predicted that people should particularly disprefer to count "reflections" in mirrors when these refer to the same items as foreground items already in the to-be-counted set (Level 1'*). A Level 1'* image references the exact same item as a Level 1 image; consequently, including both images in the same count might invoke the prohibition against counting the same item twice (the "one-to-one principle"; see Gelman & Gallistel, 1978). Finally, because this "realness" is based on cognitively complex symbolic understanding (see DeLoache, 2000; DeLoache, Miller, & Rosengren, 1997; DeLoache, Pierroutsakos, & Uttal, 2003), we predicted that people should be able to use "realness" as an explicit criterion for determining what should be counted.

Frame boundaries as a perceptual feature that may impact what counts

The perceptual features of scenes can impact items' referential depth (e.g., being contained by a picture frame can place an item at Level 2). However, the perceptual features of scenes may impact what is counted in other ways. Many studies have shown that the visual features of sets can sometimes influence people's perception of the numerosity of those sets (Allen & McGeorge, 2008; Alston & Humphreys, 2004; Hurewitz, Gelman, & Schnitzer, 2006; Mandler & Shebo, 1982; Trick & Enns, 1997a, b; van Oeffelen & Vos, 1982; Vos, van Oeffelen, Tibosch, & Allik, 1988). Visual boundaries, in particular, have been shown to negatively impact people's ability to quickly enumerate sets (Chesney & Gelman, 2012; Trick & Pylyshyn, 1993, 1994). Chesney and Gelman demonstrated that when items are arranged such that some of them are contained inside others (e.g., a set of six circles composed of two subsets of three in which one circle surrounds two others), peoples' immediate perception of set size can "miss" or "skip over" some of the circles in the containment relationship. People may fail to nonverbally "count" items that are more difficult to *perceptually* parse as separate items (e.g., visually nested circles), even when one would conceptualize these as unique items if the perceptual parsing had been successful. Given that visual boundaries can affect intuitive perceptions of set size, visual boundaries might also influence what people consider the most intuitive countable set. Thus, we predicted that people would disprefer to construct countable sets that crossed these boundaries, such as by counting cars seen both "in front of" and "through" a window, even when items are at the same level of referential depth.

The present study

When interviewed, adults assume that the questioners choose their words so as to be easily understood (Clark & Schober, 1992). Consequently, adults interpret questions as meaning whatever it is most obvious to the listener that they *should* mean. We used this "presumption of interpretability" as an investigative tool, by asking adults to count in situations in which the to-be-counted set was ambiguous. As such, participants' counts were indicative of their intuitions about what *should* be counted in the different contexts. Essentially, we made use of participants' expected strategic attempts to "guess" what the experimenter intended them to count as they complied with that intention as a way to elicit what participants considered to be the most intuitive countable set. We used this method to investigate three potential factors that might influence what people count. We hypothesized that

(1) Referential depth would influence people's counts, such that they would be less likely to count items at greater levels of referential depth: That is, Level 1 items in windows would be counted more often than Level 1' items in mirrors or Level 2 items in picture frames.

(2) People would particularly disprefer to count images at Level 1'* (i.e., reflections) that might refer to the same things as Level 1 items already included in the count, since this would invoke the prohibition against double counting.

(3) People would disprefer to count sets that crossed visual boundaries, even when items could be interpreted to be at the same level of referential depth.

Experiment 1

Method

Participants A total of 61 (38 male, 23 female; mean age = 18.9 years) undergraduates attending Rutgers University–New Brunswick participated for course credit. An additional two participants were excluded due to experimenter error.

Stimuli The experimental stimuli consisted of 12 unique scenes, each portraying a pair of identical items (chairs, lamps, trees, or cars) in the foreground and a quasi-rectangular frame in the background. The frames were drawn to look like a window, mirror, or picture frame, and contained a pair of items identical to those in the foreground, except that they were smaller in size (see Fig. 3b–d and the online supplement). The alignment of items in the mirror matched the alignment of the items in the foreground, to facilitate their interpretation as reflections of the foreground items. In the other conditions, and particularly in the picture condition, the alignment of the items was allowed to vary. These different alignments should help preclude the framed pairs from being interpreted as reflections of the foreground pairs. Two copies of each scene were printed in color on white paper, inserted individually into clear plastic protective covers, and placed in a three-ring binder in a fixed sequence that counterbalanced the order of mirror, picture-frame, and window trials. The binder's cover displayed images of the mirror, picture frame, and window (see Fig. 4).

Procedure Participants were run individually in a quiet room, and data were collected in a single session. Participants also completed other counting tasks during this session, not discussed herein. The experimenter began each session by identifying what the "frames" on the binder's cover were intended to represent. The experimental trials were then presented in two blocks. Block 1—the *standard question* block—had 24 trials, in which the experimenter showed the
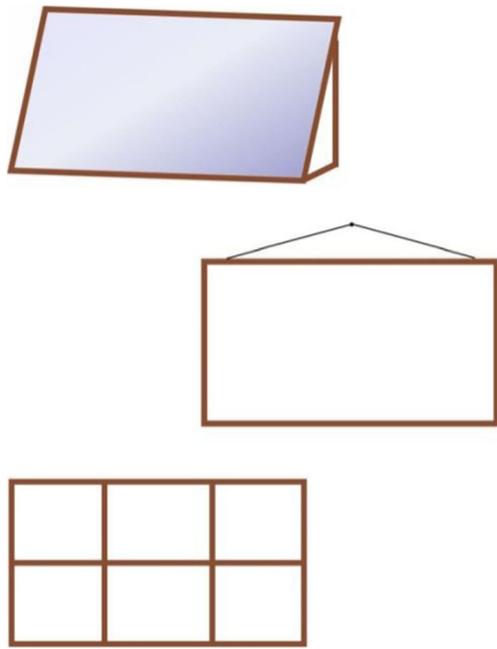
Fig. 4 Empty visual frames identified, from top to bottom, as a mirror, a picture frame, and a window

participants each scene in the binder sequentially and verbally asked a "How many_____?" question about each scene. These questions alternated between *specific questions* ("How many X?," where X corresponded to the foreground item kind—e.g., "How many cars?") and *general questions* ("How many things?"). Block 2—the *real question* block—repeated this procedure with the same 24 stimuli, except that both the specific and general questions now began "How many real . . ." (e.g., "How many real cars?," "How many real things?"). Thus, we had a 4 (item kind: chairs, lamps, trees, cars) × 3 (frame identity: mirror, picture frame, window) × 2 (specific vs. general question) × 2 (standard vs. real question) within-subjects design, with 48 trials (one trial per cell). Note that the same experimenter conducted all sessions for all four experiments presented herein.

