Graspable Objects Grab Attention
More Than Images Do

Michael A. Gomez¹, Rafał M. Skiba¹,², and
Jacqueline C. Snow¹
¹Department of Psychology, University of Nevada, Reno, and ²Department of Neuroscience,
University of Geneva

Abstract
The opportunity an object presents for action is known as an affordance. A basic assumption in previous research was that images of objects, which do not afford physical action, elicit effects on attention and behavior comparable with those of real-world tangible objects. Using a flanker task, we compared interference effects between real graspable objects and matched 2-D or 3-D images of the items. Compared with both 2-D and 3-D images, real objects yielded slower response times overall and elicited greater flanker interference effects. When the real objects were positioned out of reach or behind a transparent barrier, the pattern of response times and interference effects was comparable with that for 2-D images. Graspable objects exert a more powerful influence on attention and manual responses than images because of the affordances they offer for manual interaction. These results raise questions about whether images are suitable proxies for real objects in psychological research.

Keywords
real-world objects, affordance, attention, 2-D images, stereoscopic images

Received 6/24/16; Revision accepted 8/15/17

Natural environments contain many different objects, some of which are relevant for attention and some of which are not. Our ability to direct attention to relevant objects may be influenced by the opportunities those objects present for action, sometimes referred to as affordances (J. J. Gibson, 1979). Current models of attention posit that affordances are learned via associative relations between object properties and our actions toward them and that these high-level associations, in turn, trigger visuo-motor responses that modulate attention and behavior (Humphreys et al., 2013). Laboratory studies using images of graspable objects generally support this idea by showing that objects, such as tools, that have strong associations with action can be processed automatically (Makris, Grant, Hadar, & Yarrow, 2013), influence attentional selection (Handy, Grafton, Shroff, Ketay, & Gazzaniga, 2003; Riddoch, Humphreys, Edwards, Baker, & Willson, 2003) and manual response times (Masson, Bub, & Breuer, 2011; Tucker & Ellis, 2001), and can stimulate neural activity in the dorsal motor (Grezes & Decety, 2002; Lewis, 2006) and ventral perceptual-processing streams (Roberts & Humphreys, 2010).

To date, however, empirical studies have relied almost exclusively on 2-D images of objects to investigate affordance effects on attention. There are several potential problems with this approach. Theoretically, the notion that images are appropriate proxies for real objects diverges from classic theories of affordances, which were envisioned in the context of real objects (J. J. Gibson, 1979). Ontologically, the practice of using images to study affordances overlooks the fact that images do not afford action, yet having the potential to physically interact with an object could have a powerful

Corresponding Authors:
Michael A. Gomez, The University of Nevada, Reno, Department of Psychology, 1664 N. Virginia St., Reno, NV 89557
E-mail: michaelgomez@unr.edu
Jacqueline C. Snow, The University of Nevada, Reno, Department of Psychology, 1664 N. Virginia St., Reno, NV 89557
E-mail: jacqueline.c.snow@gmail.com
influence on attention. From an evolutionary perspective, the human brain has presumably evolved to allow us to perceive and interact with real objects and environments (Cisek & Kalaska, 2010; Heft, 2013), and images may therefore be an atypical class of stimuli with which to characterize mechanisms of naturalistic vision. Indeed, developmental psychologists have long recognized that images are abstract representations that must be learned to be fully understood (DeLoache, Pierroutsakos, Uttal, Rosengren, & Gottlieb, 1998) and have underscored the importance of physical exploratory behavior for normal perceptual development (E. J. Gibson, 1988).

There is a surprising paucity of studies that have examined whether real objects that afford physical actions are processed or represented differently than representations of objects. In humans, whereas repetition of 2-D images or photographs of objects leads to a characteristic reduction in functional MRI (fMRI) responses (fMRI adaptation), the same effect is weak if not absent for real objects (Snow et al., 2011). Priming effects may not decay over time for real objects, as is often observed with pictorial primes (Squires, Macdonald, Culham, & Snow, 2016). In addition, real objects are more memorable than matched photographs or line drawings of the same items (Snow, Skiba, Coleman, & Berryhill, 2014), are more easily recognized by agnosia patients (Chainay & Humphreys, 2001; Humphrey, Goodale, Jakobson, & Servos, 1994), and may be perceived as being more valuable and satiating than images (Romero, Compton, Yang, & Snow, in press). Importantly, when we look at real objects with two eyes, each eye receives information from a slightly different horizontal vantage point. This discrepancy between the different images (known as binocular disparity) is resolved by the brain to produce a unitary sense of depth. Three-dimensional stereoscopic images of objects can be generated using active shutter glasses that present two offset images separately to the left and right eye. Stereoscopic images are more similar to real objects than 2-D images are because they provide richer information about the shape, size, and distance of the depicted object. However, neither 2-D nor 3-D stereoscopic images afford genuine physical interaction. Differences in neural responses to stereoscopic versus real environments have been revealed in the rat brain (Aghajani et al., 2015).

