When action conditions perception: Evidence of cross-modal cue recruitment

Guy M. Wallis

Centre for Sensorimotor Performance,
University of Queensland, Brisbane, Australia

Benjamin T. Backus

Graduate Center for Vision Research,
State University of New York College of Optometry,
New York, NY, USA

Goal-directed movement, such as reaching to touch an object, relies heavily on vision. Vision guides our motor system not only during initial targeting but also during online movement correction and error-driven learning. But it is not all one-way traffic. This paper reports a situation in which this perceptual-motor interaction runs in reverse, when action affects concurrent perceptual experience. More to the point, the paper reveals that visual perception is subject to change through learned (even arbitrary) visuomotor associations. By considering a situation in which the perceptual decision is dichotomous, this paper reveals that the brain readily harnesses motor behavior to constrain the formation of a visual percept.

Introduction

The impression of our visual surroundings that our minds form is the product of a multiplicity of cues arising from multiple sources, including those not directly associated with the visual system (Ernst & Bülthoff, 2004). The information these cues provide to an observer derive from their dependency on a particular aspect of the world (Backus, 2009), and in order for a cue to be informative, it must give rise to a reliable sensory estimate (Ernst, 2007; Ernst & Bülthoff, 2004; Jacobs, 2002; Johnston, Cumming, & Landy, 1994; Körding & Wolpert, 2004; Landy, Maloney, Johnston, & Young, 1995). In cues theory (von Helmholtz, 1910/2005), these cues are then combined to yield the most accurate (probable) estimate of the state of our surroundings. For example, accurate depth perception involves a combination of cues such as binocular disparity, occlusion and motion parallax, to name just three (Landy et al., 1995). More formally, the brain has been shown to integrate cues in a manner commensurate with estimates of their reliability (Backus & Banks, 1999; Ernst & Banks, 2002; Ernst & Bülthoff, 2004).

The question this paper seeks to address is just how flexible this cue combination process is. From the time of Pavlov we have known that the mammalian brain is adept at forming associations between arbitrary stimuli, and on some level it would seem that any internal cortical process should be open to the possibility of not only modifying its reliance on known cues, but to the possibility of discovering novel cues as well. In the first half of the 20th century, there were a few notable attempts to demonstrate associative learning in perception (Brunswik & Kamiya, 1953; Howells, 1944; Wallach, O'Connell, & Neisser, 1953), but they all had to contend with two major problems: first, the difficulty of showing that effects resulted from true changes in appearance, and second, the fact that effects induced during a behavioral experiment are likely to be small, relative to those established over a lifetime of perceptual learning. In the end, Gibson and Gibson (1955) argued that perceptual learning simply does not occur in the types of experiments the researchers had used, and most of the field gradually lost interest. Modern definitions of perceptual learning often explicitly exclude the learning of new cues by means of association. Indeed, a major textbook on perceptual learning states that "contrary to associative learning, perceptual learning does not bind together two processes that were separated" (Fahle & Poggio, 2002, ix). Although unfashionable today, the empiricist philosophers clearly believed that associative learning was crucial to perception (Berkeley, 1725; Locke, 1706). And later, many of the founders of cognitive science subscribed to the idea as well (Hebb, 1949; James, 1890; von Helmholtz, 1878). It is perhaps ironic that these authors, currently held in such high regard, should

have all believed in a modern-day heresy. As it turns out, over the last ten or fifteen years a small but significant number of reports have provided evidence that these early vision scientists may, in fact, have been correct (Adams, Graf, & Ernst, 2004; Ernst, Banks, & Bülthoff, 2000; Haijiang, Saunders, Stone, & Backus, 2006; Jacobs & Fine, 1999; Sinha & Poggio, 1996; Wallis & Bülthoff, 2001).

Of the handful of recent papers on the topic, it was Backus and colleagues who specifically set out to test the hypothesis, and hence it is only their papers that directly address this work's implications for theories of associative learning and perception. On the basis of their studies, the authors coined the term "cue recruitment" to refer to situations in which the visual system adopts a signal as evidence for the veracity of a particular interpretation of a visual stimulus (Haijiang et al., 2006). The authors based their claim on experiments in which a trusted visual cue (e.g., binocular disparity, occlusion, and perspective) was added to an otherwise ambiguous figure, such as a Necker Cube, so as to force the associated percept into one of its two stable states. They then added an irrelevant signal to the figure, such as location on a display screen, but in a manner that was contingent upon the trusted cues. For example, the figure was always seen to rotate clockwise when high on the screen, and always anticlockwise when low on the screen. On test trials, appearance of the figure (inferred from the direction of perceived rotation) was seen to be affected by the presence of the newly associated signal, even in the absence of the trusted cues.

Subsequent to the 2006 study, a number of papers have been published reporting evidence of cue recruitment (Backus & Haijiang, 2007; Di Luca, Ernst, & Backus, 2010; Haijiang et al., 2006; Harrison & Backus, 2012). However, the effects were often weak, and attempts to recruit some cues failed. For example, attempts to associate an auditory cue with appearance, in the style of Pavlov, failed (Haijiang et al., 2006), calling into question the generality of the findings and their link to classical conditioning. Orhan, Michel, and Jacobs (2010) suggest that there must be an underlying or implicit knowledge of which cues can be integrated and correlated. Jain, Fuller, and Backus (2010) wondered if one source of knowledge might be the fact that the sounds tested are extrinsic to the stimulus, whereas location and direction of motion, for which association worked, are intrinsic. To test this, the authors used sounds (a camera shutter and ratchet), which are associated with rotation in the real world, to provide a more intrinsic facet to the sounds. However, recruitment once again failed. More recently, the same authors did, at last, report evidence for recruitment of an extrinsic visual cue by removing the trusted cue in the middle of a training trial. However, they once again failed to find

evidence of the recruitment working for sounds (Jain, Fuller, & Backus, 2014). Other groups have reported that sounds can be recruited as cues to visual motion (Teramoto, Hidaka, and Sugita, 2010) and that haptic feedback can change the relative reliabilities of visual cues (Ho, Serwe, Trommershäuser, Maloney, & Landy, 2009). However, the mechanism behind cue recruitment is more complex than establishing a simple contingent relationship (Di Luca et al., 2010).

One modality that has yet to be tested is that of motor behavior. It has long been known that perception and action are deeply linked. Vision provides the initial information for both action planning and action initiation, as part of the process of producing movements with the appropriate onset, duration, and force (Bekkering & Neggers, 2002). But once an action is carried out, vision also supplies the means by which the status of the action is verified (Bekkering & Neggers, 2002; Hommel, 2004, 2007; Schubotz & von Cramon, 2002). Though clearly necessary for executing fine motor corrections during grasping, for example, sensory feedback can also play a crucial role in instigating actions during multiphasic movements. In the absence of feedback, counterintuitive failures in behavior can result, as has been revealed by studies of vehicle steering control (Cloete & Wallis, 2009; Wallis, Chatziastros, & Bülthoff, 2002; Wallis, Chatziastros, Tresilian, & Tomasevic, 2007).

Of relevance to the study reported here, there is also growing evidence that visuomotor interactions run both ways, such that motor planning can influence perception (Fagioli, Hommel, & Schubotz, 2007; Wykowska, Hommel, & Schubö, 2011; Wykowska, Schubö, & Hommel, 2009). Fagioli et al. (2007), for example, demonstrated that action planning not only primes action-related dimensions; it also primes feature-related dimensions that are relevant to planning the action. In their study, participants were asked to prepare a grasping or pointing movement which was triggered by spotting deviant stimuli that appeared in a sequence of test images. Participants were faster to spot size deviants when preparing a grasping movement and faster at spotting location deviants during a pointing task. The authors argue that the results provide evidence that perceptual and motor events share an underlying mechanism or representation (Hommel, Musseler, Aschersleben, & Prinz, 2001; Schubo, Prinz, & Achersleben, 2004). Beyond motor planning, Maruya, Yang, and Blake (2007) reported evidence for a direct link between voluntary action and perception. In their task, observers were presented with binocularly rivalling images of a grating and random-dot sphere. The authors found that in the case that the motion of the dot sphere was yoked to mouse movements, the sphere was more likely to dominate perception than in
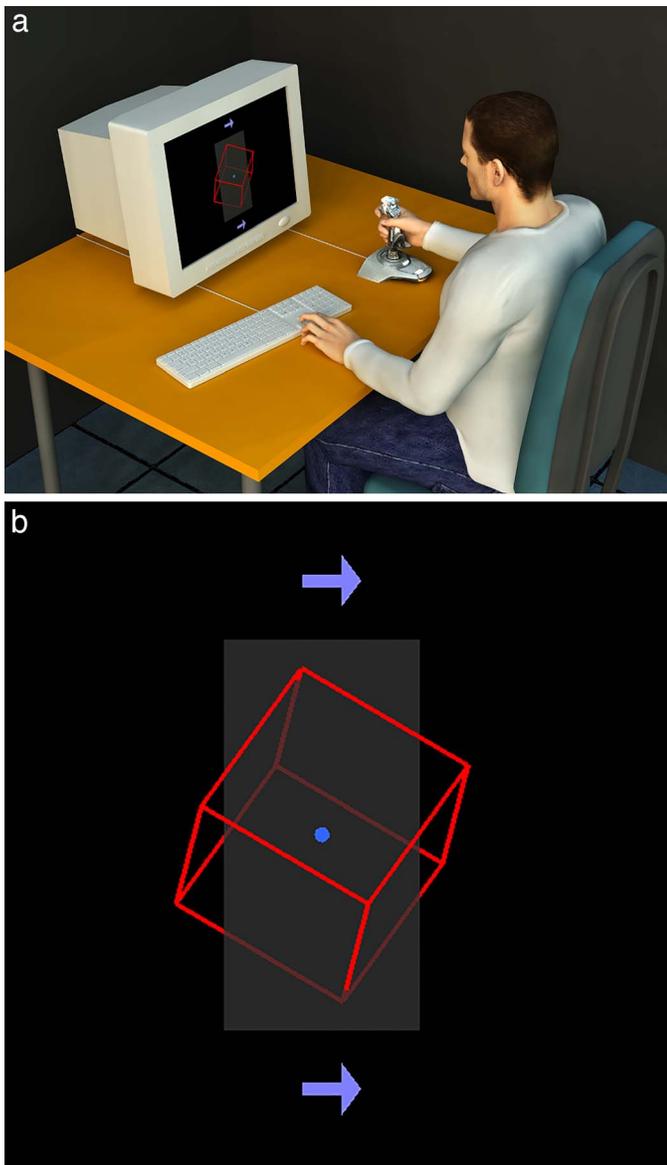
Figure 1. (a) The experimental apparatus, including the display screen, joystick, and response keyboard. (b) View of the cube stimulus seen during the training phase. The direction arrows, seen above and below the stimulus, indicated the direction in which the participant was required to move the joystick. The arrows also flashed to indicate the required duration of each joystick movement.

the case that the arm and sphere motion were not yoked.