After completing both blocks, participants were asked follow-up questions to ensure that the investigator had correctly interpreted which items were counted; for example, the experimenter would confirm that "2" referred to the foreground pair. Participants were also asked why they had decided to count those items.

Results and discussion

The "item kind" manipulation was not of theoretical interest: It was included to yield four trials per participant per cell in the 3 (frame identity) × 2 (specific vs. general question) × 2 (standard vs. real question) design. Thus, all statistics reported here are collapsed across item kinds. Figure 5 shows the rates at which participants counted all four framed and foreground

paired items for trials in each of the 12 cells of the resulting 3 × 2 × 2 design. The participants' responses within each cell showed bimodal distribution patterns: A given participant on a given kind of trial typically either always or never counted all of the paired items. Additionally, participants never counted items in mirrors in specific/real trials. This invariance made parametric analyses problematic. We therefore provide nonparametric analyses herein. However, the results from analyses of variance are provided in an online supplement.

Visual boundaries influenced counting Participants counted foreground items over 99 % of the time. In contrast, framed items were counted on only 40 % of trials (McNemar's test, $p < .001$). Even items in windows were counted less frequently than their foreground counterparts, despite these items being at the same level of referential depth (window: foreground 99 %, framed 89 %; McNemar's test, $p < .001$). Although 20 participants at some point did not count items in the window, the only instances in which foreground objects were not counted occurred for two participants who occasionally responded "zero" in the real question condition (Sign test: +18, –0, $p < .001$). Thus, all variation in paired-item counting was accounted for by the rate at which participants counted framed items.

Referential depth influenced counting Nonparametric tests demonstrated a clear rank ordering: Participants were significantly more likely to count framed items in the window condition (89 % of trials) than in the picture-frame condition (24 %), where they were likewise more apt to count framed items than in the mirror condition (7 %) (all $ps < .001$; see Table 1). Thus, it appears that referential depth indeed influenced what people counted. The items in the window had the same referential depth as those in the foreground (Level 1), and as such were counted most often. The items in the picture frame, as "pictures of pictures" (Level 2), were at a greater level of referential depth than items in the foreground, and thus were counted less frequently. The items in the mirror represented reflections of the foreground objects (Level 1'*), and thus were not only at a greater referential depth than the "real" foreground items, but also represented those same items, possibly invoking issues of double counting. Consequently, these "reflections" were counted least often.

People consciously consider referential depth when determining the countable set Participants were significantly more likely to count framed items when asked "How many ___?" (48 %) than when asked "How many *real* ___?" (33 %) ($p < .001$; see Table 1). Framed items, typically being at a greater referential depth than foreground items, are considered less "real," and thus are counted less often when the question included the qualifier "real." Moreover, this drop was greater for the mirror and picture-frame conditions, in which framed
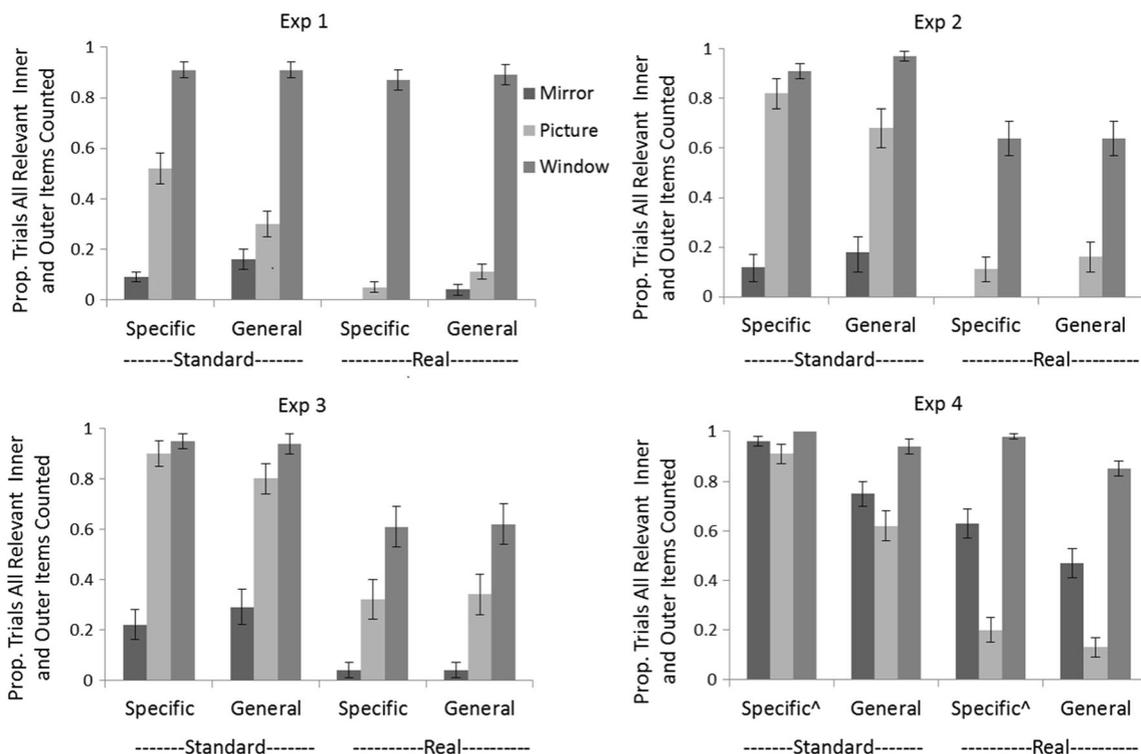
Fig. 5 Proportions of trials in which the participants counted both the items contained by the frame and (where applicable) the items in the foreground, in the various question and frame conditions for Experiments 1–4. Error bars indicate standard errors. Experiment 2 was like Experiment 1, except that the frame condition (standard vs. real) was varied between, rather than within, subjects. Experiment 3 was like Experiment 2, except that the same neutral frame was used in all frame conditions. Experiment 4 was similar to Experiment 1, except that the framed and foreground items were of different kinds, and additional trials were added so that the same number of specific questions could refer to the item kinds in each location. Note that the "specific" conditions for Experiment 4 (marked by ^) refer to trials on which participants were asked to count the kind of item that appeared inside—and not outside—the frame. Thus, the figure reflects the rates at which participants counted the two items inside the frames for these conditions, rather than all four inner and outer items. Trials on which participants were asked to count the kind of item that appeared outside the frame in Experiment 4 are not included in these proportions.