In the experiments reported here, we investigated whether real objects exert a stronger competitive influence on attention and manual responses compared with computerized images of objects. We compared interference effects for real objects with interference effects for 2-D images (Experiment 1) and 3-D stereoscopic images (Experiment 2). We measured the influence of display format on behavior using a classic flanker paradigm (Eriksen & Eriksen, 1974). When multiple objects offer the potential for interactive behavior, a number of motor-action alternatives are computed simultaneously and compete for selection (Eriksen, 2007). Resolution of the competition between different action alternatives produces a processing cost that slows response times (RTs; Jax & Buxbaum, 2010). In the flanker task, observers take longer to respond to a central target object when it is flanked by distractor objects that would elicit a different (i.e., incongruent) versus similar (i.e., congruent) manual response (Eriksen & Eriksen, 1974). The magnitude of this interference effect reflects the extent of processing of the to-be-ignored distractors. A critical test of current affordance-based models of attention (Eriksen, 2007; Cisek & Kalaska, 2010; Humphreys et al., 2013) is to determine whether real graspable objects trigger more robust action plans and therefore compete more strongly for attention and manual responses compared with their images. We predicted that participants would take longer to respond on real-object trials than on image trials because the motor plan triggered by the stimulus (i.e., grasping) conflicts with the required motor response on the task (i.e., pressing a button). Second, interference from irrelevant flankers should be greater for the real objects than for the image displays. If the effect of real objects on attention is not attributable to the presence of stereo disparity, then any differences in flanker interference between real objects and 2-D images should also be apparent in the context of 3-D stereo images. Conversely, if real objects are processed differently from images because they afford grasping (Gallivan, Cavina-Pratesi, & Culham, 2009; Iriki, Tanaka, & Iwamura, 1996; Mountcastle, Lynch, Georgopoulos, Sakata, & Acuna, 1975), then their unique effects on attention should disappear when they are positioned beyond reach (Experiment 3) or within reaching distance but behind a transparent barrier that prevents access to the objects (Experiment 4).

Experiment 1

Method

Participants. Forty undergraduate students (32 females; mean age = 20.75 years, SD = 4.27), all of whom were enrolled in an undergraduate research-methods class, participated in Experiment 1 in exchange for course credit. All participants reported having normal or corrected-to-normal vision and were right-handed. All participants provided informed consent, and the protocols used were approved by the University of Nevada, Reno Institutional Review Board (IRB). Using G*Power (Version 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007), we estimated that with
a total sample of 41 participants, we would have 80% power to detect a medium effect, \( d_z = 0.45 \), given a two-tailed test and an \( \alpha \) level of .05. Therefore, we targeted a sample size of approximately 40 participants, with final numbers reflecting the number of students who sought research-participation credit during the semester.

**Stimuli and apparatus.** In Experiment 1, we compared target RT and error rates elicited by stimuli in two different display formats: real objects and 2-D images of the same items. The stimuli in the real-object displays were three white plastic spoons (Figs. 1a and 1b). The spoons measured 12.2 cm from tip to handle and 2.6 cm at the widest point, and at a viewing distance of 60 cm, each spoon subtended 13.7° × 2.9°. One spoon (the target) was centered at the vertical and horizontal midpoint of the display. The flankers, which were also centered along the horizontal midline of the display, were positioned 3.66° above and below the target. The spoons were mounted on a black vertically oriented display board made of composite wood material. The display board was 62.5 × 37.0 cm in size (63.2° × 40.3°), matching the outer dimensions of the LCD monitor used on the image trials. The spoons were held in position using magnets, which were fixed to both the convex side of the spoon and the rear surface of the display board. Small adhesive tape
markers were positioned on the display board to ensure accurate stimulus alignment. The markers were occluded by the spoons when they were mounted and thus were not visible to participants during the trials.