The present study seeks to determine whether the impact of movement on perception described by Maruya et al. (2007), is amenable to learning. This learning, if it occurs, would confirm the flexibility of such associations and would imply that motor action can be recruited as a cue that influences perceptual appearance, in the cue recruitment framework of Backus (2011), offering the first evidence for recruitment crossing modalities.

To examine whether voluntary movement can be recruited in this way, we studied the impact of arm movements on perceptual decision-making during a pair of experiments. In the first experiment, a perceptually bistable stimulus rotated about its vertical axis, controlled by a participant's movement of a joystick. For one group of participants, arm movement was congruent with the trained visual percept, mimicking the conditions studied by Maruya et al. (2007). However, a second group experienced the reverse contingency, resulting in a less intuitive link between perceived direction of stimulus rotation and arm movement. In the second experiment, the perceptually bistable object rotated orthogonally to the direction of arm motion, completely decoupling the perceived direction of object motion from that of the arm. In all cases, arm movement strongly influenced perceptual choices in our participants. A control condition in which participants observed a moving stimulus during training without directly controlling it, consistently failed to produce recruitment. We discuss the results in the context of visuomotor theories of perception and cue recruitment.

Materials and methods

Participants

A total of sixty undergraduate students with corrected to normal vision took part in the first study. They were drawn from the student population at the University of Queensland. Participants received AU\$15 for their time. The experiment was reviewed and approved by the University's Behavioural and Social Sciences Ethical Review Committee.

Hardware and software

Stimuli were rendered using custom OpenGL and SGI Performer software running on an SGI Octane computer and presented to participants on a 24-inch Sony Trinitron monitor with a simulation and screen refresh rate of 72Hz. The objects were represented as three-dimensional objects in OpenGL but were rendered using orthographic projection to remove perspective cues. Hand movements were recorded using a Saitek Force-feedback Cyborg Evolution joystick interfaced to a PC computer which relayed the deflection angle to the display computer via a dedicated Ethernet link. The joystick was set up to produce a minor, but perceptible recentering force (see Figure 1a).
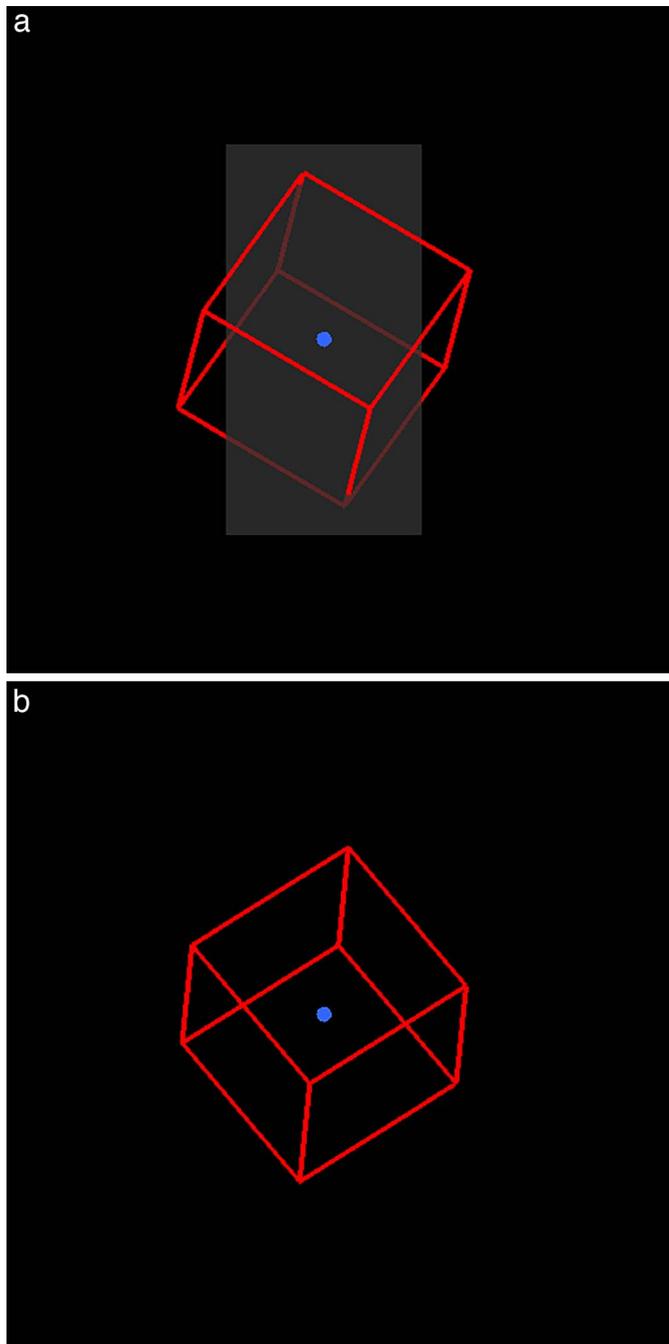
Figure 2. The Necker Cube stimulus. (a) Cube with occluder, used to produce a stable percept during training trials. (b) Ambiguous version of the cube used during test trials.
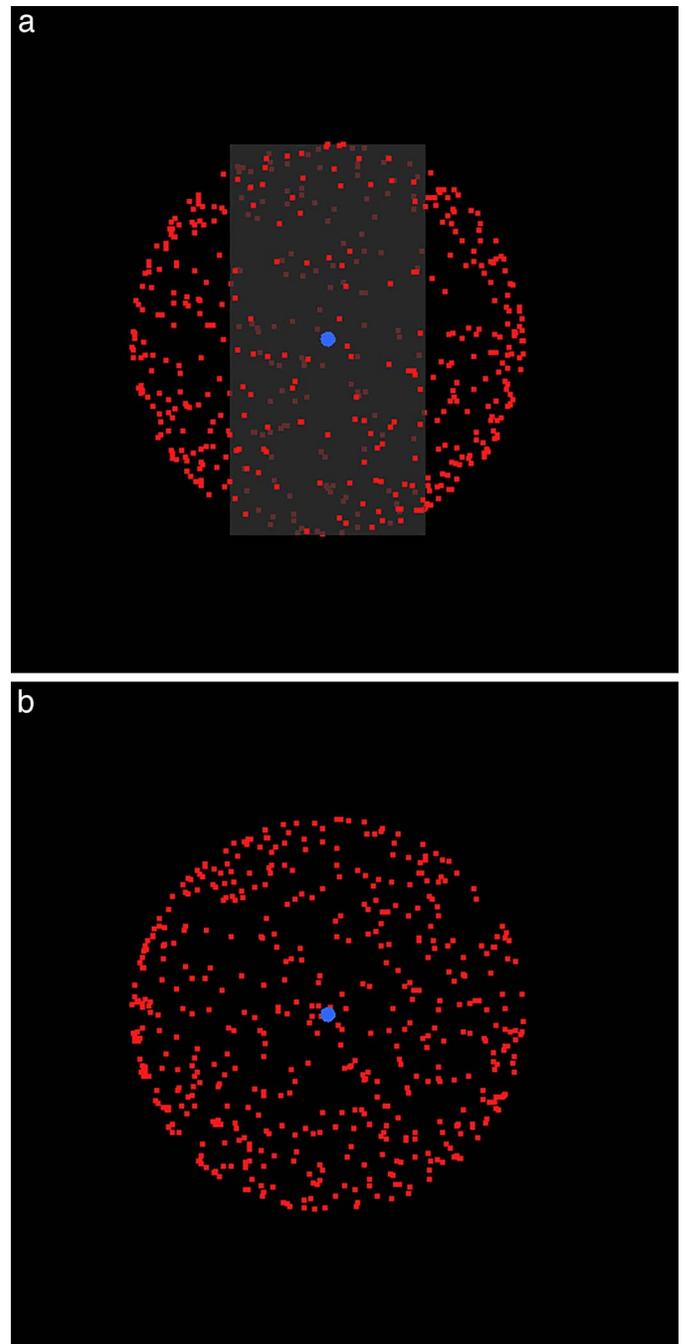

Figure 3. The dot-point sphere stimulus. (a) Sphere with occluder, used to produce a stable percept during training trials. (b) Ambiguous version of the sphere used during test trials.

