1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17 Cue-selective adaptation operates on a separable encoding of stereo and texture information
18
19 Evan Cesanek[1] and Fulvio Domini[1]
20
21 [1] Department of Cognitive, Linguistic, & Psychological Sciences
22 Brown University
23 Providence, RI, 02906

1   **Abstract**
2   To perform accurate movements, the sensorimotor system must maintain a delicate calibration of
3   the mapping between visual inputs and motor outputs. Previous work has focused on the
4   mapping between visual inputs and individual locations in egocentric space, but little attention
5   has been paid to the mappings that support interactions with 3D objects. In this study, we
6   investigated sensorimotor adaptation of grasping movements targeting the depth dimension of
7   3D paraboloid objects. Object depth was specified by separately manipulating binocular disparity
8   (stereo) and texture gradients. At the end of each movement, the fingers closed down on a
9   physical object consistent with one of the two cues, depending on the condition (haptic-for-
10   texture or haptic-for-stereo). Unlike traditional adaptation paradigms, where relevant spatial
11   properties are determined by a single dimension of visual information, this method enabled us to
12   investigate whether adaptation processes can selectively adjust the influence of different sources
13   of visual information depending on their relationship to physical depth. In two experiments, we
14   found short-term changes in grasp performance consistent with a process of cue-selective
15   adaptation: the slope of the grip aperture with respect to a reliable cue (correlated with physical
16   reality) increased, whereas the slope with respect to the unreliable cue (uncorrelated with
17   physical reality) decreased. In contrast, slope changes did not occur during exposure to a set of
18   stimuli where both cues remained correlated with physical reality, but one was rendered with a
19   constant bias of 10 mm; the grip aperture simply became uniformly larger or smaller, as in
20   standard adaptation paradigms. Overall, these experiments support a model of cue-selective
21   adaptation driven by correlations between error signals and input values (*i.e.*, supervised
22   learning), rather than mismatched haptic and visual signals.
23

1. Introduction

1.1. Sensorimotor adaptation and 3D shape perception

To ensure that actions produce their intended outcomes, our brains must keep our movements calibrated to the physical environment despite enormous moment-to-moment variability in postural dynamics and neural processing. Over the last 100+ years, research into the process of sensorimotor adaptation has produced a great deal of evidence illuminating when and how the brain adjusts motor planning processes based on sensory feedback (Helmholtz, 1909/1962; Held & Hein, 1958; Welch, 1969; 1978; Cunningham, 1989; Redding & Wallace, 1997; Krakauer, Pine, Ghilardi, & Ghez, 2000; Krakauer, Ghez, & Ghilardi, 2005; Taylor, Krakauer, & Ivry, 2014). However, the role of sensorimotor adaptation in compensating for distortions of 3D shape perception has remained largely unexplored.

When we view and interact with real physical objects, motor control relies on visual information about 3D properties (*e.g.*, slant, curvature, and depth), which are specified by multiple visual cues (*e.g.*, stereo, motion, and texture information). As viewing conditions change from one environment to the next, each of these cues is affected by specific distortions that modulate the amount of bias and noise in their neural encodings. Given the random nature of cue-specific distortions, it is unlikely that the visual system learns to predict the relative accuracies and reliabilities of all depth cues in all possible viewing situations and account for these in the cue-combination process. Thus, rather than providing unwaveringly accurate estimates of 3D shape, visual perception is subject to shifting patterns of inaccuracy as viewing conditions change. For example, an object viewed from a close distance will often be perceived to be deeper than it truly is, due to a constant bias in processing depth from binocular disparities (Johnston, 1991). In another situation, unreliable texture processing could cause significant variable error in the perceived depths of objects with unusual surface markings (Rosenholtz & Malik, 1997; Todd, 2004).

Since the visual guidance of skilled actions like grasping fundamentally depends on estimates of 3D shape, changes in viewing conditions can have real, detrimental consequences on movement accuracy. Consider an observer who suddenly changes the environment where he operates, perhaps moving from a naturally illuminated café during his lunch break to the dimly lit bar where he works. In the café, as he selects a piece of fruit from a nearby bowl, he relies on clear shading gradients and consistent texture patterning on the surfaces of the assorted apples, oranges, and bananas. Later, back in the dim and artificially lit bar, the beer glasses he delivers to patrons are transparent, with unusual specular reflections off their surfaces rather than consistent textures. To avoid fumbling with the glassware, his visuomotor system must quickly recalibrate itself in order to process this new collection of depth cues properly, transforming the available depth information into appropriate grip apertures and orientations when grasping.

The process of sensorimotor adaptation suggests one way to resolve movement errors arising from cue-specific distortions: recalibrate movement planning processes based on sensory feedback from ongoing interactions (Cesanek, Taylor, & Domini, 2019a). However, since multiple depth cues are involved, there is the additional requirement that this recalibration must occur with respect to each individual cue. For example, if stereo information reliably specifies variations in 3D structure, but texture signals are noisy or biased, the relative influence of stereo information in movement planning should be increased. Here, we probed the conditions under which adaptation processes can selectively adjust the influence of individual depth cues when performing actions that require 3D information, following up on promising findings from a previous investigation (Cesanek *et al*., 2019a). We refer to this process as *cue-selective*

1   *sensorimotor adaptation*, in contrast to the canonical form of adaptation involving a constant
2   shift of motor outputs in one direction.
3
4   *1.2. Identifying the mechanism of cue-selective adaptation*
5   As discussed above, the 3D objects that we grasp and manipulate are always specified by
6   multiple depth cues that can become distorted independently. Thus, the ability to modulate the
7   influence of individual cues from one situation to the next via cue-selective adaptation could be
8   an important capability of biological sensorimotor systems. Experiment 1 was designed to show
9   that the planned grip apertures of grasping movements are affected by cue-selective adaptation
10   when stereo or texture cues to depth suddenly become uncorrelated with physical object depth.
11   To do this, the sensorimotor system requires a credit assignment mechanism that
12   determines which cue is responsible for detected movement errors. Experiment 2 aimed
13   distinguish between two possible mechanisms for identifying reliable versus unreliable cues and
14   increasing or decreasing their respective influences on the motor output. One possible
15   mechanism involves cross-modal comparisons, where the depth signal from each cue is
16   compared to a haptic "ground truth" signal at the end of each grasp. This mechanism would
17   produce cue-selective adaptation whenever a depth cue fails to indicate the true physical depth.
18   An alternative possibility is an error-based mechanism, which should fail to produce cue-
19   selective adaptation when depth cues are affected by constant biases because these errors can be
20   quickly corrected by canonical sensorimotor adaptation. Under this mechanism, we should
21   observe cue-selective adaptation only when variable errors occur due to an unreliable cue.
22   The main benefit of the error-based mechanism is that it does not require a ground truth
23   teaching signal to be available in order to reduce the influence of faulty sources of input
24   information. Faulty inputs are identified by their positive correlation with error signals: when a
25   particular cue becomes very noisy, it will take on spurious large values and spurious low values,
26   which will cause positive and negative movement errors, respectively. On the other hand, in the
27   case of a constant bias, there is no systematic variability in the error signals, so it is not possible
28   to identify the faulty cue. As a result, an error-based mechanism would fail to selectively change
29   the influences of the cues. Experiment 2 was designed to test whether cue-selective adaptation
30   occurs (a) even in the case of cue-specific bias or (b) only in response to altered correlations
31   between individual depth cues and physical depth.
32   In the next section, we describe a model of the process of interacting with 3D objects
33   under cue-specific distortions. This modeling yields two observations that favor the error-based
34   learning mechanism on grounds of ecological function and parsimony. First, we show that
35   canonical adaptation would be sufficient to correct for movement errors when one cue is afflicted
36   by a constant bias, causing the perceived depths of all objects to be uniformly over- or
37   underestimated. In this case, cue-selective adaptation would not be necessary to achieve accurate
38   movements. Second, when canonical adaptation is rendered ineffective by conflicting error
39   signals arising from an unreliable cue, a well-known online supervised learning rule could vary
40   the influence of multiple sources of information based on their long-term correlations with error
41   feedback signals. This mechanism could enable cue-selective adaptation without assuming the
42   availability of directly comparable metric depth estimates from different sensory modalities.
43
44   *1.3. Theoretical model*
45   The visuomotor mapping that mediates grasp planning has previously been shown to be a
46   roughly linear function of object size (Jeannerod, 1981; Säfström & Edin, 2005; but see Smeets
47   & Brenner, 1999 for an alternative view). The output of this function is a target grip aperture,
48   often operationalized as the maximum grip aperture (MGA) of the movement, which occurs just