On half of all trials, the target spoon was oriented with the handle facing rightward, and on the remaining trials, the handle was oriented leftward. By virtue of the relative orientation of the target and the flankers, we manipulated target-flanker incongruence. On half of the trials, the flankers were oriented so that their handle faced the same direction as the target (congruent-flanker trials); on the remaining trials, the flankers were oriented with their handles facing the opposite direction from the target (incongruent-flanker trials). The combination of each target handle orientation (left or right) and flanker handle orientation (left or right) yielded a total of four unique display configurations, which were used to generate stimuli for the image displays. A separate LCD monitor, positioned behind the participant, was used to display an image of the stimulus configuration required for the upcoming trial. The testing room was illuminated from above with incalceous fluorescent lights.

Stimuli in the image displays consisted of high-resolution color photographs of the stimuli in each of the four real-object configurations. We used a Canon Rebel T2i digital single-lens reflex (DSLR) camera with constant f-stop and shutter speed to photograph the real-object displays. The stimuli were photographed separately for each display configuration, under identical lighting conditions. Image size was adjusted using Adobe Photoshop so that the stimuli in the 2-D images were matched in size to the real spoons. Apart from resizing, the images were not cropped or otherwise adjusted, and the resulting stimuli preserved other monocular depth cues, including shadows and specular highlights that were present in the real-object displays. The images were presented on a 62.5 \( \times \) 37.0 cm, 27-in. LCD monitor (resolution: 1,920 \( \times \) 1,080 pixels). The timing of events was controlled by a PC (Intel Core i7-4770 CPU 3.4 GHz, 16-GB RAM; Quadro graphics processing unit, NVIDIA, Toronto, Ontario, Canada) running MATLAB (The MathWorks, Natick, MA). Stimulus viewing time in both the real-object and image conditions was controlled using PLATO liquid crystal occlusion glasses (Translucent Technologies, Toronto, Ontario, Canada), which alternate between opaque (closed) and transparent (open) states. A chin rest was used to control viewing distance and to stabilize the head between all trials.

**Procedure.** During each trial, participants made a two-alternative forced-choice button-press response relating to the orientation (left vs. right) of the central target object (a spoon). The target was flanked above and below by two identical distractors whose handle orientation was the same as (i.e., congruent) or opposite from (i.e., incongruent) the target's, thereby eliciting a competing response. We did not include neutral flankers (stimuli that never appeared as targets and were therefore not associated with a motor response) in our assessment of distractor processing. Flanker studies have generally found no consistent differences between RTs on congruent and neutral flanker trials (Lavie & de Fockert, 2003), and it is customary to compare interference from incongruent flankers with interference from either congruent or neutral flankers (Avital-Cohen & Tsal, 2016). We used congruent flankers to ensure that our manipulation of incongruency did not vary the number of response alternatives (Lavie & de Fockert, 2003). The stimuli were positioned so that the center of the display was aligned at eye level when viewed from straight ahead in the chin mount. Each trial started with a 15-s waiting period, during which the PLATO glasses were closed (Fig. 1c). White noise was played during the first 14 s of the initial waiting period on all trials to mask any sounds generated during mounting of the stimuli. Next, the PLATO glasses opened to reveal the stimuli and remained transparent until the participant's response. Responses were entered on a standard computer keyboard by pressing the "A" or "L" key with the index finger of the left or right hand, respectively. An auditory tone provided feedback on incorrect trials.

Participants completed two blocks of trials for each display format. The order of blocks was counterbalanced within and between participants. The ordering of the target/flanker configurations was randomized within each block. Stimuli in each configuration were presented 10 times; with four unique configurations and two display formats; this yielded a total of 80 trials. Participants were instructed to indicate as quickly and accurately as possible whether the handle of the central target was oriented leftward or rightward. Participants were told that the flankers were irrelevant to their task, that they held no predictive information about the orientation of the target, and that they should be ignored. The experimental trials took approximately 45 min to complete, and the entire session took approximately 1 hr, including filling out consent forms and listening to task instructions.

**Results**

**Response times.** Trials in which RTs were more than 2 SD from the mean were removed from each display format and incongruency condition, and were not considered further for the analyses. Only trials with correct responses were entered into the RT analysis. The mean RT data were analyzed using a repeated measures analysis of variance
Fig. 2. Results of Experiment 1: (a) mean response time (RT) for each combination of display format (2-D image vs. real object) and target-flanker incongruency condition and (b) mean flanker interference index for stimuli in each display format. Asterisks above the data bars denote interference effects significantly greater than zero \((p < .05)\), and the asterisk above the bracket indicates that the difference between display formats was significant \((p < .05)\). Error bars represent 95% confidence intervals.