Stimuli

Two figures were used in the experiments: A Necker Cube (see Figure 2), and a dot sphere (see Figure 3). The stimuli were displayed centrally on the monitor at a distance of approximately 60 cm. Cube edges were 4.0 cm in length, resulting in a stimulus which subtended a visual angle of approximately 7.0 degrees. The sphere stimulus had a diameter of 8.5 cm. It consisted of 500 dots located randomly, but equiprobably across the sphere's surface. Each dot was square and occupied an area of approximately 1.3 mm × 1.3 mm. Both stimuli were displayed in high contrast red on an otherwise blank, black background. A small, bright blue fixation dot of 3 mm diameter was presented at the center of the displayed stimulus. This fixation dot was visible at all times.

During training an occlusion cue was created using a 4.5 cm × 8.5 cm gray semiopaque surface which acted

as a salient cue to depth. The occluder extended from above the stimulus to just below it vertically and covered approximately two thirds of the stimulus horizontally. A perceived sense of depth was created via a brightness change as portions of the stimulus passed behind the semitransparent occluder.

All stimuli were presented at a random starting angle, in order to mitigate any interaction between starting orientation and preferred interpretation of the initially static stimulus.

Procedure

During a training trial, a pair of arrows appeared above and below a disambiguated version of the cube or sphere stimulus (i.e., with the occluder present; see Figure 1b). The arrows pointed either left or right, indicating the direction in which the participant should move the joystick. The arrows appeared once every 1500 ms (108 frames) and remained visible for 83 ms (six frames). Participants were instructed to move the joystick from the neutral/upright position out to the furthest extent of the joystick's movement range, and to do so in such a way that the moment of movement initiation coincided with the onset of a flash of the arrows, and concluded at the moment of the subsequent flash. This ensured that movement, while voluntary, was somewhat constrained in terms of its speed and amplitude. Movement of the cube/sphere was directly yoked to motion of the joystick, such that the degree of rotation from its start orientation was exactly proportional to the angle of the joystick. Full deflection of the joystick (corresponding to a 22° inclination in the frontoparallel plane) produced a 100° rotation of the stimulus about the vertical axis.

After completing both an outward and return (to upright) movement of the joystick, apparent rotation of the stimulus was reported using two keypresses on a standard QWERTY keyboard. Participants used their first keypress to indicate the perceived direction of rotation of the front of the stimulus when they moved the joystick from the central position outwards, and then used the second to report the perceived direction of object rotation when they returned the joystick to the starting position. Direction was indicated via selection of the left or right cursor keys. Trials terminated once the second keypress was completed.

Each participant completed two experimental conditions, once viewing the cube and once the sphere, with order counterbalanced across members of each training group. The direction of rotation requested (as indicated by the flashing direction arrows) was random but counterbalanced, resulting in three leftward and three rightward training trials per block.

There were two movement conditions used during training, as follows:

- In the "compatible" training condition, rightward motion of the joystick was associated with a rightward rotation of the stimulus (in the sense that the frontal surface of the stimulus was seen to move rightward), and leftward movement with a leftward rotation of the stimulus.
- In the "incompatible" training condition, the contingency was reversed: A rightward joystick movement was associated with a leftward stimulus rotation, and a leftward joystick movement with rightward rotation.

The term "compatible" is derived here from its meaning in the literature on man-machine interfaces, in which an arm movement that is in the same direction as a resulting machine response is deemed "compatible." Compatibility is significant in this literature because is known to produce faster and more effective learning (Burgess-Limerick, Krupenia, Wallis, Pratim-Bannerjee, & Steiner, 2010; Worringham & Beringer, 1998). The concept of stimulus-response compatibility is essentially the same as in the psychological literature, with both linking back to the work of Fitts and Seeger (1953).

In addition to the compatible/incompatible distinction, participants were also assigned to one of three modes of learning (Active, Passive, or No Training) as follows:

- In the "active" training condition, participants were able to actively control the stimulus' rotation through movement of the joystick. Arrows located above and below the stimulus instructed the participant to move the joystick in that direction trial by trial.
- In the "passive" training condition, the stimulus rotated by itself during training, with the arrows once again indicating the direction of rotation. The joystick was not used during training.
- In a final, "no training" condition, participants only completed test trials. This served as a baseline condition against which the effects of training could be assessed. Note that this meant that participants completed half the number of trials completed by participants in the training groups.

The sixty participants were randomly assigned to one of five between-participant experimental groups: "compatible and active," "compatible and passive," "incompatible and active," "incompatible and passive," or simply "no training." Before starting the experiment, participants were familiarized with the two stimuli to ensure that they were able to see both perceptual forms in the disambiguated state clearly. In practice all of our participants reported being able to see the two forms and responded correctly to a series of five test displays

shown for each stimulus. They were then given instructions about the nature of the task, including use of the arrow keys. Participants were also given an opportunity to practice moving the joystick in time to the flashing arrows (but with the test stimulus absent).

Test trials proceeded in exactly the same manner as the active training trials, except that the occluder was removed. Hence, irrespective of the type of training, during testing all participants were required to move the joystick in a manner consistent with the flashing arrows but in the absence of the trusted, disambiguating occlusion cue. During the experiment, training trials were interleaved with test trials in fifteen blocks, each consisting of six training trials followed by six testing trials. Once a participant had completed all fifteen blocks for one of the stimuli (cube or sphere), the test was repeated for the second (cube/sphere order was counterbalanced across each training group).

One limitation of the first experiment is that it reports a link between arm and object motion in the special case that the perceived direction of motion of the frontal face of the object aligns with the horizontal axis of the fronto-parallel plane. The same is true of the motion of the hand. This alignment in the axes of motion leads to the possibility that the links are specific to this configuration, calling the generality of the results into question. For this reason, in experiment 2, the exact same protocols were followed for a new group of 60 participants, except that the stimuli now rotated about the horizontal axis. Hence the issue of compatibility now becomes moot, since there is now no a priori reason for thinking that there is a preferred or more intuitive combination of arm and object motion.

The two movement conditions used during training in experiment 2:

- In the "α" training condition, rightward motion of the joystick was associated with a downward rotation of the stimulus (in the sense that the front surface of the stimulus was seen to move downward), and leftward movement with an upward rotation of the stimulus.
- In the "β" training condition, the contingency was reversed: A rightward joystick movement was associated with an upward stimulus rotation, and a leftward joystick movement with downward rotation.

In accordance with this new orientation of perceived object motion, the response keys were switched from left and right cursor keys to the up and down keys.

Results

Preliminary analysis focussed on performance in the training trials, to determine if the training regime used was effective. Participants in the two active training groups moved the joystick in the direction indicated by the arrows on just over 98% of trials. Participants in all four training groups also reported seeing the stimulus move in a direction consistent with the trusted occlusion cue on the vast majority of trials. The actual figures break down across the four active training groups as follows: experiment 1 compatible = 94.2%, incompatible = 85.6%; experiment 2: α = 93.7%, β = 96.4%. The lower figure for the incompatible group versus the compatible group is consistent with the assertion made earlier that naive viewers prefer to perceive the compatible contingency of arm and object motion. A two-tailed, independent-sample t test revealed that although large, the difference was not statistically significant, $t(22) = 1.03$, In experiment 2, the difference between the α and β training groups was, again, not statistically significant, $t(22) = 1.45$, *ns*. The figures for the four passive training groups were as follows: experiment 1 compatible = 86.2%, incompatible = 86.2%; experiment 2: α = 90.8%, β = 92.3%. These differences were, again, not statistically significant, $t(22) = 0.0$, *ns*; $t(22) = 0.53$, *ns*.

Before analyzing the effect of training, baseline behavior was established for each stimulus using the no training condition (see Figure 4). In experiment 1, untrained participants showed a slight tendency to perceive the front face of the stimulus as moving in the same direction as their hand. In practice this "compatibility effect" was small and not statistically significant for either the cube, $t(11) = 1.18$, *ns*, or sphere, $t(11) = 2.12$, *ns*. In experiment 2, the trend was for participants to pair upward motion with an outward joystick rotation. The relation was statistically significant for the cube, $t(11) = 4.28$, $p < 0.01$, $d = 1.23$, but not for the sphere, $t(11) = 1.09$, *ns*. It appears, therefore, that despite dissociating the direction of arm motion from that of the front surface of the stimulus object, the design of experiment 2 has uncovered a new perceptual bias/preference in our participants. This might be due to a biomechanical coupling between a lifting motion and wrist/forearm supination (rotation away from the body midline) and the converse in pronation (rotation towards the body midline), in other words participants inadvertently generated a lifting force when supinating. Alternatively, participants may associate supination with a lifting action because they regularly experience this association when pulling the lid off a box, for example. This could even be taken as evidence of cue recruitment. Although the source of the bias is unknown, it highlights the importance of establishing a perceptual starting point/baseline in these types of experiment.