1 before the hand closes down on the object. At the moment of the MGA, the hand posture must
2 (a) be scaled to the size of the object and (b) provide a "safety margin" so that the fingers enclose
3 the object before clamping down to apply grip force; these two functional characteristics
4 correspond to the slope and intercept parameters of a linear function. To extend this formulation
5 to grasping 3D objects defined by multiple depth cues, we can simply introduce additional slope
6 parameters, which will capture the way that the MGA responds to variations in each of these
7 cues; the intercept remains as a way to introduce a constant offset beyond the observed visual
8 size (*i.e.*, a safety margin).
9   In this linear model, uniformly shifting the motor output—what we have called canonical
10 adaptation—corresponds to an adjustment of the intercept parameter, whereas cue-selective
11 adaptation is captured by adjustments of the separate slope parameters associated with the
12 available visual cues:
13 1. $y_n = k_{S_n} z_S + k_{T_n} z_T + x_n$
14 where y_n is the planned motor output on trial n, z_S and z_T are the depth values indicated by two
15 distinct visual cues (stereo and texture information), k_{S_n} and k_{T_n} are the slope parameters
16 associated with those cues, and x_n is the intercept parameter. The process of sensorimotor
17 adaptation adjusts the parameters of this mapping according to error feedback signals: canonical
18 adaptation adjusts the intercept, producing constant shifts of the motor output, while cue-
19 selective adaptation adjusts the slopes, altering the relative influences of the available cues on the
20 motor output.
21   Grasp adaptation is driven primarily by haptic feedback from object contact at the end of
22 each grasp. Specifically, unexpected temporal patterning of the movement, where fingertip
23 contact occurs sooner or later than expected, has been shown to correlate with adaptive changes
24 in the MGA (Säfström & Edin, 2008). Thus, for a given object size, there will be some desired
25 MGA y^* that produces the preferred temporal patterning, allowing a stable and comfortable
26 grasp. Errors arise when the actual output does not match this target value. To model this error
27 detection process, we can simply take the difference between the planned motor output y_n and
28 the target output y^* preferred for the current physical object depth z_Φ, generating a movement
29 error signal ε_n:
30 2. $y^* = f(z_\Phi)$
31 3. $\varepsilon_n = y_n - y^*$
32 Previous work on reach adaptation has clearly demonstrated that uniform shifts of the intercept
33 parameter are the result of trial-by-trial error corrections, according to some learning rate b
34 (Thoroughman & Shadmehr, 2000; Cheng & Sabes, 2006):
35 4. $x_{n+1} = x_n - b\varepsilon_n$
36 In grasping, this corresponds to a change in the safety margin, applied uniformly across all target
37 objects, regardless of the visual input information. On the other hand, changes of the cue-specific
38 slope parameters k_{S_n} and k_{T_n}, which scale the motor output in accordance with the visual inputs,
39 are not often considered in the context of sensorimotor adaptation. To gain some leverage on this
40 question, we can begin by asking when such adjustments would be necessary to support accurate
41 motor performance.
42   Applying the error-correction process described by Equations 2–4 to the standard reach
43 adaptation experiment described in section 1.2, we find that rightward errors ($y_n > y^*$) sensibly
44 cause future movements to aim more to the left, while leftward errors ($y_n < y^*$) induce shifts to
45 the right. We demonstrated previously that this same process is sufficient to correct errors that
46 occur due to simple over- or underestimation of depth from a particular cue (Cesanek *et al.*,
47 2019). This is because the faulty cue introduces a relatively constant bias to all motor outputs,

1 pushing them away from the desired output in one direction. Therefore, constant biases do not
2 demand cue-selective adaptation. Cue-selective slope changes become necessary for accurate
3 motor performance only when the strength of the correlation between a given cue and physical
4 reality is reduced, since the unreliable cue will sometimes bias outputs toward depth
5 overestimation and sometimes toward underestimation. The remedy for this more complex
6 pattern of error signals is to reduce the relative influence of this faulty cue.
7   To see how the necessary slope changes could arise through an error-based learning
8 mechanism, consider the pattern of error signals that would occur across different values of an
9 unreliable cue: spurious high values will misleadingly increase the motor output, causing
10 positive errors, whereas spurious low values will decrease the motor output, causing negative
11 errors. Thus, error signals will be positively correlated with the values of an unreliable cue. This
12 fact can be exploited to perform slope adjustments with a well-known rule for online supervised
13 learning:
14 5. $k_{S_{n+1}} = k_{S_n} - c\varepsilon_n z_S$
15 where c is a learning rate, z_S is the input from a given depth cue (in this case, stereo), and k_{S_n} is
16 the associated slope parameter on trial n. Notice that the product in the second term will be, on
17 average, positive for an unreliable cue (prior to the subtraction). This learning rule works most
18 efficiently if error signals are centered on zero, creating a sensible complementarity with the
19 rapid intercept adjustments that compensate for constant errors. Critically, this error-based
20 learning mechanism for cue-selective adaptation predicts that interfering error signals are
21 necessary to elicit slope changes (*e.g.*, the grip is too large on some trials, but too small on other
22 trials). In contrast, this mechanism would not be expected to operate when one cue is distorted by
23 a constant bias. A biased cue would result in similar movement error signals for all target
24 objects—in this case, the mechanism described by Equation 5 would produce proportional
25 decreases in the influence of each cue, which are indistinguishable from the intercept shifts of
26 Equation 4.
27   As an alternative hypothesis, it is also possible that the sensorimotor system compares
28 metric depth estimates derived from each cue with a similar metric estimate derived from haptic
29 feedback when the object is stably held. By doing so, the system could identify the biased cue
30 and subsequently reduce its influence in the visuomotor mapping. Indeed, cross-modal
31 comparisons have been tentatively suggested as a possible mechanism of adjusting relative cue
32 influences (Ernst, Banks, & Bülthoff, 2000). Under this hypothesis, we would expect to observe
33 cue-selective adaptation when participants grasp objects defined by one accurate cue and one
34 biased cue.
35
36 *1.4. Study overview*
37   In a tabletop virtual reality environment (with consistent accommodative and vergence
38 information), participants repeatedly grasped 3D objects (paraboloids) defined by stereo and
39 texture cues. At the end of each grasp, the hand closed down on a real object with a physical
40 depth that was set to match the depth specified by one or both of the cues, depending on the
41 feedback condition. Our results indicate that cue-selective adaptation reliably occurs when a
42 single depth cue suddenly becomes uncorrelated with physical reality (and is therefore positively
43 correlated with movement errors). On the other hand, when the depth specified by a particular
44 cue is biased so as to under- or overestimate the physical depth, cue-selective adaptation is
45 absent, but motor outputs are shifted uniformly toward the reinforced cue.
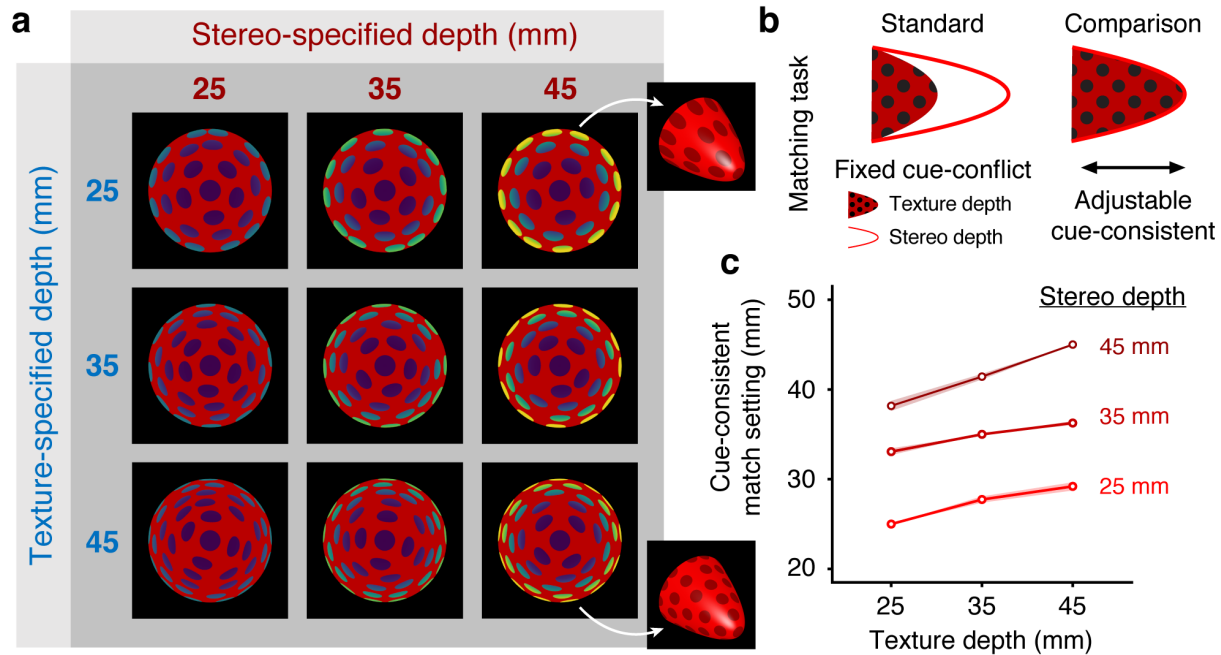
Figure 1. Experiment 1: Stereo-texture paraboloid stimuli and Matching task results. (A) Nine paraboloid objects were rendered by independently manipulating texture and stereo cues to specify depths of 25, 35, or 45 mm (base always subtended 8°). For ease of viewing, stereo depth is coded by a color gradient. The main diagonal of the matrix corresponds to the normally occurring covariation of stereo and texture information (*i.e.*, consistent cues). In this set of stimuli, the effect of variations in stereo information can be assessed independently of the effect of variations in texture via 2D linear regression (Eqn. 1). The off-diagonal objects are generated with cue conflicts. Two oblique views of the final rendered 3D objects are shown on the far right—the dots are circular on the cue-consistent stimulus (bottom-right), while the dots appear stretched on the cue-conflict stimulus (top-right) such that the frontally viewed projection of the texture specifies a shallower stimulus. (B) At the beginning of each session, participants performed a depth adjustment task to obtain a cue-consistent perceived depth match with each of the cue-conflict stimuli (polka dot pattern indicates texture depth, red dotted line indicates stereo depth). In this task, participants alternated freely between viewing one of the conflict objects (the standard, fixed during each trial) and viewing/adjusting the current setting of an adjustable cue-consistent object (the comparison). For each participant, this task identified a set of cue-consistent stimuli that were perceived to have the same depths as the cue-conflicts. In the Grasping task, these cue-consistent stimuli were presented in a Baseline phase to calibrate grasping behavior prior to introducing the cue conflicts, and afterwards in a Washout phase. (C) Average depth setting of the cue-consistent object when adjusted to match the perceived depth of each stereo-texture conflict object. The cue-consistent stimuli are plotted as reference points. Errors ribbons are ±1 SEM.