items were at a greater referential depth than the foreground items, than for the window condition, in which framed and foreground items were at the same referential depth (Fig. 5). Nonparametric analyses demonstrated that this pattern was reliable. We compared the sizes of participants' drops in counting rates between the standard and real conditions among the frame conditions. The number of participants who showed larger drops for the frame conditions at greater

referential depths was greater than the number of participants who showed the reverse pattern [picture–window, Sign test (+41, –3), $p < .001$; picture–mirror, Sign test (+36, –3), $p < .001$; mirror–window, Sign test (+27, –7), $p < .001$].

Participants counted items inside the picture frame more frequently when the question was "How many X?" (specific/standard: 52 %) than when it was "How many things?" (general/standard: 30 %). This contrasted with all five other

Table 1 Nonparametric comparison tests for Experiments 1–4

Comparison	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Mirror vs. picture	*Sign test* (+1, –40)***	$\chi^2(1) = 22.9$***	$\chi^2(1) = 25.6$***	*Sign test* (+36, –6)***
Picture vs. window	*Sign test* (+2, –56)***	$\chi^2(1) = 14.2$***	$\chi^2(1) = 5.3$*	*Sign test* (+1, –43)***
Mirror vs. window	*Sign test* (+0, –60)***	$\chi^2(1) = 61.2$***	$\chi^2(1) = 45.3$***	*Sign test* (+1, –36)***
Standard vs. real	*Sign test* (+41, –3)***	*Sign test* (+59, –4)***	*Sign test* (+50, –2)***	*Sign test* (+41, –2)***
Specific vs. general	*Sign test* (+24, –17)	*Sign test* (+20, –25)	*Sign test* (+12, –15)	*Sign test* (+38, –3)***

Experiments 1 and 4 were fully within-subjects, whereas Experiment 2 and 3 had a mixed design, with frame as the between-subjects variable. Sign test are used for within-subjects tests, whereas χ^2 s are used for between-subjects tests. χ^2 s for Experiments 2 and 3 are based on the number of participants counting at least half the specified items for all relevant trials. The *specific* condition for Experiment 4 refers to trials on which participants were asked to count the kind of item appearing inside the frame. Trials on which participants were asked to count the kind of item appearing outside the frame are not included in this analysis. * $p < .05$, *** $p < .0005$

comparisons between specific and general question conditions for the standard–real condition pairings, in which participants were as or less likely to count framed items in the specific than in the general conditions. A nonparametric test confirmed that the difference in the drops in counting rates between the specific/standard and general/standard conditions, as compared to between the specific/real and general/real conditions, was greater for the picture-frame condition than for the other frame conditions (Sign test: +31, –8, $p < .001$).

This response pattern seems somewhat paradoxical: How could there be more "cars," "trees," "lamps," or "chairs" than "things"? Explaining this, many participants who counted the framed and foreground items when asked "How many X?" (i.e., "four") instead counted the foreground items *and the picture on the wall* when asked "How many things?" (i.e., "three"). Counting the framed items therefore would have required counting both the picture on the wall and the images presented in that picture, which might be considered to violate the prohibition on double counting (see Gelman & Gallistel, 1978). We note that this drop in counting rates between the specific/standard and general/standard conditions was *only* seen when the standard questions were asked in picture-frame trials. Items "reflected" in mirrors and "seen through" windows might be less likely to invoke such issues of double counting, since they exist independently from the frame. Also, there appears to have been consensus that "pictures" are not real and should not be counted when "real" is specified, because the framed item counting rates were near floor for "real" picture-frame trials (Fig. 5).

Limitations These findings clearly demonstrate that what items adults count is influenced by referential depth. However, due to the fully within-subjects nature of Experiment 1's design, we could not determine whether these results were dependent on contrastive context, because participants were asked questions about all three frame identity conditions in the same session. We wished to determine whether the influence of referential depth seen here would persist if participants only considered one kind of frame context. In Experiment 2, we therefore replicated Experiment 1, except with a mixed model design: Frame kind was varied between, rather than within, subjects.

Experiment 2

Method

Participants A group of 109 students (63 male, 46 female; mean age = 18.8 years) at Rutgers University–New Brunswick participated for course credit. An additional three participants were excluded for experimenter error or for being younger than 18.

Stimuli In this experiment we used the same 12 unique experimental stimuli as in Experiment 1. As before, two copies of each scene were placed in a three-ring binder in fixed order, but this time they were separated by frame identity. The visual frames were identified to the participants by using the same images of an empty mirror, picture frame, and window as were used in Experiment 1. However, the participants only saw the visual frame used in the scenes that they were questioned about.