Data from the test trials in experiment 1 appear in Figure 5a. The data were compared using a sequence of four 2 × 2 between-group ANOVAs, one for each
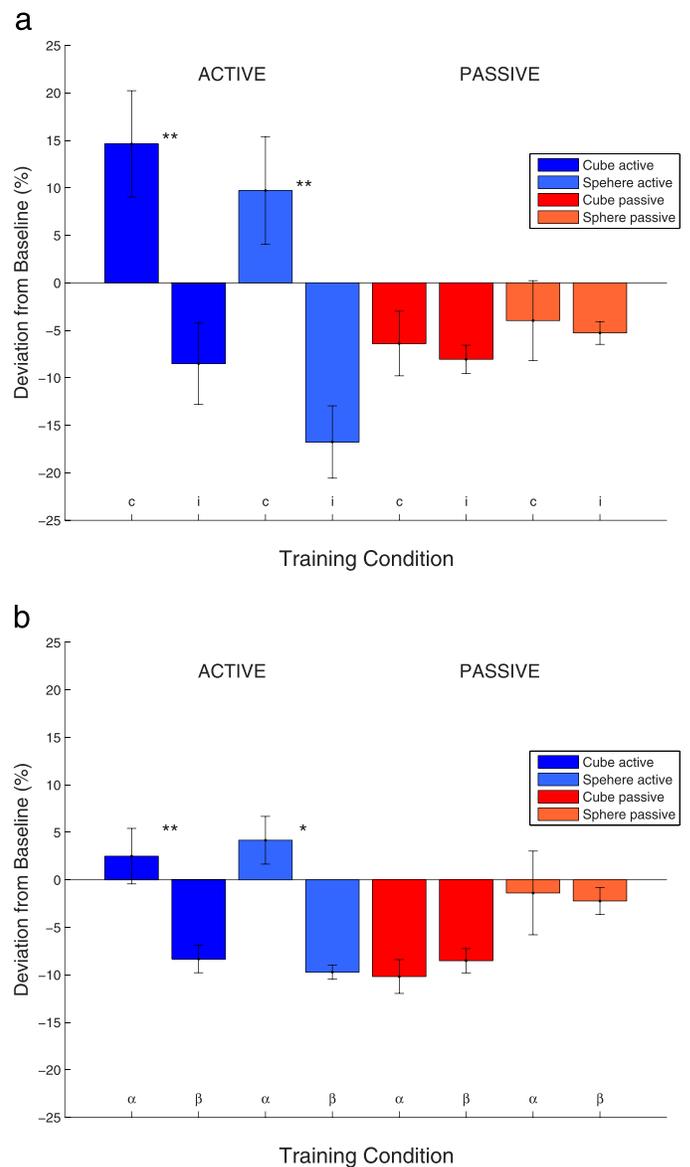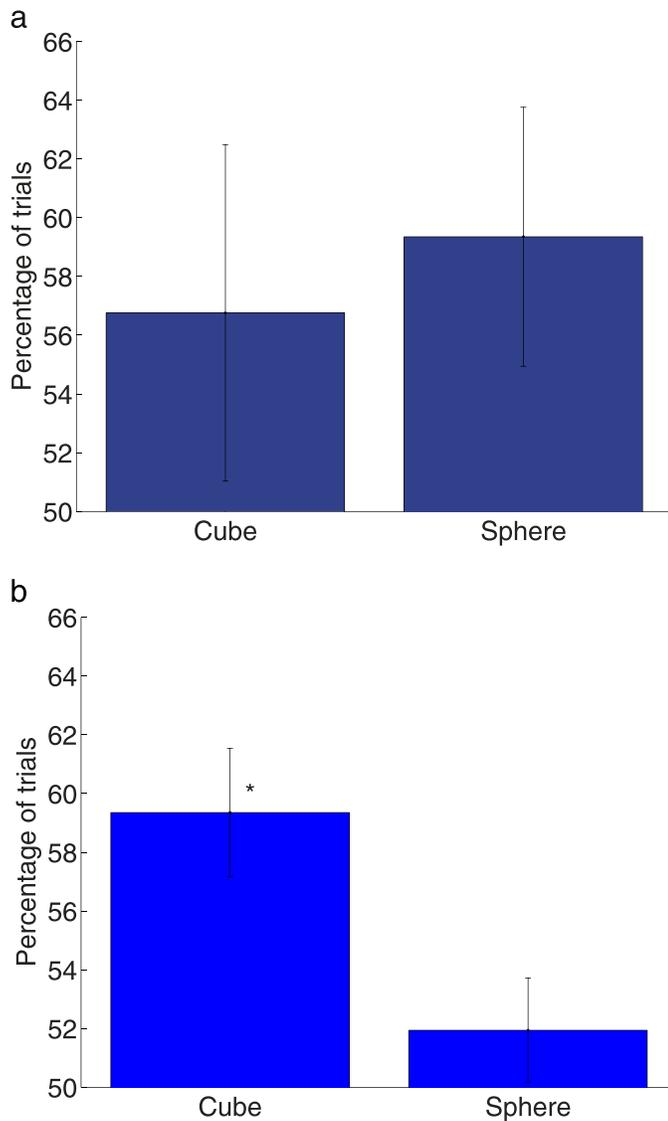
a

b

Figure 4. Baseline data for the two experiments (a) Results from the no training group in experiment 1. For both stimuli there is evidence of a preference in untrained observers to perceive the compatible relationship, such that when the hand moves rightward, so the front face of the test stimulus appears to move rightward and vice versa; however, neither was statistically significant across 12 observers. (b) Results from the no-training group in experiment 2 in which there was a preference in the untrained observers to see upward movement of the front face of the stimulus when supinating (rotating the wrist/forearm outward), and vice versa. The preference was statistically significant for the cube stimulus, based on one sample *t* tests (*$p < 0.01$. Error bars indicate *SEM*).

→

Figure 5. (a) Results from experiment 1 in which the test stimuli rotated about the vertical axis, resulting in perception of the front face of the object moving in either a leftward or rightward direction. The Figure reports the percentage of test trials in which participants reported seeing the direction of

a

b

motion of the test stimulus correspond to the direction of motion of their hand, relative to the no-training group (which acted as a baseline). Note that negative values indicate a tendency to perceive object rotation opposite to the direction of motion of the hand. Letter codes correspond to "c" = compatible motion training group and "i" = incompatible motion training group, for each combination of stimulus and passive versus active training (see Procedures). Only the active training group (blue bars) yielded significant effects of training on perceptual choices. (b) Results from experiment 2 in which the test stimulus rotated about the horizontal axis, resulting in perception of an upward or downward motion of the front face of the test stimulus. Letter codes "α" and "β" correspond to the two training groups, who experienced opposite object/arm motion contingencies during training, analogous to the compatible versus incompatible comparison of experiment 1 (see procedures; *$p < 0.05$, **$p < 0.01$. Error bars indicate *SEM*).

←

Expt.	Stimulus	$F(1, 44)$	p	η_p^2	$t(22)$	p	Hedges's g
1	Cube	7.2626	0.010	0.14	3.53	0.002	1.39
	Sphere	9.6269	0.003	0.18	3.75	0.001	1.48
2	Cube	10.1505	0.003	0.19	3.17	0.004	1.25
	Sphere	6.0177	0.018	0.12	2.74	0.012	1.08

Table 1. Results from four 2 × 2 ANOVAs testing for an interaction between training group (compatible vs. incompatible in experiment 1, or α vs. β in experiment 2) and training action (active vs. passive). *Notes*: The table also provides the outcome of a follow-up, simple main effects analysis (using t tests with adjusted alpha level of 0.025) for the difference between training groups under active training. Note that in all cases the interaction was significant, and in every case this was driven by a significant difference under active training. This effect was absent under all four of the passive training conditions, $-0.8 < t(22) < 0.8$ and $p > 0.4$.

stimulus in each experiment. The analysis included compatible/incompatible and active/passive as independent variables. In every case there was a statistically significant interaction. Analysis of the data from experiment 1 revealed that responses for the active training group were significantly different for both the cube and the sphere (see Table 1). In the case of the passive training groups, the difference was small and not statistically significant. The data for experiment 2 appear in Figure 5b. Analysis of the data from this experiment revealed a similar picture in which a significant impact of training was only measurable in the case of active training.

The design employed in these experiments also permits investigation of two other properties of the learning: (a) the speed of recruitment and (b) its stability over repeated exposure to the ambiguous version of the stimulus. A summary of trends within and across test blocks are plotted in Figures 6 and 7 for the active training groups. Data for each participant were fit using linear regression and the distribution of slopes compared between the compatible and incompatible (or α vs. β) training groups (see Table 2). Overall, data from the six trials within a block suggest that the impact of training gradually decreases. Despite this short-term drop, analysis by block suggests that the interleaved bouts of training cause the effect to increase over time. Note that one might have expected the block results to start at baseline. There are two likely reasons why they do not. First, it appears that the impact of learning is relatively quick. After just six training trials, behavior on the first test trials is different from the no-training baseline. Second, recall that participants saw two stimuli. It is likely that some carryover of learning from the first object to the second occurred, and hence that the participants were no longer operating from baseline when exposed to the second stimulus. As the block analysis focuses on the best fit slope for each individual, the precise baseline used is unimportant.
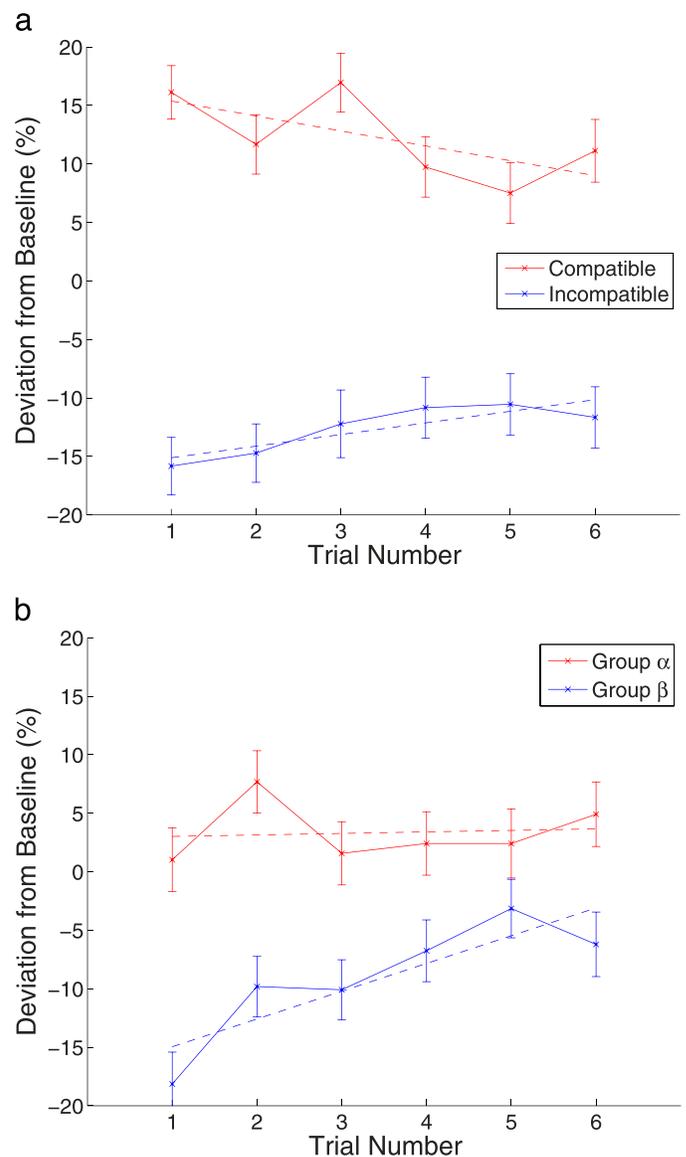
Figure 6. Impact of training on reported percepts for the ambiguous stimuli as a function of stimulus repetition (for each trial number, averaged across blocks). (a) Results for experiment 1 suggest that the effect weakens across repetition, with the size of the difference reducing. (b) In experiment 2 perceptual reports in group alpha were largely stable across trials, whereas they converged rapidly towards baseline in group β, resulting in a reduction in the difference between the two groups. For a detailed breakdown by stimulus, see the table. Error bars represent standard error of the mean.