2. Results

2.1. Experiment 1

Figure 1a depicts a three-by-three matrix of 3D paraboloid objects, where rows of the matrix correspond to different values of texture-specified depth and columns correspond to the different values of stereo-specified depth. Along the main diagonal, we obtain three *cue-consistent* stimuli, where the two cues are rendered based on the same physical depth value. The six off-diagonal stimuli are *cue conflicts*: texture depth is greater than stereo depth in the lower-left region, whereas stereo depth is greater than texture depth in the upper-right. In Experiment 1,
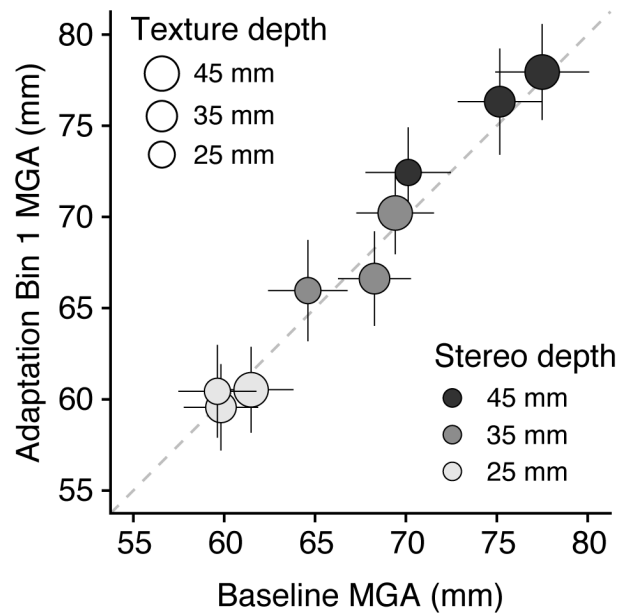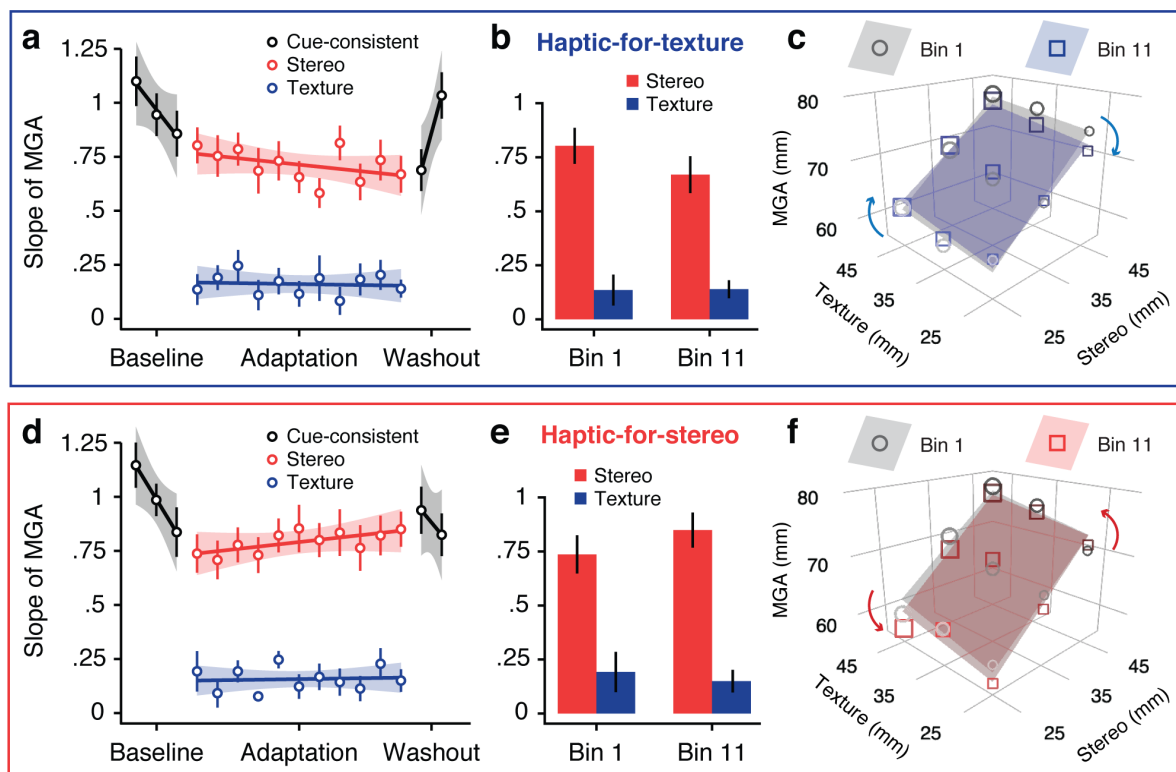
Figure 2. Comparison of maximum grip apertures (MGAs) during Baseline and in the first bin of the Adaptation phase. Across this transition, the component stereo and texture depths of each stimulus changed from consistent to conflicting, but the perceived depth of each object remained the same due to the Matching procedure (see Fig. 1 for average cue-consistent depths). The strong correlation between the MGAs supports the idea that the visuomotor system relies on the same analysis of depth as the perceptual Matching task.

1   we recruited 25 participants to repeatedly grasp these nine stimuli along their depth dimension in
2   two conditions: a haptic-for-texture condition and a haptic-for-stereo condition, where the depth
3   specified by the indicated cue always matched the physical object encountered at the end of the
4   grasp. Consequently, the other cue was uncorrelated with physical depth.
5   To obtain a set of cue-consistent stimuli that could be used to calibrate grasping behavior
6   before introducing the cue conflicts, we asked participants to perform a Matching task at the start
7   of each session (Figure 1b). For each of the six cue-conflicts from the test set (note that Matching
8   was not necessary for the three cue-consistent stimuli), they adjusted the depth of a cue-
9   consistent paraboloid until the two appeared to have the same depth. Each participant grasped
10   objects from their personalized set of perceptually matched cue-consistent stimuli in a Baseline
11   phase before the test stimuli were introduced, and again afterwards in a Washout phase. The
12   average depth settings from the Matching task are shown in Figure 1c; these settings correspond
13   to an average relative weight on stereo information w_S of 0.75 (SEM = 0.019) according to $w_S =$
14   $(z_{match} - z_T)/(z_S - z_T)$. Notice that here we have used cue weights that sum to one, instead of
15   freely varying coefficients as in our grasp planning model. This is because our psychophysical
16   procedure relied on a comparison with a fixed standard, and thus cannot indicate the exact metric
17   depth that was perceived—an independent metric probe would be required to do this. The
18   Matching procedure only allows a measurement of the relative influences of the two cues in
19   perception.
20   We adopted the Matching procedure primarily because it allowed us to obtain a precise
21   perceptual match for each of the target objects in a relatively small number of trials. Having
22   obtained these, we were then able to test in the Grasping task whether the perceived depth
23   matches were treated as such by the visuomotor system, or if a switch from cue-consistent to
24   cue-conflict stimuli would cause an immediate change in grasp performance. Figure 2 plots the
25   MGAs for the Baseline phase (cue-consistent stimuli) against those for the first nine trials (*i.e.*,
26   first bin) of the Adaptation phase (cue-conflicts). The fact that the values are nearly identical
27   across the switch suggests that the input to the visuomotor system was, by and large, the same as
28   the perceptual encoding used for the Matching task. The cue-consistent depths presented during
29   Baseline account for 97% of the variability in Baseline grip apertures, compared to 95% of the
30   variability in early Adaptation (adjusted R^2), suggesting that grasp planning was based on highly
31   similar encodings of depth information in both phases.