Procedure The participants were divided into three groups: 30 (19 male, 11 female) were shown the frame that looked like a mirror and were told that it was a mirror; 33 (15 male, 18 female) were shown the frame that looked like a picture frame and were told that it was a picture frame; and 46 (29 male, 17 female) were shown the frame that looked like a window and were told that it was a window. Next, the experimenter sequentially presented the assigned eight scenes, alternately asking "How many X?" (specific/standard) and "How many things?" (general/standard), one question per scene. The experimenter than repeated this process with the same set of pictures, alternatively asking "How many real X?" (specific/real) and "How many real things?" (general/real). Thus, each participant was asked 16 questions total in two blocks of eight trials. After completing both blocks, the participants were asked follow-up questions as in Experiment 1.

Results and discussion

As in Experiment 1, we used nonparametric analyses that collapsed across item type. Our findings replicated those of Experiment 1. Participants were more likely to count foreground items (96 %) than framed items (44 %) (Sign test: + 76, –0. $p < .001$). This was true even in the window condition, in which the difference was smallest (91 % vs. 79 %; Sign test: +17, –0, $p < .001$). With the exception of one participant in the window condition, who indicated that only the "outside" items counted (i.e., those seen through the window), and one participant who counted "the paper" as the only real item, all variation in foreground item counting was due to some participants responding "zero." Thus, almost all variation in the rate at which participants counted all of the paired items was accounted for by differences in the rates at which they counted the items inside the frames. Note that "zero" responses were only seen in the "real" conditions. Of interest, "zero" responses were more likely to occur when participants considered the framed and foreground items to be at the same level of referential depth, and consequently, the qualifier "real" should include or exclude all of these items [real/window 15 % of trials; real/picture 7 %; real/mirror, 1 %; $\chi^2(1)$ range = 11.1–35.6, all ps < .001]. Likely, this response pattern was seen here and not

in Experiment 1 because conversational pragmatics implied that participants' answers should change when the "real" qualifier was added. Lacking contrast between the frame conditions and contrast in the referential depth of the framed and foreground items, participants may have judged the items' referential depth in terms of the real world (i.e., the participant was at Level 1, and thus the stimuli would be at Level 2). Thus, they replied "zero," that none of the items were real.

Participants were significantly more likely to count all framed and foreground items in the window condition (79 %) than in the picture-frame condition (44 %), where they were likewise more apt to count framed items than in the mirror condition (8 %) (all $ps < .001$; see Table 1). Participants were significantly more likely to count all of the framed and foreground items when the question was "How many ___?" (67 %) rather than "How many real__?" (31 %) ($p < .005$; see Table 1). This difference was largest for the picture-frame condition (see Fig. 5). This frame identity by standard–real interaction was confirmed to be significant by nonparametric tests on the numbers of participants showing a drop in counting rates between the standard and real conditions [picture-frame vs. window, $\chi^2(1) = 13.0$, $p < .001$; picture vs. mirror, $\chi^2(1) = 10.0$, $p = .002$; mirror vs. window, $\chi^2(1) = 0.0$, $p = .861$]. As in Experiment 1, the drops in counting rates seen between the specific and general condition were present only for the standard-picture trials (see Fig. 5). A nonparametric test demonstrated that this pattern was reliable. For each frame condition, we found that the proportion of participants who showed a greater drop in counting rate between the specific/standard and general/standard conditions than between the specific/real and general/real conditions. This proportion was greater for the picture-frame condition than for the other groups [$\chi^2(1) = 11.5$, $p < .001$]. No other effects were detected.

Limitations These data replicated the findings of Experiment 1, demonstrating that the different count rates seen between the mirror, picture-frame, and window conditions were not dependent on contrastive context. However, one might still argue that these differences arose not at the conceptual level—from differences in referential depth—but at the perceptual level. The stimuli in the three frame identity conditions were visually different, in both the appearance of the frame and the alignment of the items. We wished to know whether these counting rates would differ in the same fashion if participants were merely verbally prompted to interpret a neutral frame as a mirror, picture frame, or window, ensuring that all differences were conceptually based. In Experiment 3, we therefore replicated Experiment 2, except that we used the same visual stimuli in all three frame identity conditions: two large items below and two small items inside a plain brown rectangle (see Fig. 3a).

Experiment 3

Method

Participants A total of 100 students (60 male, 40 female; mean age = 19.1 years) at Rutgers University–New Brunswick participated for course credit. Three additional participants were excluded for lack of English fluency or for being younger than 18.

Stimuli In this experiment, we used four unique experimental stimuli, consisting of a pair of identical items (chairs, lamps, trees, or cars) in the foreground of the image, with a plain brown rectangular frame in the background (see Fig. 3a and the online supplement). Two copies of each image were printed in color on white paper, inserted individually into clear plastic protective covers, and placed in a three-ring binder. The binder's cover displayed the brown rectangle.

Procedure The participants were divided into three groups. All were shown the plain brown rectangle, but its label varied: 34 (20 male, 14 female) were told that it was a mirror, 33 (21 male, 12 female) were told that it was a picture frame, and 33 (19 male, 14 female) were told that it was a window. The procedure in all other respects was identical to that of Experiment 2, except that the visual frame was always a plain brown rectangle.

Results and discussion

As in Experiments 1 and 2, we used nonparametric analyses that collapsed across item type. Our findings replicated those of Experiments 1 and 2. Participants were more likely to count foreground items (94 %) than framed items (51 %) (Sign test, + 61, –0, $p < .001$). This was true even in the window condition, in which the difference was smallest (92 % vs. 78 %; Sign test: +10, –0, $p = .002$). With the exception of one participant in the mirror condition, who indicated having initially thought that participants were only supposed to count items "in the box," all of the variation in foreground item counting was due to some participants responding "zero." Thus, almost all variation in the rate at which participants counted all of the paired items was accounted for by framed-item counting rates. We note that, except for two instances in the standard picture-frame condition, "zero" responses were only seen in the real-question conditions. As in Experiment 2, "zero" responses were seen more often when there was less contrast in referential depth between the framed and foreground items [real window, 11 %; real picture, 9 %; real mirror, 6 %; window vs. mirror: $\chi^2(1) = 6.0$, $p = .015$; other comparisons n.s.].