One might expect averaging across stimulus test order to introduce greater variability in the amplitude, but not the sign, of the slope.

The final data to be analyzed are the perceptual responses reported during the recentering movement of the joystick. Cue recruitment predicts that participants should report perceiving the test stimulus rotating in the opposite direction to that reported during outward
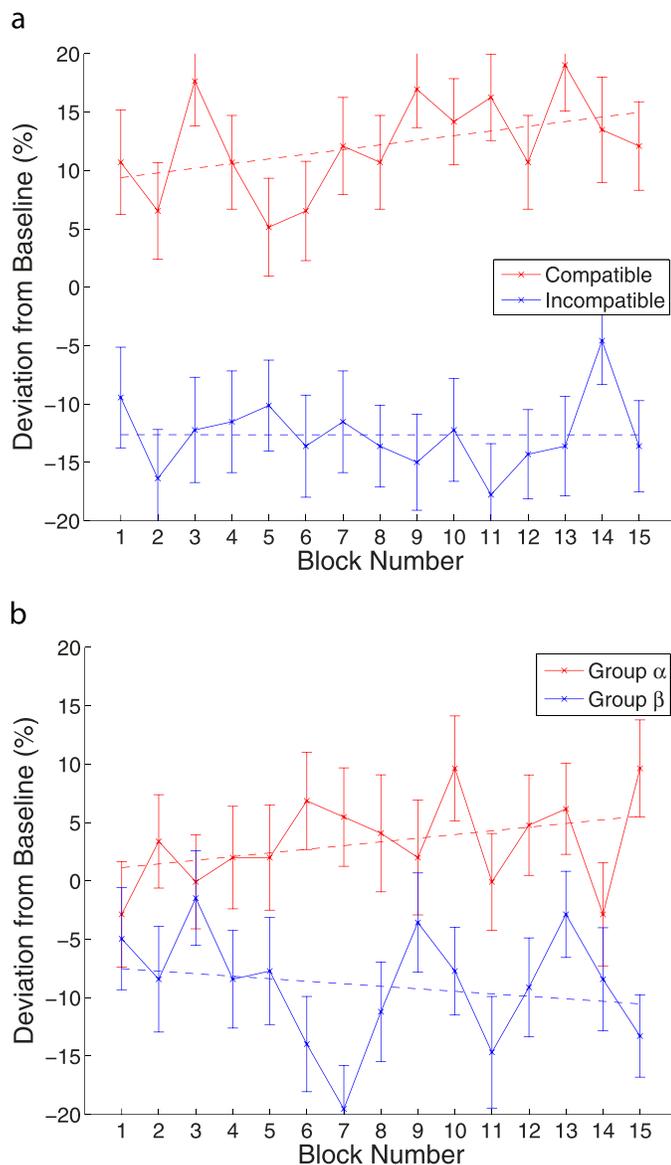
a

b

Figure 7. Impact of training on reported percepts for the ambiguous stimuli as a function of training block. (a) Results for experiment 1 suggest that the trend is for the difference to grow, on average, across blocks. (b) A similar pattern of increasing impact of training is seen in experiment 2. For a detailed breakdown by stimulus, see the table. Error bars represent standard error of the mean.

motion of the joystick, since the joystick is now being moved in the opposite direction. The results appear in Figure 8 for the active, passive, and no training conditions. Note that it is possible to collapse across compatible/incompatible trials in this case as we are only interested in a reversal of direction, not the direction per se. On initial inspection, the hypothesis appears largely confirmed for the case of the cube stimulus, since the active and passive results are well above chance (50%). However, because the reversal of direction is also seen in the no training condition, there

is no evidence that training has increased the likelihood of seeing the object reverse directions. This is possibly because reports are approaching ceiling. As described earlier, even in the presence of an occluder, the predicted percept was only evoked on around 90% of trials.

In the case of the sphere stimulus, in experiment 1 there is a tendency (although not statistically significant) for participants to report object rotation in the *same* direction as during the initial joystick movement across all three training conditions (active, passive, and no training). Note, however, that this tendency is reduced in the case of the trained groups relative to the no training baseline. In experiment 2 the difference between trained and untrained participants was large enough to be statistically significant (see Figure 8b).

One possible explanation for these mixed results is that for the sphere stimulus, at least, there is a large carryover effect from the period of outward joystick movement to the return phase. It is well established in the literature on bistable perceptual phenomena that intermittent breaks in stimulus presentation tend to lead to stabilization of a single percept (Leopold, Wilke, Maier, & Logothetis, 2002; Maloney, Dal Martello, Sahm, & Spillmann, 2005; Orbach, Ehrlich, & Heath, 1963). It is possible that the pause between outward and return joystick movements acted as a break of this kind. Nonetheless, when the impact of this "hysteresis" effect is large (as was the case for the sphere stimulus), training is seen to reduce its effect. Of course this begs the question of why hysteresis did not affect the cube stimulus in the same way. Although a much weaker effect, one might speculate that it did impact the perceptual reports for the cube as well, limiting the reports of perceived rotation reversals to around 80% of trials.

Discussion

The results of both experiments demonstrate that it is possible to couple a voluntary movement to the perceived spin direction of a synchronously rotating visual stimulus. Previous research has found that limb movement can influence perception of an ambiguous stimulus (Maruya et al., 2007), but here we demonstrate that this link can be modified. Further, the results suggest that limb movement played a crucial role in recruitment in this instance, as passive exposure to a visual cue (direction arrow) was insufficient to drive a measurable effect. This is interesting because in the strictest interpretation of cue recruitment theory, these passively observed cues also have the capacity to predict stimulus movement reliably, and hence should

Factor	Expt.	Stimulus	df	t	p	Hedges's g
Trial	1	Cube	22	−1.32	0.201	−0.52
		Sphere	22	−3.02	0.006**	−1.19
	2	Cube	22	−1.47	0.156	−0.58
		Sphere	22	−2.21	0.038*	−0.87
Block	1	Cube	22	2.30	0.032*	0.90
		Sphere	22	2.36	0.027*	0.93
	2	Cube	22	0.93	0.367	0.36
		Sphere	22	2.45	0.023*	0.96

Table 2. Statistical tests for difference in mean regression slopes for the active training conditions (*compatible* vs. *incompatible* in experiment 1 and *contingency α* vs. *contingency β* in experiment 2). *Notes*: The difference in slope is seen to always be negative across trials, consistent with converging responses and hence a weakening of the effect of training across repetitions, although the differences do not achieve statistical significance for the Cube stimulus. Conversely, the difference in slope is seen to be always positive across blocks, consistent with diverging responses and hence an increase of the effect across time. Three of these trends were statistically significant (*$p < 0.05$, **$p < 0.01$).

also have been recruited. Instead, only after active training was a strong and reliable effect observed.

So why did the participants in the passive conditions show little or no cue recruitment? As mentioned in the Introduction, Orhan et al. (2010) have argued that there must be a mechanism for determining which cues can be combined to create a meaningful interpretation of the sensory array. In the context of this study it would appear that passive viewing of the rotating object and the contingent arrows, does not satisfy this mechanism's requirements. This is perhaps because the arrows are regarded as extrinsic to the stimulus and hence unfavored for recruitment (Jain et al., 2014). However, it would appear that coincident stimulus and hand motion does favor recruitment, providing the first evidence of successful motor-visual cue recruitment (Jain et al., 2014). Of course, what the current study cannot yet tell us is what particular aspect of the limb movement facilitated cue recruitment in this case. The active task differs from the passive one in numerous ways. In the active case, motor planning, motor execution, and proprioceptive feedback are all provided. It is a matter for future studies to determine which, if any, of these aspects are crucial to successful recruitment. That said, it seems unlikely that proprioception alone would suffice. Proprioceptive feedback would take time to be received and processed which would likely be after the first perceptual impression of form and motion had already been established.