Figure 3. Cue-selective adaptation in Experiment 1. Top panel: Haptic-for-texture condition. Bottom panel: Haptic-for-stereo condition. (a, d) Slope parameters estimated by linear regression on maximum grip apertures (MGAs) as a function of depth information in each bin. For the cue-consistent stimuli presented during Baseline and Washout, we computed a single slope in a simple linear regression. For the cue-conflicts presented during Adaptation, we computed independent slopes with respect to the rendered stereo and texture depths in a multiple regression (Eqn. 1). To evaluate whether the influence of stereo and texture information changed in response to the haptic feedback within the Adaptation phase, we fit a further linear regression on these estimated slopes as a function of bin number. (b, e) Stereo and texture slopes in the first and last bins of Adaptation. (c, f) Surface plots of the 2D regression results during Baseline (gray) and Adaptation (blue or red), plotted over the average MGA for each of the nine target objects. In the haptic-for-texture condition, MGAs shifted down in regions of the stimulus space where stereo misleadingly specified more depth than texture, and shifted up in regions where stereo misleadingly specified less depth than texture. The haptic-for-stereo condition elicited opposite shifts.

1   Most importantly, cue-selective adaptation was revealed by changes in the slope of the
2   MGA with respect to stereo-specified and texture-specified depth over the course of Adaptation
3   (Fig. 3). A three-way repeated-measures ANOVA (Condition $\times$ Bin $\times$ Cue) revealed a
4   significant main effect of Cue ($F(1, 24) = 60.98$; $p < 0.0001$), representing the stronger influence
5   of stereo information, and a significant three-way interaction ($F(1, 24) = 5.36$; $p = 0.029$). The
6   latter statistic is the critical one with respect to cue-selective adaptation: it reflects our finding
7   that the difference between stereo and texture slopes becomes smaller over time in the haptic-for-
8   texture condition, and larger over time in the haptic-for-stereo condition. A follow-up two-way
9   ANOVA restricted to the texture slopes yielded no significant effects (all $p > 0.5$), whereas
10   restricting the test to the stereo slopes yielded a significant interaction of Condition $\times$ Bin ($F(1,$
11    $24) = 5.14$; $p = 0.033$). This indicates that the three-way interaction of the omnibus test was
12   driven primarily by opposing changes in the stereo slope, with no noticeable changes in the
13   texture slope. This finding was supported by a bin-by-bin analysis of the coefficients, shown in

Figures 3a and 3d. When analyzing the difference in these estimates of stereo slope change per bin across haptic feedback conditions, we found that the rate of change was significantly modulated by condition (one-tailed t-test; $t(24) = 2.20$; $p = 0.019$). Linear regressions on the stereo slopes as a function of bin number estimated an average change of +0.01 per bin in the haptic-for-stereo condition, and an average change of −0.01 per bin in the haptic-for-texture condition.

Our bin-by-bin slope analysis also indicates a small aftereffect on the slope of the MGA with respect to the cue-consistent depths presented in Baseline and Washout, where the changes in the stereo coefficient that occurred during Adaptation appear to have carried over into Washout, also reducing or increasing the slope of the MGA across the cue-consistent stimuli. In the haptic-for-texture condition, the first bin of Washout showed a smaller slope than Baseline, whereas in the haptic-for-stereo condition, the first bin of Washout showed a slope similar to Baseline. We found a trend toward a slope aftereffect, measured as the effect of feedback condition on the change in cue-consistent slope from Baseline to the first bin of Washout ($t(24) = 1.45$; $p = 0.079$). Additionally, the slopes in the first bin of Washout differed significantly across condition ($t(24) = 2.00$; $p = 0.028$). Since aftereffects are typically considered a hallmark of sensorimotor adaptation, this trend provides converging evidence for our main claim of an adaptive change in the relative influence of stereo and texture information within the visuomotor mapping.

2.2. Experiment 2

As discussed in the Introduction, cue-selective adaptation can be modeled as a change in the slope parameters of the visuomotor mapping, whereas standard adaptation paradigms have focused on shifts of the intercept. Intercept adjustments are sufficient for eliminating movement errors that arise when depth from one or more cues is over- or underestimated. In contrast, slope adjustments are necessary only when a cue changes the strength of its correlation with physical reality, as this results in movement errors that tend to be positive for spuriously large values of the noisy cue and negative for spuriously low values. Faced with conflicting error signals across the domain of visual inputs, uniform shifts of the intercept would oscillate unhelpfully, but slope adjustments can produce increases in some regions of the visual input space and decreases in other regions, as seen in Experiment 1 (Figs. 3c, 3f).

In Experiment 2, we tested a key prediction of the error-based learning mechanism of cue-selective adaptation (Eqn. 5) against an alternative mechanism that involves checking the metric depth estimates derived from each cue against a haptic "ground truth" signal. If an error-based mechanism is responsible for cue-selective adaptation, then it should be observed only during exposure to the stimulus set of Experiment 1 (Fig. 1a; call this an *uncorrelated* set), and not during exposure to a stimulus set where the faulty cue always specifies less depth than the reinforced cue (or more depth; call these *biased* sets). In contrast, if cross-modal metric comparisons allow the system to increase the relative influence of the reinforced cue, then we should observe cue-selective adaptation during exposure to either set.

Biased stimulus sets were comprised of six cue-conflict stimuli where texture depth and stereo depth differed by 10 mm across all objects. The shallower cue always ranged from 20 to 45 mm, the deeper from 30 to 55 mm. We ran two groups of participants in this experiment. For one group ("Adapt+"), the reinforced cue was the deeper of the two cues; for the other group, ("Adapt−"), the reinforced cue was the shallower of the two cues. Each participant performed a haptic-for-texture condition and a haptic-for-stereo condition in separate sessions at least three hours apart. Participants began each session by creating perceptual matches between cue-consistent paraboloids and the six conflict stimuli in the biased set. They then performed

grasping movements through five phases: (1) Baseline grasping of the six perceptually matched cue-consistent stimuli; (2) Pre-test grasping of the uncorrelated set from Experiment 1 to estimate cue slopes prior to exposure, with haptic feedback matching the reinforced cue to maintain consistency with Experiment 1; (3) Adaptation grasping of the relevant biased set (10 bins for the Adapt+ group; 5 bins for Adapt−); (4) Post-test grasping of the uncorrelated set from Experiment 1 to estimate cue slopes after exposure, with haptic feedback still remaining consistent with the reinforced cue; and (5) Washout grasping of the perceptually matched cue-consistent stimuli.

Figure 4 depicts the main results of the experiment, with one panel for each feedback condition (haptic-for-texture, haptic-for-stereo) of each group (Adapt+, Adapt−). In the middle of each panel, we present the Baseline-centered average MGAs for each bin of the Adaptation phase (right-hand y-axis, open circles). The dashed red and blue lines spanning the Adaptation phase represent the rendered stereo and texture depths in the biased sets, with the constant 10-mm cue-conflict; one of these cues was consistent with haptic feedback. In Figure 4, the positions of the dashed lines with respect to the average Baseline MGA (zero) reflect the changes in texture and stereo depth from Baseline to Adaptation. Notice that the dashed red line is slightly closer to zero; this is because cue-consistent depths were set closer to the stereo depths than to the texture depths of the cue-conflicts during perceptual matching, consistent with the stronger influence of stereo information on perceived depth.

During the Adaptation phase, MGAs increased (Adapt+: $t(21) = 4.18$, $p = 0.00021$) or decreased (Adapt−: $t(17) = 3.21$, $p = 0.0026$) from their Baseline values in order to target the reinforced cue. We were surprised, however, to find that the time course of these data did not reflect the exponential learning curve that is characteristic of adaptation to a constant bias. Even in the very first bin of Adaptation (six trials), grasp planning had already compensated for most or all of the change in the haptic feedback. Originally, we expected to observe a more gradual shifting of the MGAs, as participants in previous grasp adaptation experiments required approximately ten trials to fully adapt in response to similar perturbations (Cesanek & Domini, 2017; Cesanek *et al.*, 2019). It is likely that the inclusion of the Pre-test phase between Baseline and Adaptation disrupted the typical time course. In any case, the key result of the Adaptation phase is that MGAs were significantly altered from Baseline, appearing to specifically target the reinforced cue by the end of the phase.

In the Pre-test and Post-test phases, we measured the influences of stereo and texture information during 18 grasps toward the uncorrelated set. Even during these Test trials, haptic feedback remained consistent with the reinforced cue. As in Experiment 1, we estimated a slope parameter for each cue using multiple linear regression (left-hand y-axis, bar graphs). We then performed a mixed-design ANOVA on these slopes with a single between-subjects factor (Group: Adapt+ or Adapt−) and three within-subjects factors (Condition: haptic-for-stereo or haptic-for-texture; Test Phase: Pre-test or Post-test; Cue: stereo or texture). This analysis revealed a highly significant main effect of Cue ($F(1, 38) = 106.72$, $p < 0.001$), as well as a three-way interaction of Condition $\times$ Test Phase $\times$ Cue ($F(1, 38) = 9.41$, $p = 0.0040$).