Participants were significantly more likely to count all framed and foreground items in the window condition (78 %) than in the picture-frame condition (59 %), where they were more likely to count framed items than in the mirror condition (15 %) (all $ps < .001$; see Table 1). Participants were also

significantly more likely to count framed items when the question was "How many ___?"(68 %) than when it was "How many real __?" (33 %). As before, this drop was largest for the picture-frame condition. This interaction was confirmed to be significant by nonparametric tests on the number of participants showing a drop in counting rates between the standard and real conditions [picture vs. window, $\chi^2(1) = 4.9$, $p = .026$; picture vs. mirror, $\chi^2(1) = 3.4$, $p = .064$; mirror vs. window, $\chi^2(1) = 0.2$, $p = .695$]. As before, only for the standard picture-frame trials was a drop in counting rate seen between the specific and general conditions (see Fig. 5). A nonparametric test demonstrated this pattern was marginally reliable. For each frame condition, we found the proportion of participants who showed a greater drop in counting rate between the specific/standard and general/standard conditions than between the specific/real and general/real conditions. This proportion was greater for the picture-frame condition than for the other groups [$\chi^2(1) = 3.3$, $p = .069$]. No other effects were detected.

Limitations These results clearly demonstrated that the influence of frame identity on counting rates was due to conceptually based differences in referential depth: Framed items were counted less often when the frame was identified as a mirror or picture frame than when it was identified as a window, despite the fact that the exact same visual stimuli were presented in these conditions. However, all of the experiments discussed so far were limited in that counting all four of the foreground and framed items required crossing the visual boundary created by the frame. Counting across *conceptual* levels of referential depth, such as by counting framed and foreground items in the mirror and picture-frame conditions, might have decreased the rate at which "less real" items were counted. Additionally, crossing the *perceptual* boundary of the frame may have contributed to the relative lack of framed-item counting. If so, one should expect that eliminating the need to cross these boundaries would result in higher framed-item counting rates.

In Experiment 4, we investigated this possibility by placing different kinds of items inside and outside the now familiar frames. This allowed the interior and exterior items to be separately specified by the questions. Thus, we additionally varied whether the specific-question trials referred to the kind of item inside or outside the frame, yielding a 3 (frame) × 3 (specific/framed, specific/foreground, general) × 2 (standard vs. real) within-subjects design.
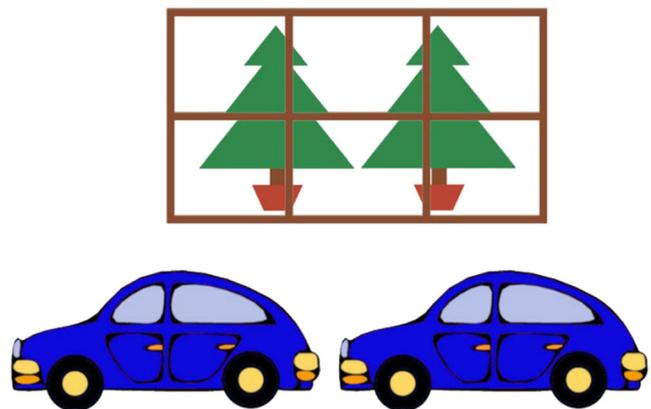
Experiment 4

Method

Participants A group of 46 (25 male, 21 female; mean age = 18.5 years) students at Rutgers University–New Brunswick

participated for course credit. Three additional participants were excluded for being younger than 18.

Stimuli In this study, we used the same introductory display showing the three frames that had been used in Experiment 1 (see Fig. 4). We also created 36 unique experimental displays of the same general composition as those used in Experiment 1, with one crucial difference: The matched pair of items in the foreground was *never* of the same kind as the matched pair inside the frame (see Fig. 6 and the online supplement).

Procedure Participants were assigned to one of three order conditions, and viewed different stimulus orders in the standard and "real" question blocks. Each participant answered 96 questions. Otherwise, the procedure was similar to that used in Experiment 1, though modified so that the specific-question conditions alternately referred to the framed and foreground items. The experimenter began each session by identifying the empty frames on the binder's cover. The experimenter then presented the scenes one at a time, in one of three orders, asking one "How many _____?" question per scene. Over the first 48 scenes (the standard question block), the experimenter first asked "How many X?," where "X" referred to the kind of item inside the frame (specific/framed/standard; e.g., "How many trees?" for Fig. 6); then "How many things" (general/standard); then "How many X?," where "X" referred to the kind of item in the foreground (specific/foreground/standard; e.g., "How many cars?" for Fig. 6); then "How many things?" (general/standard) once again. The experimenter cycled through these four questions 12 times, yielding 12 specific/framed/standard trials, 12 specific/foreground/standard trials, and 24 general/standard trials. As was the case in Experiments 1–3, the questioning was much the same for the next 48 trials (the "real" block), except that the qualifier "real" was added to the questions. This yielded 12 specific/framed/real

Fig. 6 Example stimulus scene used in Experiment 4: Two trees framed by a window, with two cars in the foreground. A participant could be asked about the specific items inside the frame (e.g., "How many trees?"), the specific items outside the frame (e.g., "How many cars?"), or things in general (e.g., "How many things?")

trials, 12 specific/foreground/real trials, and 24 general/real trials. The question kinds were evenly distributed among the three frame conditions, and the first 24 stimuli in the standard block used scenes identical to the last 24 stimuli used in the "real" block. After the participants had completed both blocks, they were asked follow-up questions as in Experiment 1.