One potential criticism of any study of the kind described in this paper, concerns the reliability of the subjective reports upon which the analysis is based.
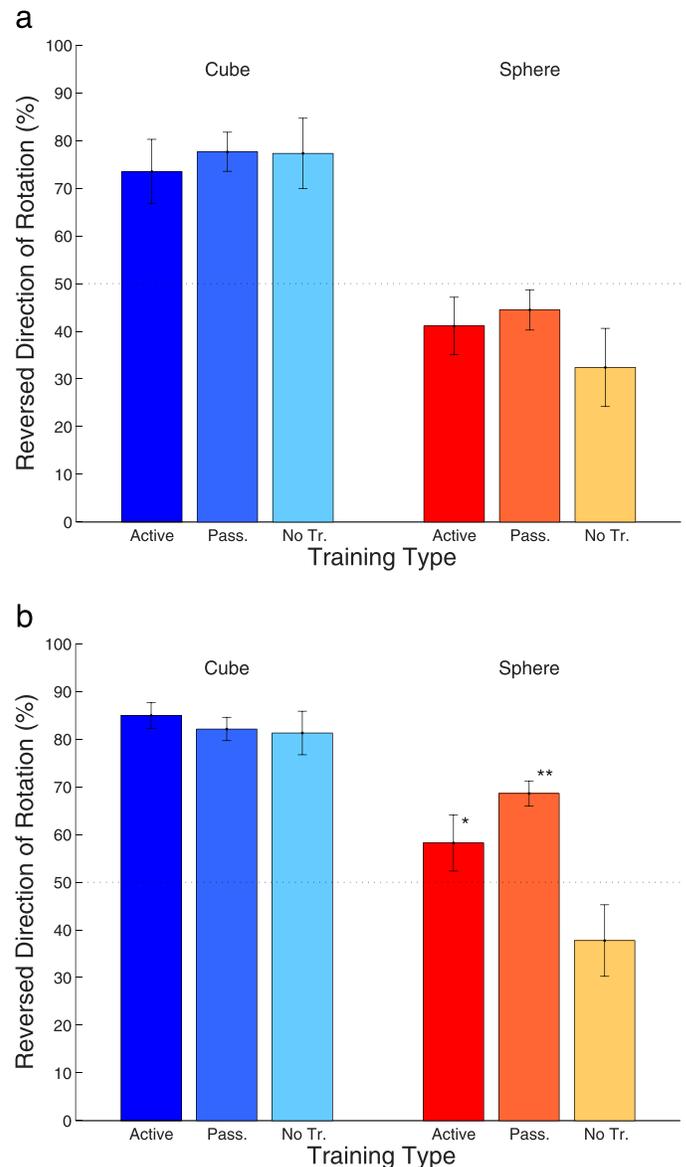

Figure 8. Percentage of trials in which the perceived direction of stimulus rotation during the second (recentering) movement of the joystick was opposite to that seen during the initial outward movement of the joystick. (a) Results for experiment 1 suggest that the perceived direction was opposite for the cube stimulus but not for the sphere stimulus. (b) A similar pattern of results emerged in experiment 2. Asterisks indicate cases in which the proportion of active or passive trials differed significantly from the associated baseline (no training) condition (*$p < 0.05$, **$p < 0.01$).

Firestone and Scholl (in press), for example, do an admirable job of dismantling much of the literature on top-down influences on perception on the grounds of response bias. As part of their detailed, wide-ranging review, they include a section on action-based influences on perception. Their critique focuses largely on subjective reports of perceived distance or of a hill's slope as a function of perceived task difficulty (Proffitt

& Linkenaugher, 2013; Proffitt, Stefanucci, Banton, & Epstein, 2003; Sugovic & Witt, 2013). The authors argue that much of this literature is polluted with response bias and that suitable control studies have undermined many of the earlier findings. Although we regard cue-recruitment as a cognitively impenetrable, bottom-up process, the question remains as to whether their criticisms might also apply to this work. For example, what about response bias? Perhaps the participants were simply responding in a manner they thought the experimenters expected them to. One simple counter to this is that if participants were merely trying to respond in a manner they thought we might expect, there is no reason why they would not have done so in the passive conditions too, since the test trials were identical for both groups.

One striking feature of the data is the time course of the learning which mirrors a previous finding in the cue recruitment literature: The learned contingency has at least two phases, a fast phase and a slow phase. The fast phase is very fast indeed, such that just a few training trials are sufficient to cause the appearance of bias on subsequent test trials. In the current work, this effect decayed somewhat over the course of a few repeat trials (see Figure 6). However, a slow phase causes the effect to grow over time (see Figure 7). In general, it becomes increasingly difficult to reverse the learning of a new cue the longer it is trained (Backus, 2011; Harrison & Backus, 2014). Another issue raised earlier, and worth reiterating here is that repeated presentation of perceptually bistable stimuli tends to lead to stabilization of a percept (Leopold et al., 2002; Maloney et al., 2005; Orbach et al., 1963). Hence, one possible reason for the decaying impact of training trial by trial, is that training was attempting to combat this well-known perceptual stabilization effect.

Another possible criticism of the study is that although the participants' reports are a reliable reflection of their perceptual experience, the direction arrows cued them to attend to a specific part of the stimulus. As Firestone and Scholl also point out, it is well known that the decision of what is front and what is back in the Necker cube is heavily influenced by spatial attention (Peterson & Gibson, 1991; Toppino, 2003). Whereas we do not dispute this fact, this spatial attention cannot explain the results described here because the cube was presented in a random starting orientation; hence, it was not possible for the participants to know which edge should be regarded as front and which as back. Any attempt to assign depth order to the edges before movement occurred would only have served to introduce noise into the results. Indeed, although many ambiguous figures are subject to attentional biases (Long & Toppino, 2004), in the case of the (transparent) dot sphere, the choice of which points

to attend to is even less clear. An explanation along these lines would also need to spell out why it did not happen in the passive training case. In practice, one major advantage of the approach used in this paper is that apparent rotation direction at stimulus onset is extremely difficult to control through volition, and the task of distinguishing the direction of perceived rotation is relatively easy and clear-cut. In other words, during this study a participant's visual system presumably made a cognitively impenetrable dichotomous decision, and the observer simply and effortlessly reported the result (Backus, 2009).

An alternative and, in our opinion, more important reason for being guarded in interpreting the results, concerns the decision to use the label "cue recruitment." Whereas the effect we observed is formally cue recruitment, using this term promotes the idea that a ubiquitous mechanism exists for linking visual perception to all manner of sensory experiences. Since this study reports the first successful example of a motor-visual effect, it may be premature to make such a claim. Perhaps there is something special about the visuomotor system that lends itself to this type of association. Some participants reported envisaging a machine containing pulleys and ropes linking the joystick to the stimulus object. Experience with the "machine" produced a strong expectation/contingency in their mind between direction of hand motion and direction of object rotation. The influence of their expectation on sensory integration could clearly also be couched in Bayesian terms, and one attempt to move beyond a merely qualitative Bayesian framework is provided by Backus (2009).

In conclusion, the present study presents evidence that a voluntary action (arm movement) can influence visual perceptual processes. We suggest that this relationship may develop through an already functional link between motor behavior and the visual system (Cisek & Kalaska, 2010; Fagioli et al., 2007; Wohlschläger & Wohlschläger, 1998). Through the associative learning paradigm used here, this relationship can be modified to enable arbitrary relationships between limb movement and perceived motion of a perceptually ambiguous stimulus. Passive exposure to the contingency was not sufficient to generate a reliable change in the visual percept, suggesting that motor behavior can exert a strong influence over perceptual choice.

Keywords: form perception, visuomotor behavior, conditioning, cue recruitment

Acknowledgments

Commercial relationships: none.
Corresponding author: Guy Wallis; Benjamin Backus.
Email: gwallis@uq.edu.au; gwallis@hms.uq.edu.au; ben.backus@gmail.com.
Address: Centre for Sensorimotor Performance, University of Queensland, Brisbane, Australia; Graduate Center for Vision Research, State University of New York College of Optometry, New York, NY, USA.

References

Adams, W., Graf, E., & Ernst, M. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, *7*, 1057–1058.

Backus, B. T. (2009). The Mixture of Bernoulli Experts: A theory to quantify reliance on cues in dichotomous perceptual decisions. *Journal of Vision*, *9*(1):6, 1–19, doi:10.1167/9.1.6. [PubMed] [Article]

Backus, B. T. (2011). Recruitment of new visual cues for perceptual appearance. In J. Trommershauser, K. Kording, & M. S. Landy (Eds.), *Sensory cue integration* (pp. 101–119). Oxford, UK: Oxford University Press.

Backus, B. T., & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception*, *28*(2), 217–242.

Backus, B. T., & Haijiang, Q. (2007). Competition between newly recruited and pre-existing visual cues during the construction of visual appearance. *Vision Research*, *47*, 919–924.

Bekkering, H., & Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychological Science*, *13*, 370–374.

Berkeley, G. (1725). *Three dialogues between hylas and philonous : The design of which is plainly to demonstrate the reality and perfection of human knowledge, the incorporeal nature of the soul, and the immediate providence of a deity: In opposition to sceptics and atheists*. London, UK: Innys.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of proximity and of other Gestalt factors. *American Journal of Psychology*, *66*, 20–32.

Burgess-Limerick, R., Krupenia, V., Wallis, G., Pratim-Bannerjee, A., & Steiner, L. (2010). Directional control-response relationships for mining equipment. *Ergonomics*, *53*, 748–757.

Carter, O., & Cavanagh, P. (2007). Onset rivalry: Brief presentation isolates an early independent phase of perceptual competition. *PLoS ONE*, *2*(4), e343.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*, 269–298.

Cloete, S., & Wallis, G. (2009). Limitations of feedforward control in multiple-phase steering movements. *Experimental Brain Research*, *195*, 481–487.

Di Luca, M., Ernst, M., & Backus, B. (2010). Learning to use an invisible signal for perception. *Current Biology*, *20*, 1860–1863.