At first glance, these results appear to suggest that, contrary to our predictions, cue-selective adaptation did in fact take place during exposure to biased set during the Adaptation phase. However, this conclusion overlooks the possibility that these tests captured a gradual accumulation of cue-selective adaptation in the Pre-test and the Post-test. Recall that within each Test phase, participants performed two bins of nine grasps toward the stimuli of the uncorrelated set, with haptic feedback continuing to reinforce the reliable cue in each condition. The exposure during these phases was therefore identical to that in the Adaptation phase of Experiment 1, although considerably briefer. Accordingly, we also evaluated the possibility that cue-selective
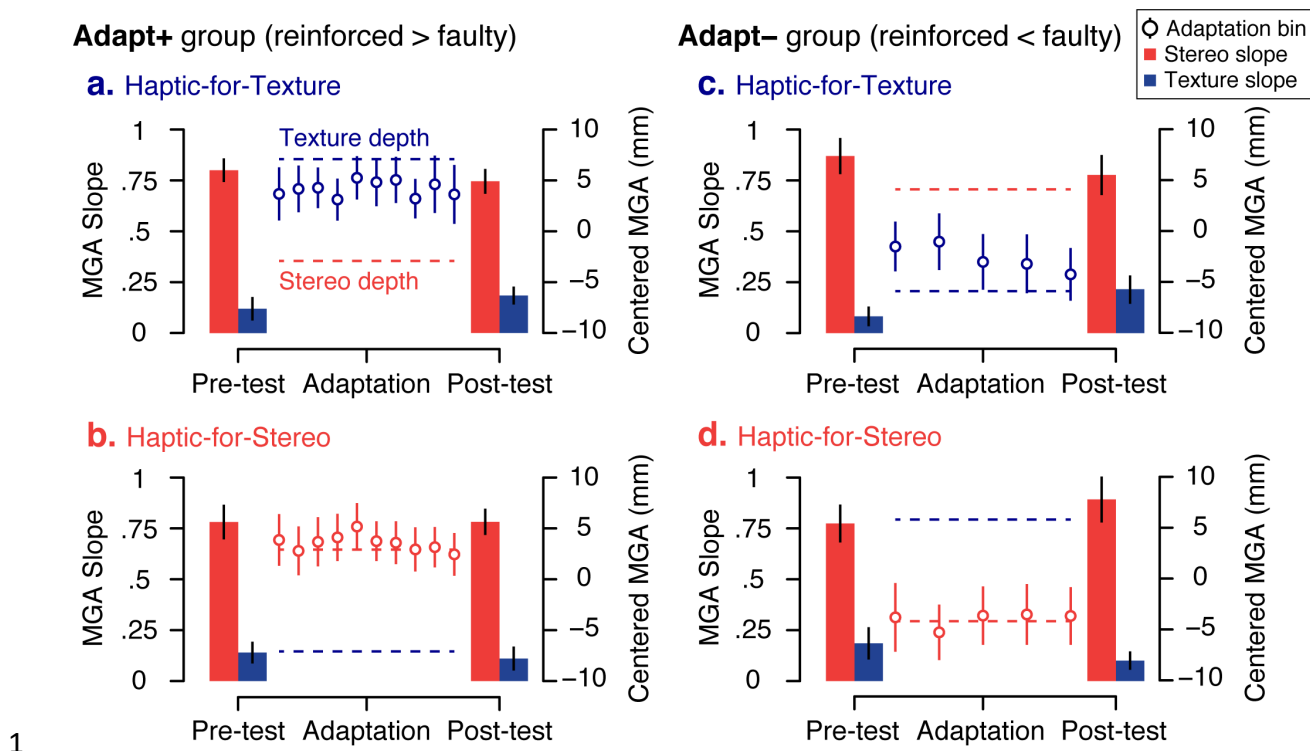
Figure 4. Experiment 2 results. Each panel shows the results of one group-condition pairing. Panels a and b depict the two conditions of the Adapt+ group, while panels c and d show the conditions of the Adapt− group. In the central Adaptation phase, participants grasped objects with a constant cue-conflict: 10-mm separation between texture depth and stereo depth (blue and red dashed lines, respectively). For each bin (6 trials) of the Adaptation phase, we depict the Baseline-centered average maximum grip apertures (MGAs; right y-axis); the symbol color corresponds to the haptically reinforced cue. The length of the Adaptation phase for the Adapt− group was shortened by half based on the rapid adaptation observed for the Adapt+ group. Flanking the main Adaptation phase, the bar graphs indicate the slope of the MGA with respect to stereo and texture information (left y-axis) during the Pre-test and Post-test phases, where we presented the uncorrelated set of stimuli (matrix of Fig. 1).

2 adaptation occurred *within* the Pre-test and the Post-test, during exposure to the uncorrelated set,
3 and not *across* the central Adaptation phase during exposure the biased sets.
4   To obtain a finer temporal resolution, we fit multiple linear regressions to measure the
5 influences of stereo and texture information in each of the two bins of Pre-test and Post-test, so
6 we could compare slope changes that occurred *within* the Test phases, from Bin 1 to Bin 2, with
7 those that occurred *across* the Adaptation phase, from Bin 2 of Pre-test to Bin 1 of Post-test (Fig.
8 5). This is appropriate because Bin 2 of the Pre-test provides the most up-to-date measure of cue
9 influences on grasping prior to any exposure to the biased set. Recall that the error-based update
10 rule in our model predicts no cue-selective adaptation during exposure to the biased set, so the
11 relative influence of the reinforced cue should not be enhanced across the Adaptation phase. On
12 the other hand, we might expect some cue-selective adaptation within the Test phases.
13   First, we used a mixed-design ANOVA as an omnibus test of the slope-change data
14 displayed in Figure 5, with one between-subjects factor, Group (columns of Fig. 5), and two
15 within-subjects factors, Condition (rows of Fig. 5) and Order (x-axis of Fig. 5; changes occurring
16 *within* the Test phases versus those occurring *across* the Adaptation phase). Since cue-selective
17 adaptation is marked by opposing changes in the slopes of the two cues, we have simply taken
18 the difference between the slope changes, change in reinforced minus change in faulty, as our
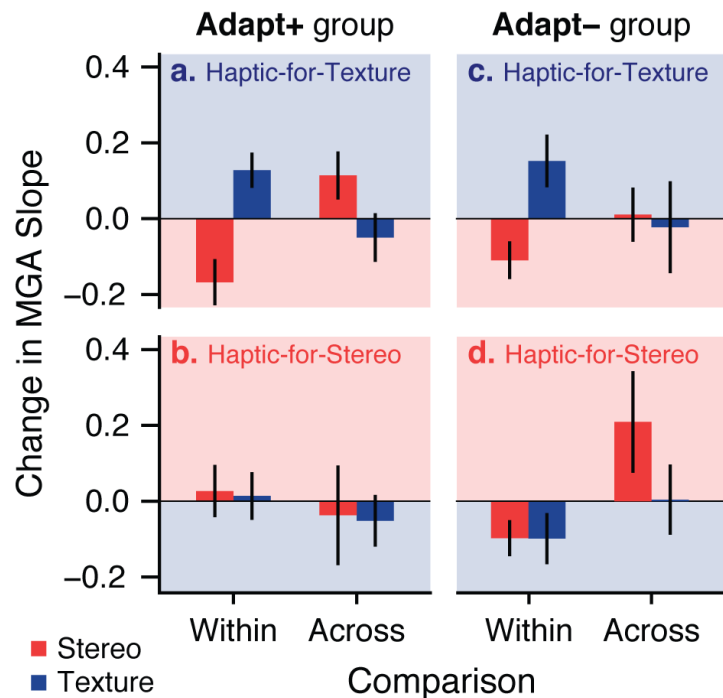
Figure 5. Changes in slope parameters observed *within* the Test phases, as a result of exposure to the uncorrelated set, versus those observed *across* the Adaptation phase, as a result of exposure to the biased sets. The shading of the background indicates the expected direction of slope change for each cue, if cue-selective adaptation took place. For example, in a haptic-for-texture condition, cue-selective adaptation would be marked by an increase in the slope of the MGA with respect to texture information (blue) and/or a decrease in the slope with respect to stereo (red).