Results and discussion

The data were collapsed across the various item kind and order conditions, and nonparametric analyses were used. As in Experiments 1–3, framed items were counted less often than foreground items. This was made unambiguous by the difference in the rate at which participants counted specific items outside the frame (>99 %) and the rate at which they counted specific items inside the frame (78 %) (Sign test: +42, –0, $p < .001$). In contrast to the prior experiments, the different rates of counting specific framed and foreground items did not extend to the window condition (>99 % vs. > 99 %; Sign test: +2, –0, $p = .500$). This indicated that, as suspected, counting across the visual boundary might have decreased the rate at which framed items were counted in Experiments 1–3. Note that all of the variation in whether participants counted items outside the frame in the relevant conditions was due to two participants who occasionally gave "zero" responses in the "real" condition.

In order to best interpret the results of this experiment, all other analyses excluded the specific/foreground trials-trials in which participants were asked to count the kinds of items outside the frame. This best paralleled the analyses in Experiments 1–3, since the rate at which participants counted all paired items in those experiments had been almost entirely accounted for by framed-item counting rates. Our analyses focused on the rate at which participants counted all *relevant* paired items—that is, items inside the frames for specific/framed trials, and both the foreground and framed pairs for general trials. These proportions are presented in Fig. 5.

As before, referential depth influenced what was counted. Participants were significantly more likely to count the framed pair in the standard (86 %) than in the real (54 %) question conditions ($p < .001$; see Table 1). Frame condition also influenced what was counted: Items inside windows (94 %) were counted significantly more often than were those in mirrors (70 %), which in turn were counted significantly more often than those in picture frames (46 %) (all $ps < .001$; see Table 1). Note that the rank ordering of the mirror and picture-frame conditions here was the reverse of those found in Experiments 1–3. The ratio of participants who counted items in mirrors more often than items in picture frames to participants showing the reverse pattern (mirror > picture, 36; picture > mirror, 6) was 240 times that same ratio seen in Experiment 1 (mirror > picture, 1; picture > mirror, 40) [$\chi^2(1) = 58.1$, $p < .001$]. Recall that items in picture frames represent pictures in the

portrayed scene (Level 2), whereas items in mirrors represent reflections of "real" items (Level 1'). In this experiment, unlike in Experiments 1–3, the "real" items that these reflections referred to were *not* represented elsewhere in the stimuli, so the issue of double counting was not invoked. One might consider these reflections more "real" than the "pictures," since they uniquely referred to "real" objects. Indeed, the largest drop in counting rates between the standard and "real" question conditions was seen in the picture-frame condition, despite the mirror and window conditions having "farther to fall." Nonparametric tests comparing the sizes of the drops in counting rates between the standard and "real" conditions among the various frame conditions demonstrated that this interaction was significant (picture–window, Sign test: +41, –3, $p < .001$; picture–mirror, Sign test: +34, –7, $p < .001$; mirror–window, Sign test: +26, –10, $p < .05$).

Another divergence of this experiment's results from those of Experiments 1–3 was a significant effect of the specific/framed versus the general question condition (Fig. 5). Items inside the frame were counted significantly more often when that kind of item was specified (78 %) (e.g., "How many trees?") than when general "things" questions were asked (63 %). The ratio of participants who counted more framed items in the specific/framed than in the general condition to those who showed the reverse pattern (specific > general, 38; general > specific, 3) was greater here than the parallel ratios in any of the other experiments [range $\chi^2(1)$s: 13.1–22.7, all $ps < .001$].

Since foreground items were nearly always counted in the general condition (>99 %), the specific/framed versus general manipulation here doubled as a manipulation of whether counting framed items would (general condition) or would not (specific/framed condition) require one to count across the perceptual boundary of the frame, and thus across the different levels of referential depth that the frame might impart. This difference was significant for the window condition, in which items were at the same referential depth (99 % vs. 90 %, Sign test: +18, –1, $p < .001$). This further supported the hypothesis that crossing the perceptual frame reduced the rate at which framed items were counted, regardless of referential depth. However, we also noted that this difference was larger for the mirror and picture-frame conditions than for the window condition, indicating that the need to cross the *conceptual* boundary and count items at different levels of referential depth was more likely to reduce counting than was crossing physical frame boundaries. Nonparametric tests demonstrated that this interaction was significant. We compared the number of participants showing greater drops in counting rates between the specific/framed and general conditions for the frame conditions that imparted greater referential depth to the number of participants showing the opposite pattern. This difference was significant when contrasting the mirror and window conditions, but was a nonsignificant trend when comparing the picture and window conditions (mirror–window: Sign test,

+24, –10, p = .024; picture–window, Sign test: +24, –14, p = .143; picture–mirror, Sign test: +18, –23, p = .533).

Note that there was a drop in counting rates between the specific/framed and general conditions for all six frame/standard–real conditions; in Experiment 1–3, this drop was only seen in the standard picture-frame condition. However, the drop in Experiment 4 was *largest* for the standard picture-frame condition, paralleling the three-way interaction seen in Experiments 1–3. The same test used in Experiment 1 demonstrated that this pattern was reliable: The difference in the drops in counting rates between the specific/standard and general/standard conditions, relative to between the specific/real and general/real conditions, was greater for the picture-frame condition than for the other frame conditions (Sign test: +31, –9; p < .001).

General discussion

These experiments addressed three questions: (1) Does referential depth influence what adults count? (2) Are adults conscious of referential depth when determining what to count? (3) Do visual boundaries influence what adults count? Our data indicated that the answer to all of these questions is "yes."