Ernst, M. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5):7, 1–14, doi:10.1167/7.5.7. [PubMed] [Article]

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Ernst, M., Banks, M., & Bülthoff, H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*(1), 69–73.

Ernst, M., & Bülthoff, H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169.

Fagioli, S., Hommel, B., & Schubotz, R. (2007). Intentional control of attention: Action planning primes action related stimulus dimensions. *Psychological Research*, *71*, 22–29.

Fahle, M., & Poggio, T. (Eds.). (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Firestone, C., & Scholl, B. (in press). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, E-pub ahead of print.

Fitts, P. M., & Seeger, C. M. (1953). S-r compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*(3), 199–210.

Gibson, J., & Gibson, E. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*, 32–41.

Haijiang, Q., Saunders, J.A., Stone, R.W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of pavlovian conditioning. *Proceedings of the National Academy of Sciences, USA, 103*, 483–488.

Harrison, S. J., & Backus, B.T. (2012). Associative learning of shape as a cue to appearance: A new demonstration of cue recruitment. *Journal of Vision, 12*(3):15, 1–17, doi:10.1167/12.3.15. [PubMed] [Article]

Harrison, S. J., & Backus, B. T. (2014). A trained perceptual bias that lasts for weeks. *Vision Research, 99*, 148–153.

Hebb, D. O. (1949). *The organization of behavior*. New York, NY: Wiley.

Hentschke, H., & Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *European Journal of Neuroscience, 34*(12), 1887–1894.

Ho, Y. X., Serwe, S., Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2009). The role of visuohaptic experience in visually perceived depth. *Journal of Neurophysiology, 101*(6), 2789–2801.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8*(11), 494–500.

Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research, 71*, 42–63.

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioural and Brain Sciences, 24*, 849–937.

Howells, T. H. (1944). The experimental development of color-tone synesthesia. *Journal of Experimental Psychology, 34*, 87–103.

Jacobs, R. (2002). What determines visual cue reliability. *Trends in Cognitive Sciences, 6*, 345–351.

Jacobs, R., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research, 39*(24), 4062–4075.

Jain, A., Fuller, S., & Backus, B. (2010). Absence of Cue-recruitment for extrinsic signals: Sounds, spots and swirling dots fail to influence perceived 3D rotation direction after training. *PLoS ONE, 5*, 1–7.

Jain, A., Fuller, S., & Backus, B. T. (2014). Cue-recruitment for extrinsic signals after training with low information stimuli. *PLoS ONE, 9*(5), e96383.

James, W. (1890). *The principles of psychology* (1983 ed.).Cambridge, MA: Harvard University Press.

Johnston, E., Cumming, B., & Landy, M. (1994).

Integration of motion and stereopsis cues. *Vision Research, 34* , 2259–2275.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*(6971), 244–247.

Landy, M., Maloney, L., Johnston, E., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35*, 389–412.

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature Neuroscience, 5*(6), 605–609.

Locke, J. (1706). *An essay concerning human understanding* (5th ed.). London: printed for Awnsham and John Churchill, and Samuel Manship.

Long, G., & Toppino, T. (2004). Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin, 130*(5), 748–768.

Maloney, L., Dal Martello, M., Sahm, C., & Spillmann, L. (2005). Past trials influence perception of ambiguous motion quartets through pattern completion. *Proceedings of the National Academy of Sciences, USA, 102*(8), 3164–3169.

Maruya, K., Yang, E., & Blake, R. (2007). Voluntary action influences visual competition. *Psychological Science, 18*(12), 1090–1098.

Orbach, J., Ehrlich, D., & Heath, H. A. (1963). Reversibility of the Necker Cube: I. An examination of the concept of satiation of orientation. *Perceptual and Motor Skills, 17*(2), 439–458.

Orhan, A. E., Michel, M. M., & Jacobs, R. A. (2010). Visual learning with reliable and unreliable features. *Journal of Vision, 10*(2):2, 1–15, doi:10.1167/10.2.2. [PubMed] [Article]

Peterson, M., & Gibson, B. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape description. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 170.

Proffitt, D. R., & Linkenauger, S. A. (2013). Perception viewed as a phenotypic expression. In W. Prinz, M. Beisert, & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 171–197). Cambridge, MA: MIT Press.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science, 14*(2), 106–112.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Schubo, A., Prinz, W., & Achersleben, G. (2004).

Perceiving while acting: Action affects perception. *Psychological Research*, 68, 208–215.

Schubotz, R. I., & von Cramon, D. Y. (2002). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: An fMRI study. *NeuroImage*, 15, 787–796.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature*, 384(6608), 460–463.

Sugovic, M., & Witt, J. K. (2013). An older view on distance perception: Older adults perceive walkable extents as farther. *Experimental Brain Research*, 226(3), 383–391.

Toppino, T. C. (2003). Reversible-figure perception: Mechanisms of intentional control. *Perception & Psychophysics*, 65(8), 1285–1295.

Teramoto, W., Hidaka, S., & Sugita, Y. (2010). Sounds move a static visual object. *PLoS One*, 5(8), e12255.

van Dam, L. C., & Ernst, M. O. (2010). Preexposure disrupts learning of location-contingent perceptual biases for ambiguous stimuli. *Journal of Vision*, 10(8):15, 1–17, doi:10.1167/10.8.15. [PubMed] [Article]

von Helmholtz, H. (1878). *Die Tatsachen in der Wahrnehmung* [Translation: Treatise on Physiological Optics]. Berlin, Germany: Hirschwald.

von Helmholtz, H. (2005). Handbuch der physiologischen Optik. [Treatise on physiological optics] (J. P. C. Southall, Ed., & Trans.). Lowell, MA: Courier Corporation. (Original work published 1910).

Wallach, H., O'Connell, D., & Neisser, U. (1953). The memory effect of visual perception of three-dimensional form. *Journal of Experimental Psychology*, 45(5), 360–368.

Wallis, G., & Bülthoff, H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, USA, 98*(8), 4800–4804.

Wallis, G., Chatziastros, A., & Bülthoff, H. (2002). An unexpected role for visual feedback in vehicle steering control. *Current Biology*, 12(4), 295–299.

Wallis, G., Chatziastros, A., Tresilian, J., & Tomasevic, N. (2007). The role of visual and non-visual feedback in a vehicle steering task. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1127–1144.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 397.

Worringham, C. J., & Beringer, D. B. (1998). Directional stimulus-response compatibility: A test of three alternative principles. *Ergonomics*, 41(6), 864–880.

Wykowska, A., Hommel, B., & Schubö, A. (2011). Action-induced effects on perception depend neither on element-level nor on set-level similarity between stimulus and response sets. *Attention, Perception, & Psychophysics*, 73(4), 1034–1041.

Wykowska, A., Schubö, A, and Hommel, B. (2009). How you move is what you see: Action planning biases selection in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1755–1769.

Appendix

There are many examples in the literature on bistable perception, of long-lasting individual differences and biases, not least for the Necker Cube stimulus used in this paper (Carter & Cavanagh, 2007; van Dam & Ernst, 2010). For completeness, this appendix provides individual data for participants in the active training groups and in the no-training group. These data are useful in determining whether inherent or early biases are retained and the extent to which early biases can be overcome though the cue recruitment training. To establish an estimate of initial biases, the responses from the first five blocks of testing (30 trials) were compared to responses in the final five blocks.

Figure A1 reports the shifts in responses in the no training trials. This offers a baseline for comparison to behavior seen in the training trials. Figures A2 and A3 provide data from actively trained individuals who took part in experiments 1 and 2 respectively. Reviewing the data as a whole, one can occasionally see evidence of widening in the distribution of responses over time, e.g., Group β viewing the sphere in experiment 2, consistent with an entrenching of earlier preferences. At the same time, there is also evidence of a *sharpening* of the distributions e.g., in the incompatible training group viewing the cube in experiment 1. Overall there is no consistent pattern of shift in the distributions of responses across the two stimuli, the two experiments, or across training type (compatible vs. incompatible or α vs. β). The one consistent pattern that emerges from the data is that the distribution peaks diverge over time, a fact which is entirely consistent with the predictions of cue recruitment theory.
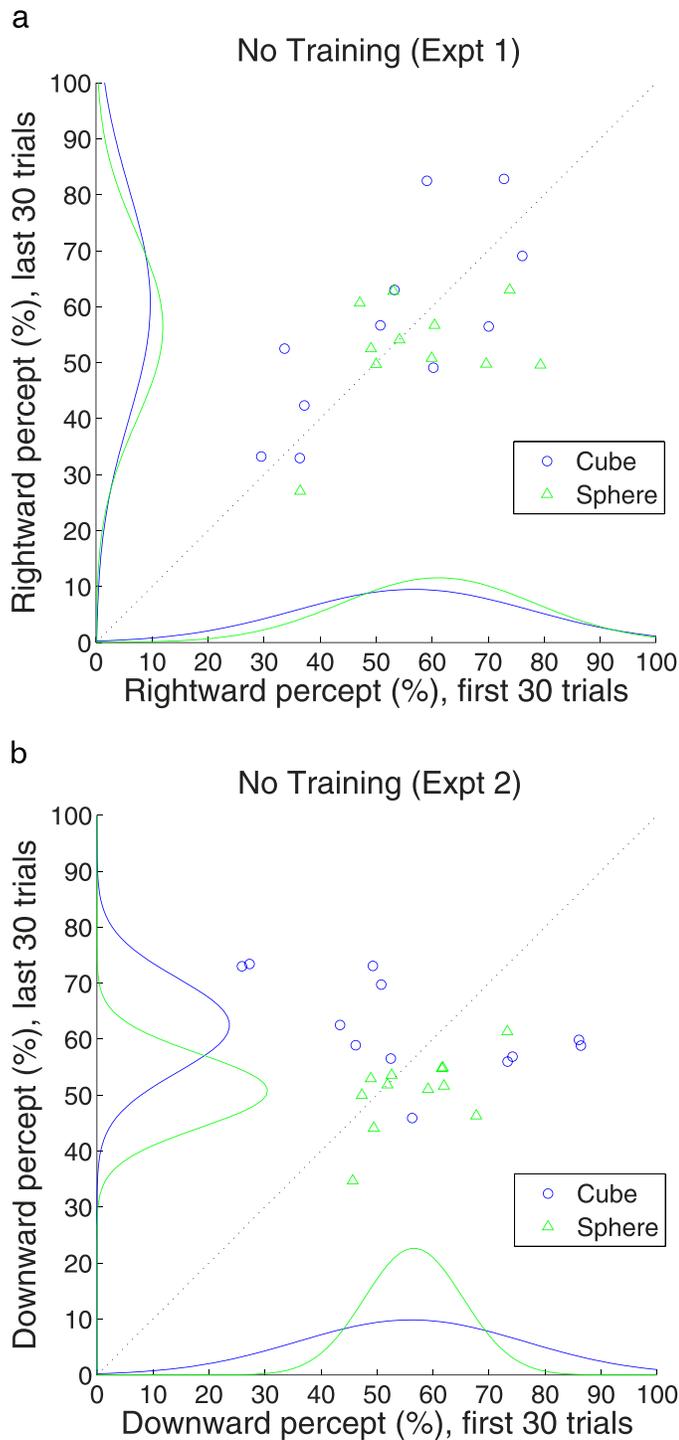
a