1   dependent variable. This analysis revealed a significant interaction of Condition $\times$ Order (F(1,
2   38) = 6.39, p = 0.016), indicating that the within-versus-across difference that we were interested
3   in varied as a function of the feedback condition. Accordingly, we followed up with two specific
4   paired *t*-tests, one for each condition. In the haptic-for-texture condition (Figs. 5a & 5c), we
5   found that cue-selective adaptation was significantly greater within the Test phases than across
6   the Adaptation phase (t(39) = 2.92, p = 0.0058). No such difference was found in the haptic-for-
7   stereo condition (p = 0.47)—Figures 5b and 5d reveal mostly negligible slope changes in this
8   condition. An apparent exception can be spotted in Figure 5d (Adapt−, haptic-for-stereo), where
9   it appears that the strength of stereo information did increase considerably across the Adaptation
10   phase. However, on closer inspection we found the Pre-test of this condition to be somewhat
11   anomalous, with unusually low stereo and texture slopes in Bin 2 of Pre-test (0.67 and 0.12,
12   compared to 0.89 and 0.27 in the preceding bin). The low slopes in this bin were accompanied by
13   a very large intercept parameter (39.8 mm, compared to 27.4 mm in the preceding bin),
14   suggesting that participants had adopted a uniformly larger grip aperture and temporarily reduced
15   their normal reliance on depth information. The correction of the abnormally low stereo slopes in
16   Bin 1 of Post-test should not necessarily be taken as evidence of cue-selective adaptation—in
17   fact, a post-hoc test of this subset of the data reveals that the observed increase in stereo slope
18   across Adaptation was not significant (p = 0.075).
19   Overall, by breaking down our Pre-test and Post-test phases into their constituent bins,
20   we find evidence that the overall cue-selective adaptation effect from Pre-test to Post-test
21   (reported above, see Fig. 4) actually resulted from cumulative exposure to the uncorrelated set in
22   the two Test phases. These data show that the constant perceptual bias introduced during the
23   Adaptation phase was handled by simply increasing or decreasing the grip aperture, rather than
24   changing the relative influences of the cues. Yet during this phase, participants had plenty of
25   exposure to a systematic mismatch between physical depth, felt via haptic feedback, and the
26   depth specified by the faulty visual cue. If a metric, cross-modal comparison of the haptic
27   "ground truth" with each visual cue is capable of reducing the relative influence of the faulty cue
28   in grasp planning, we should observe this as a shift in the regression coefficients from the second

1   bin of Pre-test to the first bin of Post-test, which immediately followed Adaptation. In contrast,
2   we see that exposure to the biased sets produced no cue-selective adaptation. Ultimately, we
3   ended up with additional evidence that exposure to the uncorrelated set, where participants
4   experience conflicting movement errors that are correlated with the faulty cue, is necessary to
5   trigger cue-selective adaptation. These results are consistent with the predictions of our error-
6   based mechanism, and inconsistent with the notion that haptic feedback causes cue-selective
7   adaptation by identifying the inaccurate cue.
8

9 **3. Discussion**
10   Both of the reported experiments reliably induced cue-selective adaptation for 3D targets
11   defined by stereo and texture information, as measured by changes in the slope of the MGA with
12   respect to each cue. This occurred only when the correlation between a given cue and physical
13   reality was altered, and not in response to simple over- or underestimation of depth from that
14   cue, which instead produced a constant shift of the planned grip aperture. As a complement to
15   canonical adaptation, cue-selective adaptation is a powerful mechanism to keep movements
16   accurate when interacting with real 3D objects under variable viewing conditions.
17   With respect to the neurophysiology of the visuomotor system, perhaps the most notable
18   implication of our findings is that individual depth cues must remain separable in the visual
19   encoding used for motor planning. Cue-selective adaptation would not be possible if available
20   cues were already combined into a single estimate, because movement error signals could not be
21   related back to the input signals received from each cue (Eqn. 5). Thus, we can conclude that
22   texture and stereo signals are separably encoded in the visuomotor mapping.
23   In apparent contrast to the separable encoding needed to support cue-selective adaptation,
24   a seminal paper from Hillis, Ernst, Banks, and Landy (2002) concluded that "mandatory fusion"
25   of stereo and texture cues occurs in conscious 3D perception. Specifically, their study showed
26   that individual cues to slant cannot be isolated to aid perceptual discriminations between cue-
27   consistent and cue-conflict stimuli. In the remainder of the Discussion, we develop an account of
28   cue-selective adaptation that reconciles the phenomenon of cue-selective adaptation with the idea
29   of mandatory fusion of depth cues in perception. To do so, we will speculate on the relationship
30   between the visual encoding affected by cue-selective adaptation and how perceived 3D shape
31   may emerge from this encoding.
32

33 *3.1. Distributed encoding of depth signals*
34   In the Introduction, we posed the visuomotor mapping as a global function with specific
35   parameters (cue-specific slopes and an intercept) and assumed that these parameters are modified
36   by sensorimotor adaptation. However, some previous research suggests it is more appropriate to
37   model the effects of adaptation using a distributed representation of the visuomotor mapping,
38   where a population of locally-tuned units ("neurons") encode relevant features of the desired
39   response (*e.g.*, target location, movement velocity; Thoroughman & Shadmehr, 2000; Donchin,
40   Francis, & Shadmehr, 2003; Hwang, Donchin, Smith, & Shadmehr, 2003). In the case of
41   grasping 3D objects defined by multiple depth cues, we are concerned with visual features,
42   namely stereo and texture information. When preparing a grasp, we propose that input signals
43   related to these features establish a pattern of activity across the population of stereo- and
44   texture-sensitive units. For instance, some units might respond maximally when a strong stereo
45   signal is paired with a weak texture signal, while others respond maximally to strong signals
46   from both stereo and texture. With this type of tuning, the units could encode the entire space of
47   possible depth signals, allowing them to produce a response for nearly any cue-specific distortion
48   that might arise. Together, these units and their associated weights specify a particular output at

1   every location in the depth signal space, producing a smooth function akin to, but less
2   constrained than the linear surface depicted in Figure 2.
3   When planning to grasp a paraboloid object with a large depth-from-stereo and a small
4   depth-from-texture, nodes that are sensitive to this particular combination of visual information
5   may respond preferentially, while nodes sensitive to other combinations are relatively silent. This
6   pattern of activity is then transformed into a planned grip aperture by (1) scaling the input
7   activity of each node by a vector of weight parameters, which determine the influence of each
8   node on the motor output, and (2) pooling the weighted node activities. This approach to
9   computing arbitrary functions is called a local function approximation because the output is
10   constrained to vary smoothly with changes in the input dimensions (stereo and texture
11   information), but unlike Equation 1 there are no explicit global parameters. Instead, the
12   parameters are the node weights, and adjusting the weight of a node based on error feedback will
13   only raise or lower the motor output in a localized region of the input domain. As a result, this
14   format provides a certain amount of selectivity along the feature dimensions encoded by the
15   individual units, that is, it provides the separability of stereo and texture information that is
16   required for cue-selective adaptation.
17   On this view, cue-selective adaptation is a straightforward result of the same error-
18   correction mechanisms that underlie canonical sensorimotor adaptation, parsimoniously
19   eliminating the *ad hoc* distinction between one form of adaptation that adjusts the intercept and
20   another that adjusts the slope of the visuomotor mapping. In this model, cue-selective adaptation
21   is achieved entirely by incremental, local adjustments of the approximated function. Since the
22   units are selective for specific combinations of depth cue signals, error signals will most strongly
23   affect motor outputs in the region of signal space that encodes the target object, with the
24   corrective effect falling off in more distant regions. As these corrections accumulate, the
25   resulting changes in planned grip apertures can be captured in the parameters estimated by
26   multiple linear regression. Intercept shifts occur when node weights all change in the same
27   direction, as when experiencing constant errors regardless of the available depth information. On
28   the other hand, slope changes occur when node weights change in opposite directions in different
29   regions of the input space, as when large values of an unreliable depth cue cause positive errors
30   and small values cause negative errors.
31
32   *3.2. Cue-selective adaptation and 3D perception*
33   In this study, we did not test whether the observed cue-selective adaptation is actually the
34   result of a deeper change in 3D visual perception. However, a previous study by Ernst *et al*.
35   (2000) suggests that the same kind of visuomotor training does produce changes in perception of
36   3D slant. In that study, participants touched a variety of slanted surfaces in which either stereo or
37   texture information was consistent with the physical surface slant; the other (unreliable) cue
38   varied dramatically around the physical slant ($\pm30°$). After exposure, participants' slant
39   judgments were less influenced by the unreliable cue. The authors, seeking to justify their
40   "small" observed changes in perception, point out that their unreliable cue remained somewhat
41   correlated with physical slant ($r = 0.59$). However, this may be a misleading view of the
42   experimental conditions, as their stimuli would have actually produced a strong correlation of the
43   unreliable cue with movement error signals, which we have suggested as the driver of cue-
44   selective adaptation according to the learning rule of Equation 5 (note that this update rule is
45   perfectly suited for tuning node weights in the distributed representation discussed above). Given
46   the highly similar patterns of error feedback in the present study and the study of Ernst *et al*.
47   (2000), an intriguing possibility is that the process of cue-selective adaptation actually produces
48   changes in visual perception, and not just changes in the motor output.