(ANOVA) with the factors of display format (real objects vs. images) and flanker incongruency (congruent vs. incongruent). As expected on the basis of previous flanker-paradigm studies, there was a significant main effect of incongruency, \(F(1, 39) = 775.85, p < .001, \eta^2_p = .95\); participants responded faster when the flankers were congruent \((M = 442\, ms, \text{SE} = 3)\) than when they were incongruent \((M = 474\, ms, \text{SE} = 4)\) with the target. Critically, however, we also observed a significant main effect of display format, \(F(1, 39) = 69.75, p < .001, \eta^2_p = .64\); participants were faster to respond to the 2-D image targets \((M = 446\, ms, \text{SE} = 4)\) than the real-object targets \((M = 471\, ms, \text{SE} = 4)\). Longer RTs were observed for real objects than for images in both the congruent trials (real objects: \(M = 452\, ms, \text{SE} = 4\); images: \(M = 431\, ms, \text{SE} = 4\), \(t(39) = 15.93, p < .001, d = 1.03\), and incongruent trials (real objects: \(M = 489\, ms, \text{SE} = 4\); images: \(M = 460\, ms, \text{SE} = 4\), \(t(39) = 19.38, p < .001, d = 3.47\). Moreover, we found a significant two-way interaction between display format and incongruency, \(F(1, 39) = 10.39, p = .005, \eta^2_p = .21\). Figure 2a displays mean RTs, separately for each combination of display format and incongruency condition.

To break down this interaction, we quantified flanker effects using an interference index, which estimated the RT difference between the incongruent and the congruent conditions and permitted a direct comparison of the relative strength of interference effects across display formats. The index was calculated using the following formula:

\[
\text{interference index} = \frac{(\text{RT}_I - \text{RT}_C)}{(\text{RT}_I + \text{RT}_C)} \times 100,
\]

where RT\(_I\) is the mean RT obtained on incongruent-flanker trials and RT\(_C\) is the mean RT obtained on congruent-flanker trials. Positive index values indicate longer RTs to incongruent than to congruent displays, negative values indicate the opposite, and values around zero indicate an absence of flanker interference (Fig. 2b). One-sample \(t\) tests against zero confirmed that flanker interference effects were significantly greater than zero for both the 2-D image displays \((M = 3.21, \text{SE} = 0.21), t(39) = 15.32, p < .001, d = 2.4,\) and the real-object displays \((M = 3.92, \text{SE} = 0.19), t(39) = 20.24, p < .001, d = 2.78.\) Critically, a paired-samples
Table 1. Mean Percentage of Errors (and Standard Errors) for Each Condition in Experiment 1

<table>
<thead>
<tr>
<th>Target-flanker incongruency</th>
<th>2-D image</th>
<th>Real object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>Congruent</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Incongruent</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

*t test revealed that the interference index for the real-object displays was significantly greater than that for the 2-D image displays, *t*(39) = 2.11, *p* = .041, *d* = 0.42.

Error rates. For all conditions, we used the filtered data (with RTs > 2 SD from the mean removed) in the analysis of error rates, although the results were the same when all data were analyzed. In all conditions, accuracy was very high, and none of our participants produced error rates more than 2 SD from the mean. Table 1 presents the mean percentage of errors in each condition in Experiment 1. A repeated measures ANOVA with the factors of display format and incongruency was performed on the mean error rates and revealed no significant main effects or interactions (all *p*s > .05).

Experiment 2

Method

Participants. Thirty-eight right-handed undergraduate students (26 females; mean age = 22.0 years, SD = 4.96), who did not participate in Experiment 1, completed Experiment 2 for course credit. All reported normal or corrected-to-normal vision. Participants provided informed consent, and the protocols used were approved by The University of Nevada, Reno IRB. Sample size was determined as in Experiment 1.

Stimuli, apparatus, and procedure. The stimuli, apparatus, and procedure for Experiment 2 were identical to those of Experiment 1, with the following exceptions. The image displays consisted of 3-D stereo images of the spoons in each of the four target/flanker display configurations. We created stereo images by photographing each of the real-object displays with a forward-facing camera, positioned 60 cm from the screen and 3.2 cm to the left and right of midline, respectively. Participants viewed the stimuli through glasses in both display formats. The real objects were viewed using PLATO glasses, as described in Experiment 1. In the stereo trials, the stimuli were viewed binocularly through active shutter glasses (3D Vision 2, NVIDIA) and displayed on an LCD monitor (120 Hz; Model VG278HE, ASUS, Beitou District, Taipei, Taiwan) with a screen resolution of 1,920 x 1,080 pixels.