b

Figure A1. Baseline data for the two experiments at the beginning and after substantial exposure to the stimuli (a) Results for the cube and sphere in experiment 1 for the first 30 trials (*x* axis) and last 30 trials (*y* axis) across all twelve participants. The averages and overall distribution remain comparable across the whole experiment. (b) Results for the cube and sphere in experiment 2 for the first 30 trials (*x* axis) and last 30 trials (*y* axis) across all twelve participants. The averages remain comparable across the whole experiment, but →

←

compared to experiment 1 the distribution of responses are sharper, especially after considerable exposure to the stimuli. Note that random jitter with a maximum amplitude of 1% was applied to all plot symbols to ensure that data from all twelve participants is visible.
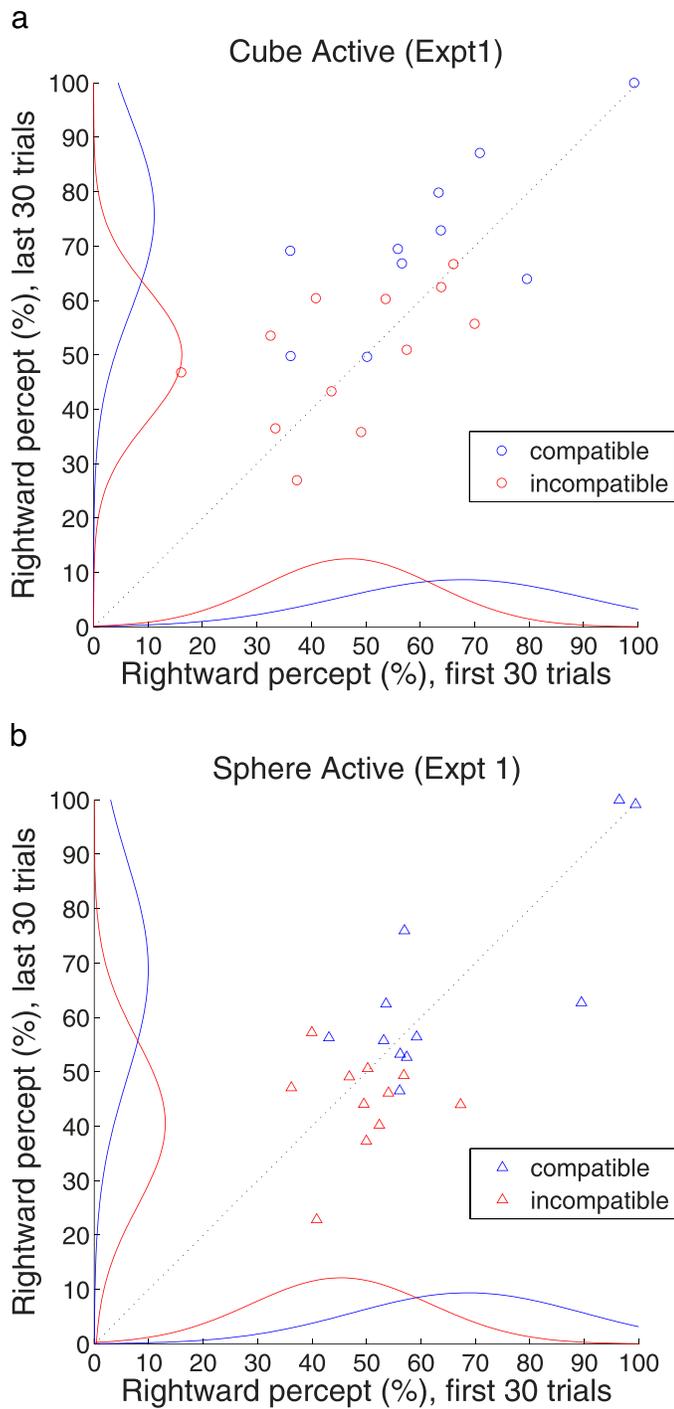
Figure A2. Individual data for the two stimuli in the active training condition averaged over the first 30 trials (*x* axis) and the last 30 trials (*y* axis) in experiment 1. (a) Results for the cube stimulus revealing that the two distributions of responses for compatible (blue) versus incompatible (red) remain comparable in form over time. However, separation between the two distributions increases over time. (b) Results for the sphere stimulus, revealing a similar pattern of results. Note that random jitter with a maximum amplitude of 1% was applied to all plot symbols to ensure that data from all twelve participants are visible.
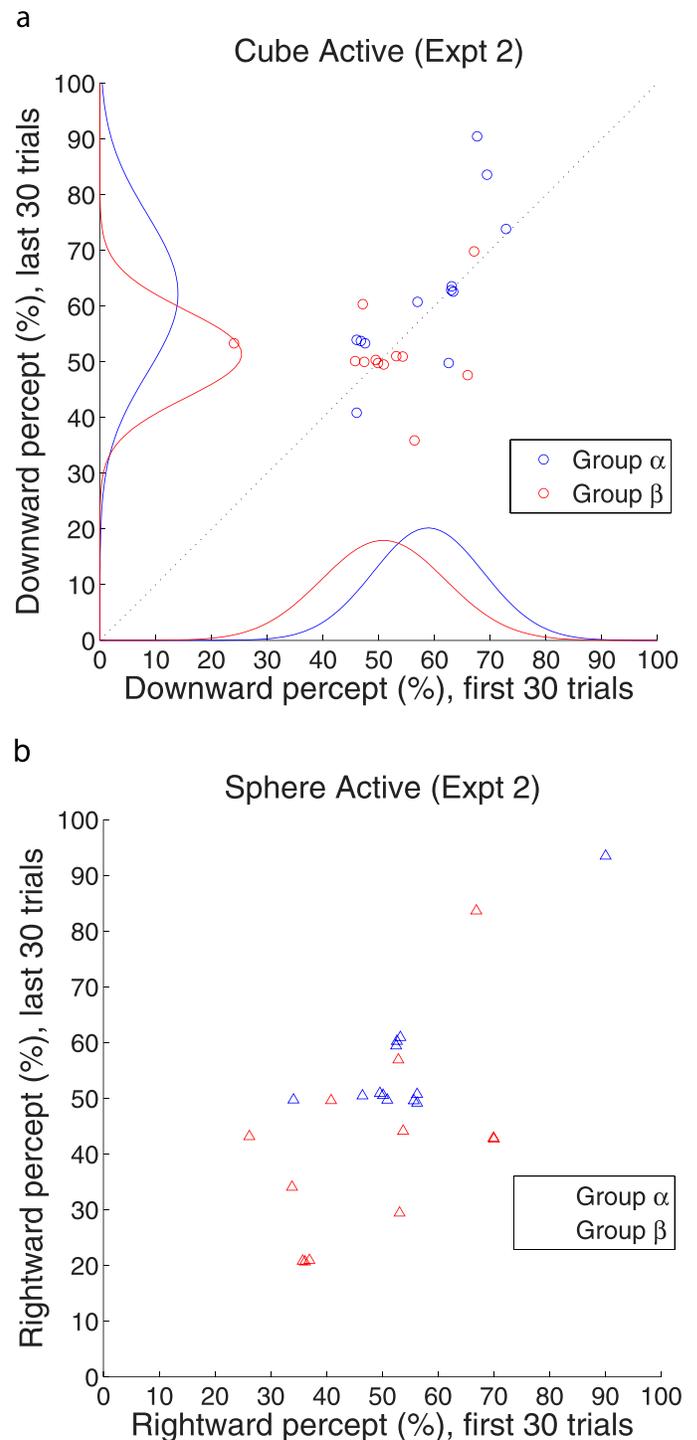
Figure A3. Individual data for the two stimuli in the active training condition averaged over the first 30 trials (*x* axis) and the last 30 trials (*y* axis) in experiment 2. (a) Results for the Cube stimulus revealing that the distribution of responses for α (blue) training widened slightly with training, whereas it narrowed in the β (red) case. As in experiment 1, separation between the two distributions increases over time. (b) Results for the Sphere stimulus, revealing changes in response distribution across time, but the same growth in peak separation seen for the Cube. Note that random jitter with a maximum amplitude of 1% was applied to all plot symbols to ensure that data from all twelve participants are visible.