Referential depth influences what adults count

In all four experiments, items at greater levels of referential depth were counted less frequently. Participants were less inclined to count items "in" a mirror (Level 1' or 1'*) or "on" a painting (Level 2) than items "seen through" a window (Level 1). This effect was quite robust, impacting participants' counts even when the frame conditions were not contrasted (Exps. 2 and 3) and when there was no need for the counts to cross visual boundaries (Exp. 4). This held even when the frames were visually identical, only differentiated by an initial verbal label (Exp. 3). It did not matter that the "mirror," "picture frame," or "window" was a simple, undifferentiated rectangle; the kind of object symbolized by the rectangle was what influenced participants' counts, not that symbol's instantiation.

Adults can consciously consider referential depth when counting

It appears that adults are aware of referential depth as a factor in their counting decisions. Verbal contexts that emphasized "realness" as a counting criterion shifted counting behavior: "Pictures" and "reflections" were counted less often when the qualifier "real" was added to "How many ____?" Even participants in the window conditions of Experiments 2 and 3—who did not see the mirror and picture-frame conditions for contrastive context—displayed changes in counting behaviors when "real" was specified. Some participants shifted from counting all of the framed and foreground items to counting none of them, demonstrating both that they understood that both the framed and foreground items represented "real" items within the context of the stimulus scene, and that these images were equally nonreal, in that they were all drawings.

Visual boundaries influence what adults count

In all four experiments, participants were less likely to count the framed items than the foreground items, even when the items were at the same level of referential depth. Experiment 4 demonstrated explicitly that the framed items were counted less often when foreground items might also be included in the count. We concluded framed-item counting was partially inhibited by resistance to constructing sets across visual boundaries. We note that the frame was not the only visual feature that separated the framed and foreground items. Framed items were also differentiated from the foreground items by size, location, and—in Experiment 4—shape, invoking Gestalt grouping features of similarity and proximity (Rock, 1984) to establish the framed and foreground items as separate sets. Consequently, we cannot conclude that crossing the frame boundary per se inhibited framed-item counting. Nevertheless, the more general hypothesis—that people disprefer constructing countable sets that cross perceptually established sets—was clearly supported.

Implications

These results offer important insight regarding how adults decide what "counts" in a given context. Although the definition of "countable items" is extremely broad, it is not so broad as to violate basic counting principles. When items in mirrors are interpreted as reflections of counted foreground items (Level 1'*), most people do not count them. Although it may be considered legitimate to count foreground "cars" and framed "reflections of cars," it is not legitimate to count the cars *referenced by* the reflections when those reflections reference already-counted foreground cars. Doing so would violate the one-to-one principle (Gelman & Gallistel, 1978). This is highlighted by the low rate of counting such Level 1'* reflections Experiments 1–3 as compared to Experiment 4, in which the Level 1' reflections instead referenced off-screen items. This highlights an important distinction between "reflections" and "pictures" in these scenes. Reflections, by definition, are caused by the interaction of an object and a reflective surface (i.e., a mirror). However, entities drawn on a "picture," even when they are copies of foreground items, are not portrayed as having been caused by them (cf. DeLoache, 2000). Thus, in contrast to pictures, whether reflections "count" is partially determined by whether one also counts what they are reflecting.

Expanding on this point, what our participants counted provides insight into what they thought "counted" in a given scene. Our participants were quite adept in their ability to take the subtle pragmatics of a given setting and description into account when establishing the countable set. This was reflected in how the various context changes over trials and experiments were matched by shifts in what was counted. It appears that all of these adjustments in interpreting items' referential depth happened online, quickly, and without effort. Individuals readily integrated different kinds of labels for the same neutral item (i.e., the rectangle's label) and behaved in ways consistent with these on-the-spot assignments. Similarly, individuals' could switch between the general term "thing" and a particular term (e.g., "cars") and easily applied a delimiter regarding existential status ("real") to two-dimensional drawings.

These results have implications for many areas of research. "Framing problems"—determining what particular elements are relevant in particular contexts—have long posed a puzzle to those interested in cognition (Pylyshyn, 1996; see also Gelman, 1969). As we have seen here, what a person literally counts can indicate what "counts" in a given scene. Indeed, a common research tactic is to have individuals count the items in displays and to treat the counted set as directly indicative of what individuals mentally "counted" as relevant. For example, participants' counts of "blocks" composing target items have been used to study object segmentation (Bravo & Farid, 2003). Although counting is a useful investigative tool, it is important to note that counting itself is not immune to contextual ambiguities. As Experiment 3 demonstrated, individuals can draw different conclusions about what "counts" in perceptually identical scenes, given different verbal cues: Chairs seen "through" windows may count, whereas chairs seen "in" mirrors do not. This occurs despite there being many similarities in the ways that mirrors and windows are perceived. People are similarly accurate at perceiving the sizes of "real" objects seen through windows or in mirrors, and similarly inaccurate when estimating the size of reflections appearing on mirrors and the sizes of images as they pass through windows (Bertamini, Lawson, & Liu, 2008; Lawson, Bertamini, & Liu, 2007). Thus, frame identity impacts counting by providing conceptual information regarding items' referential depth, rather than by altering how perceptual cues are processed.

Further research will be needed to determine the extent to which decisions as to what "counts" are automatic and intuitive, and to what extent they are based on deductive reasoning. It may be fruitful for future studies to investigate children's developmental trajectories for sensitivity to the conceptual and perceptual variables that impact adult counting behavior. For example, the impact of visual boundaries on what is counted may be seen early in developmental trajectories, since it seems to arise from core perceptual processes. In contrast,

one can predict that the effect of referential depth on counting will emerge relatively late, since it requires understanding of symbolic representations. Young children can have specific difficulties understanding the relationships between symbols and their referents (DeLoache, 2000; DeLoache et al., 1997, 2003). Pilot data supporting these hypotheses have been collected.

Limitations

Although the present study indicated that people are conscious of referential depth when deciding what to count, we did not investigate whether people were conscious of the influence of visual boundaries on these decisions. It maybe that people are more aware of the impact of referential depth on counting judgments than they are of the impact of visual boundaries. Participants who counted foreground but not framed items in the window condition commonly justified their choice by claiming that the framed items were "reflections off the glass," invoking referential depth. Other answers were more locative, such as "I thought you meant the ones inside [the building]," but even those cases seemed to focus on location within the depicted "world" rather than the drawn boundary itself. Further study will be needed to determine whether this is the case.