Results

Response times. A repeated measures ANOVA with the factors of display format and incongruency showed a significant main effect of incongruency, *F*(1, 37) = 115.64, *p* < .001, η² = .76; participants responded faster when flanking objects were congruent (M = 431 ms, SE = 8) than when they were incongruent (M = 458 ms, SE = 8) with target orientation. There was a significant main effect of display format, *F*(1, 37) = 9.31, *p* = .004, η² = .20; participants were faster to respond to the stereo targets (M = 435 ms, SE = 9) than to the real-object targets (M = 455 ms, SE = 9). Significantly longer RTs for real objects (M = 475 ms, SE = 9) than for images (M = 444 ms, SE = 9) were observed in the incongruent displays, *t*(37) = 3.74, *p* < .001, *d* = 0.61, and there was a similar trend for congruent displays (real objects: *M* = 438 ms, SE = 9; images: *M* = 425 ms, SE = 9), *t*(37) = 3.62, *p* = .120, *d* = 0.29. Critically, we found a significant two-way interaction between display format and incongruency, *F*(1, 37) = 7.62, *p* = .009, η² = .17. Mean RTs are shown in Figure 3a, separately for each condition. We examined the magnitude of the flanker effect by calculating an interference index for stimuli in each display format, as described in Experiment 1 (Fig. 3b). Although flanking objects interfered with RTs in both the stereo displays (M = 2.22, SE = 0.32), *t*(37) = 6.83, *p* < .001, *d* = 1.10, and the real-object displays (M = 3.85, SE = 0.50), *t*(37) = 7.68, *p* < .001, *d* = 1.27, the interference index was again significantly greater for the real objects than for the stereo images, *t*(37) = 2.65, *p* = .012, *d* = 0.51.

Error rates. Table 2 shows the mean percentage of errors for each condition in Experiment 2. A repeated measures ANOVA for mean error rates with the factors of display format and incongruency revealed no significant main effects or interactions (all *p*s > .05).

The results of Experiments 1 and 2 show that real grasping objects compete more strongly for attention than do 2-D and 3-D computerized images of grasping objects: RTs were slower overall and flanker interference effects were greater for real spoons than for both 2-D and 3-D images of the spoons. Experiment 2 confirmed our prediction that the influence of real objects on attention is not attributable to stereoscopic disparity cues (which are present in real objects but not 2-D images), because the same pattern of greater interference effects for real objects was observed in comparison with 3-D stereoscopic images. The key question is whether real objects interfere with attention because they afford grasping (Gallivan et al., 2009; Iriki et al., 1996; Mountcastle et al., 1975) or whether they require more detailed and richer visual processing (and incongruence in the visual domain results in greater competition). Under conditions in which real objects no longer afford...
grasping, they should not stimulate networks involved in on-line action planning and execution, and so their effects on behavior should be similar to images (i.e., overall RTs, and flanker interference effects, should be comparable for real objects and images). Conversely, if real objects and images engage similar underlying cognitive processes but real objects simply amplify effects that would otherwise be observed for images, then manipulations of graspability should have similar effects on the processing of real objects and images (i.e., overall RTs and flanker interference effects should be greater for real objects than for images).

Experiment 3

Method

Participants. Forty-eight right-handed undergraduate students (36 females; mean age = 21.2 years, SD = 7.66), who had not participated in Experiments 1 or 2, completed the experiment in exchange for course credit. All participants reported having normal or corrected-to-normal vision. Each participant provided informed consent, and the protocols used were approved by The University of Nevada, Reno IRB. Sample size was determined as in the previous experiments.

Stimuli, apparatus, and procedure. The stimuli, apparatus, and procedure for Experiment 3 were identical to those of Experiment 1, except that viewing distance was increased from 60 cm to 80 cm, so that the stimuli were no longer within grasping distance of the seated participant. Manipulations of graspability change the relationship between the body and the target without modifying the body or the target.
Fig. 4. Results of Experiment 3: (a) mean response time (RT) for each combination of display format (2-D images vs. real objects) and target-flanker incongruency condition and (b) mean flanker interference index for stimuli in each display format. Asterisks above the data bars denote interference effects significantly greater than zero ($p < .05$). Error bars represent 95% confidence intervals.