1   Though this is a speculative proposal, consistency with empirical results indicates that it
2 is not unreasonable to believe that perceived depth is shaped by the process of tuning weights on
3 a population of visual units, as described in the previous section. Specifically, we proposed that
4 the units in this population are responsive to particular *combinations* of signals from earlier
5 stereo processing and texture processing. To explain the results of Ernst *et al*. (2000), we propose
6 that the adjustment of weights on these units does not merely modify the motor output, like the
7 planned grip aperture, but actually determines the resulting perceptual interpretation of the
8 present assortment of depth information. Intriguingly, this view also fits with the mandatory
9 fusion of depth cues in perceptual judgments, as demonstrated by Hillis *et al*. (2002). In creating
10 a particular pattern of activity across the full population of units, available depth cues have
11 already been combined—although the pattern of activity does indicate a specific combination of
12 cue signals, which provides separability, the component signals can no longer be individually
13 accessed. Therefore, if perceived 3D shape is determined by the learned "weights" on the
14 presently activated units, which connect them to downstream systems for action planning, then
15 when two distinct combinations of depth signals activate units with similar weights, the resulting
16 percepts will not be discriminable, even if the individual cues might be discriminable when
17 presented in isolation.
18
19 *3.3. The teaching signal: Haptic information versus movement errors*
20   As a final note, it is also worth pointing out that the earlier connection with the study by
21 Ernst *et al*. (2000) helps to illuminate a subtle but important difference between movement errors
22 and haptic estimates of 3D shape as drivers of cue-selective adaptation. Haptic estimates of
23 physical shape are often suggested as a ground truth against which to calibrate visual processing
24 of depth information, following Berkeley's maxim that "touch teaches vision". The gist of the
25 claim is that the relative reliabilities of depth cues can be estimated by tracking correlations
26 between haptic estimates of depth and visual estimates of depth obtained independently from
27 each cue. In practice it would be extremely difficult to robustly estimate the relative reliabilities
28 of two noisy signals, stereo and texture, based on their correlations with a third, even noisier
29 signal, haptics. However, many authors have assumed that relative reliabilities can be accurately
30 estimated by comparison with haptic information and subsequently leveraged to combine the
31 individual depth estimates from each cue in statistically optimal fashion (Ernst & Banks, 2002;
32 Hillis, Watt, Landy, & Banks, 2004). Perhaps a bigger issue with this mechanism for cue
33 combination is that it depends on yet another questionable assumption, that individual cues
34 provide unbiased metric estimates of physical depth. There is a large body of psychophysical
35 evidence that directly conflicts with this assumption (Todd, 2004).
36   In contrast, the learning rule we have suggested for adjusting the interpretation of
37 different combinations of depth cue signals (Eqn. 5) could be driven by the same error signals
38 that have been implicated in sensorimotor adaptation and does not require any controversial
39 assumptions about the availability of unbiased estimates. Rather, it represents a way in which
40 naturally occurring fluctuations in the reliability of individual depth cues could be handled by the
41 sensorimotor system, through iterative online adjustments that gradually reduce the influence of
42 signals that are positively correlated with error. The end result is an interpretation of depth
43 information from multiple cues, each of which may be expressed in arbitrary units, that reflects
44 the distinctions that are most important for producing accurate actions. This may be viewed as a
45 less intelligent mechanism of cue combination, as it does away with the idea that the visual
46 system recovers accurate depth estimates from each individual cue. However, this intelligence,
47 inherited from the naïve realist view that perception is firmly moored in the metrics of physical

1 reality, is replaced by a powerful short-term flexibility that allows perception to effectively drive
2 action across changing environmental conditions.
3
4 **4. Methods**
5
6 *4.1. Participants*
7 Sixty-five participants were recruited for Experiments 1 (N = 25) and 2 (N=40; 22 in
8 Adapt+, 18 in Adapt−). Participants were between 18 and 35 years old and right-handed, with
9 normal or corrected-to-normal vision. They were either granted course credit or paid $8/hour as
10 compensation. Informed consent was obtained from all participants prior to any participation, in
11 accordance with protocol approved by the Brown University Institutional Review Board and
12 performed in accordance with the ethical standards set forth in the Declaration of Helsinki.
13
14 *4.2. Apparatus*
15 Figure S1 presents a few photographs of the lab setup. Participants were seated in a
16 height-adjustable chair so that the chin rested comfortably in a chinrest. Movements of the right
17 hand were tracked using an Optotrak Certus motion-capture system. Small, lightweight posts
18 containing three infrared emitting diodes were attached to the fingernails of the index finger and
19 thumb, and the system was calibrated prior to the experiment to track the extreme tips of the
20 distal phalanxes of each finger. This motion-capture system was coupled to a tabletop virtual
21 reality environment: participants looked into a half-silvered mirror slanted at 45° relative to the
22 sagittal body midline, which reflected the image displayed on a 19" CRT monitor placed directly
23 to the left of the mirror at the correct distance to provide consistent accommodative and vergence
24 information.
25 Participants viewed stereoscopic renderings of 3D paraboloid objects in this tabletop
26 virtual reality setup, where stereo and texture information were controlled independently using a
27 3D graphics technique known as backprojection. The paraboloids were rendered with their tips at
28 a viewing distance of 40 cm at eye level. This arrangement made the rendered 3D objects appear
29 to be floating in space beyond the mirror. The bases of the paraboloids always subtended 6.5° of
30 visual angle. Stereoscopic presentation was achieved with a frame interlacing technique in
31 conjunction with liquid-crystal goggles synchronized to the frame rate. Stereoscopic visual
32 feedback of the thumb only was provided throughout the experiment, to help participants keep
33 track of their hand position. We presented the thumb only in order to prevent visual comparison
34 of the stereo-rendered grip aperture with the stereo-specified depth of the object, which might
35 have unintentionally reinforced stereo information in our haptic-for-texture conditions.
36 To provide haptic feedback, a custom-built motorized apparatus was placed in the
37 workspace. This apparatus consisted of a stepper motor with its shaft extended by a long screw.
38 On the end of this screw, we attached a round metal nub to simulate the rounded tip of the
39 paraboloid objects—perfect alignment between the physical and rendered paraboloid tips was
40 established during the calibration phase at the start of each session. To simulate the flat, round
41 rear end of the paraboloids, we threaded a metal washer (approximately 6 cm in diameter, equal
42 to the average base diameter of the rendered objects) onto the screw. As the stepper motor spun,
43 the washer would travel back and forth along the length of the screw, anchored by a heavy
44 wooden piece on the left side to ensure that one full rotation of the stepper motor would move
45 the washer by one thread pitch (*i.e.*, the distance between threads). On every trial, the resulting
46 depth of the physical object was double-checked using additional Optotrak markers mounted on
47 the physical apparatus and corrected if necessary.
48
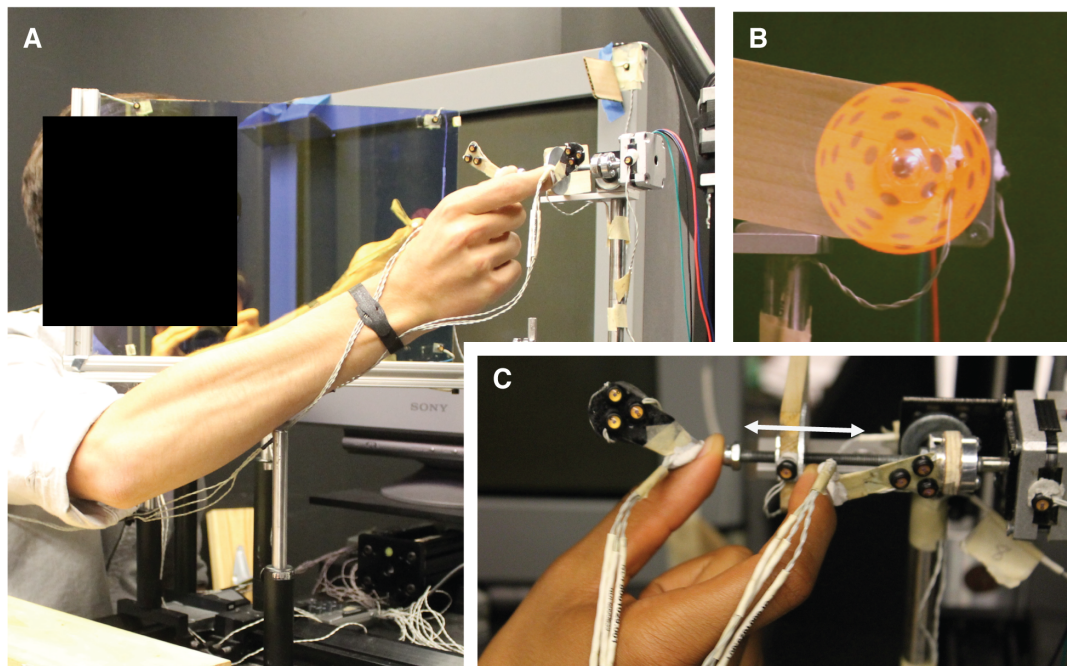
Figure S1. Photographs of the tabletop virtual reality setup. (A) The observer (face hidden for pre-print) looks into a slanted mirror while wearing stereoscopic glasses, seeing a compelling 3D object on the far side of the mirror, aligned with an automated physical apparatus in the workspace that provides haptic feedback of objects with different depths. During the experiment, the room is completely dark and a back panel is placed on the mirror to prevent any visual stimulation other than the rendered 3D object. The participant reaches with the right hand to grasp the rendered object so the thumb lands on the tip and the index finger lands on the base. Optotrak infrared-emitting diodes attached to the fingernails provide precise location information about the fingertips, allowing us to compute the in-flight grip aperture. (B) Frontal view of a rendered paraboloid. This is a cyclopean view, rather than stereoscopic, for visualization purposes. The tip of the paraboloid is perfectly centered on a small rounded nub to provide haptic feedback of the tip. (C) Side view of the physical apparatus for providing haptic feedback. A stepper motor spins a screw in order to slide a large round washer back and forth along the screw. This allowed us to create a physical object of any depth on each trial. The thumb landed on the rounded nub aligned with the tip of the paraboloid, while the index finger pinched down on the rear surface, which could be aligned with either the stereo- or texture-specified depth when grasping a cue-conflict paraboloid.