Additionally, the binary coding used herein to indicate whether participants counted framed and foreground pairs did not fully capture the range of responses. Notably, this coding obscures whether participants counted the frame itself—that is, whether they counted the window, picture frame, or mirror when asked to count "things." Interestingly, the frames were counted less often than foreground items, even though these items were at the same level of referential depth. We consider this result worthy of further investigation, particularly because perceptual boundaries influence both counting behavior (as demonstrated here) and bias estimation (Chesney & Gelman, 2012).

Additionally, whereas the vast majority of participants limited their counts to the framed and foreground items and the frame itself, there were a few "creative counters." For example, two participants responded "seven" when asked "How many things?" in Experiments 4's mirror conditions. Their counts included the two foreground items, the mirror, the two "reflections," and the two objects "off screen" that those reflections referenced. Such creative counts may be worthy of further consideration, particularly because their rarity demonstrates how closely adults adhere to a narrow range of typical countable sets.

Conclusions

Gelman and Gallistel's (1978) counting principles allow people to count any set of individual things. Nevertheless, what

people typically count appears to be quite constrained. For example, all of the participants in these four experiments at some point counted the foreground images, but none ever counted the plastic sheet protectors that they had to peer through to view the printed scenes. Consistent patterns in how conceptual and perceptual contexts influenced the counting judgments were seen across four experiments. These results demonstrate that both visual boundaries and referential depth impact what adults count. We hold no illusions that this account is complete. Future investigation will be necessary to determine what other factors may influence what "counts." We believe that such investigations have the potential to offer insight in many diverse fields, including attention, symbolic understanding, and pragmatics.

References

Allen, R., & McGeorge, P. (2008). Subitizing: Shape information and expertise. *Acta Psychologica, 129,* 26–31. doi:10.1016/j.actpsy.2008.04.003

Alston, L. W., & Humphreys, G. W. (2004). Subitization and attentional engagement by transient stimuli. *Spatial Vision, 17,* 17–50. doi:10.1163/156856804322778251

Bertamini, M., Lawson, R., & Liu, D. (2008). Understanding 2D projections on mirrors and on windows. *Spatial Vision, 21,* 273–289. doi:10.1163/156856808784532527

Bravo, M. J., & Farid, H. (2003). Object segmentation by top-down processes. *Visual Cognition, 10,* 471–491. doi:10.1080/13506280244000177

Chesney, D. L., & Gelman, R. (2012). Visual nesting impacts approximate number system estimation. *Attention, Perception, & Psychophysics, 74,* 1104–1113. doi:10.3758/s13414-012-0349-1

Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 16–45). New York, NY: Sage.

Craine, T. V. (1994). Activities: Counting embedded figures. *Mathematics Teacher, 87,* 524–528.

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development, 71,* 329–338.

DeLoache, J. S., Miller, K. F., & Rosengren, K. S. (1997). The credible shrinking room: Very young children's performance with symbolic and non-symbolic relations. *Psychological Science, 8,* 308–313. doi:10.1111/1467-8624.00148

DeLoache, J. S., Pierroutsakos, S. L., & Uttal, D. H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science, 12,* 114–118. doi:10.1111/1467-8721.01244

Gelman, R. S. (1969). Conservation acquisition: A problem of learning to attend to relevant attributes. *Journal of Experimental Child Psychology, 7,* 167–187. doi:10.1016/0022-0965(69)90041-1

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Giralt, N., & Bloom, P. (2000). How special are objects? Children's reasoning about objects, parts, and holes. *Psychological Science, 11,* 497–501. doi:10.1111/1467-9280.00295

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103,* 19599–19604. doi:10.1073/pnas.0609485103

Lawson, R., Bertamini, M., & Liu, D. (2007). Overestimation of the projected size of objects on the surface of mirrors and windows. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1027–1044. doi:10.1037/0096-1523.33.5.1027

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111,* 1–22. doi:10.1037/0096-3445.111.1.1

Markman, E. M. (1979). Classes and collections: Conceptual organization and numerical abilities. *Cognitive Psychology, 11,* 395–411. doi:10.1016/0010-0285(79)90018-5

Pylyshyn, Z. (1996). The frame problem blues: Once more, with feeling. In K. M. Ford & Z. W. Pylyshyn (Eds.), *The robot's dilemma revisited: The frame problem in artificial intelligence* (pp. xi–xvii). Norwood, NJ: Ablex.

Rock, I. (1984). *Perception.* New York, NY: Scientific American Library.

Shipley, E. F., & Shepperson, B. (1990). Countable entities: Developmental changes. *Cognition, 34,* 109–136.

Trick, L. M., & Enns, J. T. (1997a). Clusters precede shapes in perceptual organization. *Psychological Science, 8,* 124–129. doi:10.1111/j.1467-9280.1997.tb00694.x

Trick, L. M., & Enns, J. T. (1997b). Measuring preattentive processes: When is pop out not enough? *Visual Cognition, 4,* 163–198. doi:10.1080/713756754

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 331–351. doi:10.1037/0096-1523.19.2.331

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review, 101,* 80–102. doi:10.1037/0033-295X.101.1.80

van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots. *Memory & Cognition, 10,* 396–404. doi:10.3758/BF03202432

Vos, P. G., van Oeffelen, M. P., Tibosch, H. J., & Allik, J. (1988). Interactions between area and numerosity. *Psychological Research, 50,* 148–154. doi:10.1007/BF00310175

Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition, 90,* 163–191. doi:10.1016/S0010-0277(03)00143-4

Witkin, H. A., Moore, C. A., Goodenough, R. D., & Cox, P. W. (1977). Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research, 47,* 1–64.