**Results**

**Response times.** A repeated measures ANOVA with the factors of display format and incongruency showed a significant main effect of incongruency, $F(1, 47) = 145.57$, $p < .001$, $\eta_p^2 = .76$; where participants responded faster when flankers were congruent than when they were incongruent with target orientation. Unlike in Experiments 1 and 2, however, there was no main effect of display format, $F(1, 47) = 1.21$, $p > .250$, $\eta_p^2 = .03$, and no interaction between display format and incongruency (congruent real object: $M = 457$ ms, $SE = 9$; congruent image: $M = 452$ ms, $SE = 10$; incongruent real object: $M = 492$ ms, $SE = 9$; incongruent image: $M = 483$ ms, $SE = 10$), $F(1, 47) = 0.860$, $p > .250$, $\eta_p^2 = .02$. Mean RTs in Experiment 3 are shown in Figure 4a separately for each condition, and Figure 4b displays the interference indices for stimuli in each display format. Although flanker interference effects were significantly greater than zero for both the 2-D images ($M = 3.26$, $SE = 0.43$), $t(47) = 6.73$, $p < .001$, $d = 1.06$; and the real objects ($M = 3.73$, $SE = 0.36$), $t(47) = 9.01$, $p < .001$, $d = 1.44$; paired-samples $t$ tests revealed that there was no significant difference in interference effects for real objects positioned out of reach versus 2-D images, $t(47) = 0.87$, $p > .250$, $d = 0.13$.

**Error rates.** Table 3 shows the mean percentage of errors for each condition in Experiment 3. A repeated measures ANOVA with the factors of display format and incongruency was performed on the mean error rates and revealed no significant main effects or interactions (all $ps > .05$).

**Experiment 4**

**Method**

**Participants.** Forty-eight right-handed undergraduate students (29 females; mean age = 24.3 years, $SD = 7.12$), who had not participated in any of the previous experiments, completed the study in exchange for course credit. All participants reported having normal or corrected-to-normal vision. Participants provided informed consent, and the protocols used were approved by The University of Nevada, Reno IRB. Sample size was determined as in the previous experiments.

**Stimuli, apparatus, and procedure.** The stimuli, apparatus, and procedure for Experiment 4 were the same as those of Experiment 1, except that the stimuli were presented behind a transparent barrier so that they no longer...
Table 3. Mean Percentage of Errors (and Standard Errors) for Each Condition in Experiment 3

<table>
<thead>
<tr>
<th>Target-flanker incongruency</th>
<th>2-D image</th>
<th>Real object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>Congruent</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Incongruent</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

afforded in-the-moment grasping (Morgado, Gentaz, Guinet, Osuira, & Palluel-Germain, 2013). The barrier was a large acrylic sheet (36 × 48 × ¼ in.) positioned vertically between the participant and the stimuli, subtended at the left and right edges by wooden beams. The barrier was centered 28 cm from the edge of the table and 32 cm from the stimulus arrays (and it is unlikely that participants would focus at the distance of the barrier rather than at the distance of the stimuli). Importantly, viewing distance (and retinal size) was identical to Experiment 1.

Results

Response times. A repeated measures ANOVA with the factors of display format and incongruency revealed a significant main effect of incongruency, \( F(1, 47) = 187.86, p < .001, \eta_p^2 = .80 \). There was no main effect of display format, \( F(1, 47) = 0.08, p > .250, \eta_p^2 = .01 \), and no interaction between display format and incongruency (congruent real object: \( M = 466 \text{ ms, } SE = 12 \); congruent image: \( M = 463 \text{ ms, } SE = 13 \); incongruent real object: \( M = 503 \text{ ms, } SE = 12 \); incongruent image: \( M = 501 \text{ ms, } SE = 12 \), \( F(1, 47) = 0.03, p > .250, \eta_p^2 = .01 \). Figure 5a shows the mean RTs separately for each condition in Experiment 4. Figure 5b shows the interference indices for stimuli in each display format. Although flanker interference effects were significantly greater than zero for both the 2-D image (\( M = 4.10, SE = 0.43 \), \( t(47) = 9.08, p < .001, d = 1.43 \), and real-object displays (\( M = 3.83, SE = 0.42 \), \( t(47) = 8.00, p < .001, d = 1.27 \), there was no difference in the magnitude of the interference indices between the two display formats when the stimuli were positioned behind a transparent barrier, \( t(47) = 0.63, p > .250 \).