4.3. Procedure

Both experiments began with a Matching task, where participants matched through a psychometric procedure the perceived depth of the cue-conflict paraboloids with cue-consistent objects, where stereo and texture specified the same depth. In Experiment 1, the target stereo-texture conflict stimuli were the six off-diagonal objects in our uncorrelated set (see Fig. 1 for more details); the three cue-consistent objects of the uncorrelated set did not require matching because they were already composed of consistent cues. In Experiment 2, the target stimuli were six cue-conflict objects where stereo depth and texture depth differed by a constant conflict of 10 mm (see Section 2.2 for more details). On each Matching trial, we allowed participants to switch freely between the fixed cue-conflict stimulus and an adjustable cue-consistent stimulus, using keypresses to make incremental changes to the depth of the cue-consistent stimulus until it appeared to match the depth of the cue-conflict stimulus. To prevent the use of motion information, we displayed a blank screen with a small fixation dot for an inter-stimulus interval

1   of 750 ms whenever the stimulus was changed. Participants performed two repetitions on each of
2   the six fixed cue-conflicts in each Experiment, for a total of 12 Matching trials.
3   The resulting sets of stimuli (six pairs of matched cue-conflict and cue-consistent
4   paraboloids) were then presented as stimuli in the Baseline, Adaptation, and Washout phases of
5   the Grasping task. During the Grasping task, participants used a precision grip to grasp the
6   paraboloid objects from front to back. Trials were presented in a pseudo-random "binned" trial
7   order, where each of the target objects in a given phase of the experiment was presented once
8   before any one was presented again; as a result, each bin contains one presentation of each target
9   object. Since there were nine target objects in the uncorrelated set, each bin of Experiment 1
10   contained nine trials. Since there were six target objects in the biased sets, each bin of
11   Experiment 2 contained six trials, except in the Pre-test and Post-test phases where we presented
12   the uncorrelated set. On each trial, participants were shown the target object for 500 ms, then
13   heard the "go" signal, and reached to grasp the target as shown in Fig. S1. There was no explicit
14   time limit on these grasps, but the total elapsed time from movement onset to object contact
15   rarely exceeded 1.5 seconds.
16   Following the Matching task, the Grasping procedure of Experiment 1 was as follows. In
17   the Baseline phase, participants grasped their personalized set of nine cue-consistent paraboloids,
18   perceptually matched to the nine objects of the uncorrelated set, for three bins of trials.
19   Participants then proceeded immediately into the Adaptation phase, where the cue-consistent
20   paraboloids were suddenly replaced by the perceptually matched cue-conflict paraboloids, and
21   the depth of the physical object providing haptic feedback matched the depth of whichever cue
22   was being reinforced during that session (haptic-for-stereo or haptic-for-texture). Following
23   eleven bins of exposure to the uncorrelated set, Experiment 1 concluded with a two-bin Washout
24   phase, identical to Baseline.
25   The procedure of Experiment 2 was designed to be similar to Experiment 1, but with
26   presentations of the biased set, with its constant cue-conflict of 10 mm, rather than the
27   uncorrelated set of Experiment 1, with its variable positive and negative cue-conflicts. As in
28   Experiment 1, participants began with a Baseline phase, grasping their personalized set of six
29   cue-consistent paraboloids, which were perceptually matched to the six objects of the biased set,
30   for three bins of trials. Instead of proceeding directly into the Adaptation phase, where they
31   would interact with the cue-conflict objects of the biased set, they first completed a Pre-test
32   phase consisting of two bins of trials where we presented the uncorrelated set. Next, in the
33   Adaptation phase, we presented the six objects of the biased set for 10 bins of trials when testing
34   the Adapt+ group, but only for 5 bins of trials when testing the Adapt− group. We shortened the
35   Adaptation phase for Adapt− because this version of the experiment was run after the Adapt+
36   group, where we already observed rapid convergence on the reinforced cue—longer adaptation
37   periods were clearly not necessary to eliminate movement errors. Following Adaptation,
38   participants completed a two-bin Post-test, identical to the Pre-test, and concluded with a two-bin
39   Washout phase, identical to Baseline.
40
41   *4.4. Analysis*
42   Raw motion-capture position data was processed and analyzed offline using custom
43   software. Missing frames due to marker dropout were linearly interpolated, and the 85-Hz raw
44   data was smoothed with a 20-Hz low-pass filter. The time series data from each trial was
45   cropped by defining the start frame as the final frame where the thumb was more than 25 cm
46   from its contact location on the tip of the object, and the end frame as the first frame where (a)
47   the thumb came within 1 cm of its contact location or (b) the index finger entered into a 3 cm
48   wide by 3 cm high bounding box, extending 10 cm in depth (well beyond the rear edge of the

deepest object). The grip aperture profile was computed for each trial by taking the vector distance between the index finger and thumb locations on each frame. The maximum grip aperture (MGA), a widely used kinematic measure of grasp planning (Jeannerod, 1981), was extracted from this time series.

Two criteria were used for trial exclusion: the proportion of missing frames due to marker dropout exceeded 90% or fewer than 5 frames were not missing. In Experiment 1, neither of these criteria were met for any of the movements, so no trials were excluded from analysis. In Experiment 2, 72 out of a total 9480 trials were excluded by these criteria (~0.7%).

The factorial design of the uncorrelated set of stimuli allowed us to measure the relative influence of stereo and texture information in the Grasping task by estimating coefficients (slopes) for each cue via multiple linear regression according to Equation 1, with the MGA as the response variable. A regression was computed for each bin of nine trials within the Adaptation phase of Experiment 1, and in the Pre-test and Post-test phases of Experiment 2. In the Baseline phase of Experiment 1, we computed the slope of the MGA with respect to the perceptually matched cue-consistent depths using simple linear regression in each bin.

Acknowledgments

References

Cesanek, E. & Domini, F. (2017). Error correction and spatial generalization in human grasp control. *Neuropsychologia*, *106*, 112–122.

Cesanek, E., Taylor, J. A., & Domini, F. (2019). Sensorimotor adaptation compensates for distortions of 3D shape information. *bioRxiv*. doi: 10.1101/540187.

Cheng, S. & Sabes, P. N. (2006). Modeling sensorimotor learning with linear dynamical systems. *Neural Computation*, *18*(4), 760–793.

Donchin, O., Francis, J. T., & Shadmehr, R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control. *Journal of Neuroscience*, *23*(27), 9032–9045.

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

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

Held, R. & Hein, A. V. (1958). Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, *8*(3), 87–90.

von Helmholtz, H. E. F. (1962). Treatise on physiological optics. (J. P. C. Southall, Ed. and Trans.). New York: Dover. (Original work completed in 1867, published in 1909).

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, *298*(5598), 1627–1630.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 967–992.

Hwang, E. J., Donchin, O., Smith, M. A., & Shadmehr, R. (2003). A gain-field encoding of limb position and velocity in the internal model of arm dynamics. *PLoS Biology*, *1*(2), 209–220.

Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In: J. Long and A. Baddeley (Eds.), *Attention and Performance IX* (pp. 153–168). Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*(7-8), 1351–1360.

Krakauer, J. W., Ghez, C., & Ghilardi, M. F. (2005). Adaptation to visuomotor transformations: Consolidation, interference, and forgetting. *Journal of Neuroscience*, *25*(2), 473–478.

Krakauer, J. W., Pine, Z. M., Ghilardi, M. F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, *20*(23), 8916–8924.

Pouget, A. & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, *3*(11s), 1192–1198.

Redding, G. M. & Wallace, B. (1997). Adaptive spatial alignment. Mahwah, NJ: Lawrence Erlbaum Associates.

Rosenholtz, R. & Malik, J. (1997). Surface orientation from texture: isotropy or homogeneity (or both)?. *Vision Research*, *37*(16), 2283–2294.

Säfström, D. & Edin, B. B. (2005). Short-term plasticity of the visuomotor map during grasping movements in humans. *Learning & Memory*, *12*(1), 67–74.

Säfström, D. & Edin, B. B. (2008). Prediction of object contact during grasping. *Experimental Brain Research*, *190*(3), 265–277.

Smeets, J. B. & Brenner, E. (1999). A new view on grasping. *Motor Control*, *3*(3), 237–271.

Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *Journal of Neuroscience*, *34*(8), 3023–3032.

Thoroughman, K. A. & Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, *407*(6805), 742–747.

Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, *8*(3), 115–121.

Welch, R. B. (1969). Adaptation to prism-displaced vision: The importance of target-pointing. *Perception & Psychophysics*, *5*(5), 305–309.

Welch, R. B. (1978). Perceptual modification: Adapting to altered sensory environments. New York, NY: Academic Press, Inc.