Fig. 5. Results of Experiment 4: (a) mean response time (RT) for each combination of display format (2-D images vs. real objects) and target-flanker incongruency condition and (b) mean flanker interference index for stimuli in each display format. Asterisks above the data bars denote interference effects significantly greater than zero (\( p < .05 \)). Error bars represent 95% confidence intervals.
Table 4. Mean Percentage of Errors (and Standard Errors) for Each Condition in Experiment 4

<table>
<thead>
<tr>
<th>Target-flanker incongruency</th>
<th>2-D image</th>
<th>Real object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>Congruent</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Incongruent</td>
<td>0.6</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Error rates. Table 4 displays the mean percentage of errors for each condition of Experiment 4. A repeated measures ANOVA with the factors of display format and incongruency was performed on the mean error rates and revealed no significant main effects or interactions (all ps > .05).

Analysis of the Relationship Between RT and Flanker Interference

The results from Experiments 1 through 4 demonstrate that the relative effect of display format on RTs and on flanker interference changes depending on whether or not the stimuli afford in-the-moment manual interaction. Given that action-related effects may decline at different rates throughout an experiment for stimuli in different display formats (Squires et al., 2016), we were careful to keep the number of trials and the task duration constant across all the experiments. Different participants completed each experiment, but the viewing conditions were always equivalent for stimuli in the different display formats. Although RTs tended to be longer overall in Experiments 3 (mean RT = 470 ms) and 4 (mean RT = 508 ms) than in Experiments 1 (mean RT = 458 ms) and 2 (mean RT = 445 ms), perhaps because of reduced retinal size, visual crowding, or minor reflections on the barrier, Figures 2 through 5 show there was no consistent relationship between RT and flanker interference effects across experiments. Moreover, Pearson’s product-moment correlations confirmed that participants’ RTs did not vary systematically with the magnitude of flanker interference indices across experiments (Table 5). Interestingly, despite the equivalent viewing conditions within each study, the increase in RTs in Experiments 3 and 4 tended to be greater for 2-D images than for the real objects, consistent with the idea that there was a relative facilitatory effect on RTs for the real objects (vs. 2-D images) when they were no longer graspable. It is also the case, however, that different groups of undergraduate participants were tested in each experiment, and so differences in overall RTs may simply reflect intersubject variability.4

Discussion

We used a flanker task to examine whether real objects exert a stronger competitive influence on attention and manual responses than images of objects. We hypothesized that compared with images, real objects that afford genuine action should trigger grasp-related visuo-motor plans that conflict more strongly with button-press responses, and to-be-ignored graspable flankers should compete more powerfully for selection with a central target. Compared with RTs for matched 2-D computerized images of objects, RTs for real graspable objects were slower overall, and real-object flankers interfered more strongly with responses to the central target. The same pattern of effects was observed for real objects compared with 3-D stereoscopic images, indicating that the effect of real objects on attention is not attributable to depth cues that provide richer information about the shape, size, and distance of the object from the participant. Critically, however, RTs and flanker interference effects were equivalent for real objects and images when the stimuli were placed out of reach of the participant, as well as when the stimuli were presented within reach but behind a large transparent barrier that prevented manual interaction with the stimuli. Together, the results demonstrate that real objects exert

Table 5. Pearson’s Product-Moment Correlations Between Response Times and Flanker Interference Effects, Separately for Each Display Format and Experiment

<table>
<thead>
<tr>
<th>Experiment</th>
<th>N</th>
<th>Pearson’s r</th>
<th>p</th>
<th>Pearson’s r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Image displays</td>
<td></td>
<td>Real-object displays</td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td>40</td>
<td>−.281</td>
<td>.079</td>
<td>.458</td>
<td>.003</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>38</td>
<td>.125</td>
<td>.455</td>
<td>−.018</td>
<td>.916</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>48</td>
<td>−.026</td>
<td>.859</td>
<td>−.118</td>
<td>.425</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>48</td>
<td>−.340</td>
<td>.018</td>
<td>−.042</td>
<td>.777</td>
</tr>
<tr>
<td>All experiments</td>
<td>174</td>
<td>−.091</td>
<td>.232</td>
<td>−.061</td>
<td>.427</td>
</tr>
</tbody>
</table>
a more powerful influence on attention and manual responses than representations of objects and that this effect is primarily due to the affordances that real objects provide the perceiver for physical interaction.

Our stimuli, procedure, and design preclude a number of alternative explanations. Because the stimuli were presented in the same display format on each trial, differences in interference effects cannot be due to the apparent depth or relative conspicuity of the target versus the flankers. Our stimuli were matched closely for size, distance, viewpoint, and illumination, and viewing time was computer controlled on all trials. The influence of graspable objects on attention cannot be attributed to differences in scaling that could arise as a result of longer RTs on real-object trials, because absolute interference measures on congruent versus incongruent trials were scaled by overall RTs, separately for each display format. Although previous studies have found that RTs to name objects are faster for photographs than for more impoverished line drawings that provide fewer shape and monocular depth cues (Humphrey et al., 1994; Salmon, Matheson, & McMullen, 2014), and that patients with visual agnosia are faster to name real objects than 2-D photos of the same items (Chainay & Humphreys, 2001; Humphrey et al., 1994), our results suggest that when the task requires a manual response, the processes engaged by richer graspable objects lead to interference rather than facilitation. Explanations based on visual richness do not, however, explain why responses to real objects were equivalent to responses to images in our experiments when the real objects no longer afforded grasping. Although conflicting depth cues from ocular vergence and lens accommodation to images (but not real objects) become less apparent with increasing distance (as in Experiment 3), cue conflicts are unlikely to explain our finding of equal RTs and interference effects for real objects and 2-D images in the context of a barrier (Experiment 4) or why a decrease in cue conflicts should predict longer overall RTs for real objects in Experiments 1 and 2. Finally, we confirmed that the magnitude of flanker interference effects was not explained by differences in overall RTs.

Our results provide strong support for the affordance-competition hypothesis (Cisek, 2007) and cognitive models that emphasize the importance of action constraints on attention (Humphreys et al., 2013). In particular, the findings underscore that the goal of attention is to ensure that one acts on the right object at any given moment and that capacity limits in attention reflect physical constraints imposed by the number of actions that can be performed coherently on an object at a given time (Humphreys et al., 2013). Whether real objects trigger a greater number of competing action plans than images, or whether these plans and their associated feedback gains are stronger, more highly elaborated (Gallivan, Logan, Wolpert, & Flanagan, 2016), or temporally distinct, awaits further investigation. Our data from adults align with studies showing that children habituate differently to (Gerhard, Culham, & Schwarzer, 2016) and maintain fixation longer (Mustafar, De Luna, & Rainer, 2015) on real objects compared with matched 2-D images. Further, our findings add to an emerging literature showing that cognitive processes, such as object recognition (Chainay & Humphreys, 2001; Humphrey et al., 1994), memory (Snow et al., 2014), and decision making (Mischel & Moore, 1973; Romero et al., in press), as well as neural responses (Snow et al., 2014), differ between real objects and 2-D images. Critically, our results demonstrate that real objects are processed differently than both 2-D and 3-D images because they afford physical action.

Outstanding questions for future research will be to isolate the underlying cognitive and neural mechanisms that subserve action-related effects in the context of real objects and images (which could reflect the differential recruitment of ventral perceptual and dorsal action-related cortical networks; Freud, Plaut, & Behrmann, 2016), determine whether outcome measures reflect quantitative or qualitative differences in the underlying mechanisms (Camerer & Mobbs, 2017), and identify whether effects of display format are modulated with increased stimulus distance, different object types (i.e., tools vs. nontools), or other manipulations of reachability (such as using a tool to extend reachable space). Whether similar signatures of "realness" can be elicited using immersive virtual reality (Wamain, Gabrielli, & Coello, 2016) or augmented-reality displays—particularly those that allow goal-directed actions with representations—is a question of empirical and philosophical importance. Because normal development in humans (Kretch & Adolph, 2015) and animals (Held & Hein, 1963) relies on having active manual control of the environment in response to visual inputs, studying vision with naturalistic graspable objects will yield important insights into the relationship between active observers and their physical environment.
Acknowledgments
We thank Marlene Behrmann and Michael Groggme for their comments on the manuscript.

Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding
This research was supported by grants to J. C. Snow from the National Eye Institute of the National Institutes of Health (Grant R01EY026701). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Notes
1. In all subsequent experiments, trials in which RTs were more than 2 SD from the mean were also removed from each display format and incongruency condition, and were not considered further for the analyses.
2. In the subsequent experiments, we again used filtered data (with RTs > 2 SD from the mean removed) in the analysis of error rates (results were the same when all data were included).
3. As in Experiment 1, accuracy was high in all conditions, and all error rates were less than 2 SD from the mean.
4. The image stimuli used for Experiments 1 through 4 can be downloaded at http://wolfweb.unr.edu/~snow/Materials/Stimuli.zip.
5. The RT and error data for Experiments 1 through 4 can be downloaded at http://wolfweb.unr.edu/~snow/Materials/Data.xlxx.

References


Romero, C. A., Compton, M. T., Yang, Y., & Snow, J. C. (in press). The real deal: Willingness-to-pay and satiety expectations are greater for real foods versus their images. *Cortex*


