Perceptual Co-Reference

Michael Rescorla

Abstract: The perceptual system estimates distal conditions based upon proximal sensory input. It typically exploits information from multiple cues across and within modalities: it estimates shape based upon visual and haptic cues; it estimates depth based upon convergence, binocular disparity, motion parallax, and other visual cues; and so on. Bayesian models illuminate the computations through which the perceptual system combines sensory cues. I review key aspects of these models. Based on my review, I argue that we should posit *co-referring perceptual representations* corresponding to distinct sensory cues. For example, the perceptual system represents a distal size using a representation canonically linked with vision and a distinct representation canonically linked with touch. Distinct co-referring perceptual representations represent the same denotation, but they do so under different *modes of presentation*. Bayesian cue combination models demonstrate that psychological explanation of perception should attend to mode of presentation and not simply to denotation.

§1. Sensory cue combination

A familiar picture holds that mental activity involves computation over mental representations (Fodor, 1975, 2008; Gallistel and King, 2009; Pylyshyn, 1984). This paper will discuss representations that figure in computations executed by the perceptual system.

My focus is *sensory cue combination*. The perceptual system estimates distal conditions (e.g. shapes, sizes, colors, and locations of distal objects) based upon proximal sensory input

(e.g. retinal stimulations, or vibrations in the inner ear, or muscle spindle activity). When estimating distal conditions, the perceptual system typically exploits relevant information from multiple sensory modalities. If you see a barking dog across the park, then you receive both visual and auditory information relevant to the dog's location. If you look at an apple while holding it in your hand, then you receive both visual and haptic information relevant to the apple's shape and size. Multiple relevant cues can also arise within a single sensory modality. For example, the visual system estimates depth based upon convergence, binocular disparity, monocular linear perspective, motion parallax, occlusion, texture gradients, and other cues. Multiple cues pose a challenge to the perceptual system, because they usually conflict if only to a slight degree. The perceptual system must resolve conflicts, integrating sensory cues into a unified estimate of distal conditions. A striking illustration is *ventriloquism*, which generates a conflict between visual and auditory cues to location. The perceptual system mistakenly resolves the conflict in favor of the visual cue (the puppet's moving lips).[1]

I want to explore what scientific research into sensory cue combination reveals about perceptual representation. My main thesis is that, in many cases, we should posit multiple perceptual representations representing a single distal property. For example, the perceptual system represents a given distal size using a representation canonically linked with vision and a distinct representation canonically linked with touch. Similarly, the perceptual system represents a given depth using a representation canonically linked with binocular disparity, a distinct representation canonically linked with motion parallax, a distinct representation linked with convergence, and so on. In general, the perceptual system represents a single distal property using distinct perceptual representations canonically linked with distinct cues. The

[1] O'Callaghan (2012) gives a helpful philosophical introduction to multimodal aspects of perception, including ventriloquism and other cross-modal illusions.

representations are *co-referring*, in that they represent a single distal property. We can clarify the nature of these co-referring representations by studying their roles within perceptual computation, including canonical links to privileged information sources.

§2 surveys current research into sensory cue combination. §3 defends my main thesis: we should posit co-referring perceptual representations corresponding to distinct sensory cues. §4 compares my position with a similar position espoused centuries ago by Berkeley. §5 elucidates perceptual co-reference by deploying Frege's insights into mental representation. §6 draws morals regarding psychological explanation. §7 explores the conditions under which mental activity instantiates a given perceptual representation.

§2. Bayesian modeling of cue combination

Helmholtz (1867) proposed that perception involves an "unconscious inference" from proximal sensory input to estimates of distal conditions. Recently, perceptual psychologists have developed Helmholtz's viewpoint using Bayesian decision theory (Knill and Richards, 1996). On a Bayesian approach, the perceptual system encodes a *prior probability*, which assigns probabilities $p(h)$ to hypotheses h regarding distal conditions (e.g. hypotheses regarding the size of a perceived object). The perceptual system also encodes a *prior likelihood*, which assigns a conditional probability $p(e \mid h)$ to sensory input e given hypothesis h (e.g. the probability of certain retinal input given that an object has a certain size and depth). Upon receiving input e, the perceptual system computes the *posterior probability* $p(h \mid e)$, where

$$p(h \mid e) = \eta\, p(h)\, p(e \mid h).$$

Here η is a normalizing constant to ensure that probabilities sum to 1. From the posterior, the perceptual system computes a privileged hypothesis $\hat{h}$ (e.g. a privileged estimate of size). In the models we will consider, $\hat{h}$ is the hypothesis that maximizes the posterior.

Researchers have elaborated the Bayesian framework into mathematically precise, well-confirmed models of numerous perceptual phenomena. The models are extremely idealized, yet even so they yield satisfying explanations for many illusions and constancies.[2] For an overview, see (Rescorla, 2015). My goal here is to highlight how the Bayesian paradigm illuminates sensory cue combination.

§2.1 Weighted averaging and beyond

Suppose you must determine the size of an object while holding it in your hand. Your perceptual system forms an estimate based upon both visual and haptic feedback. For simplicity, let us assume a relatively "flat" prior probability $p(s)$ over possible sizes. In this case, the optimal Bayesian estimate is determined almost entirely by the prior likelihoods $p(e_V \mid s)$ and $p(e_H \mid s)$, where e_V is visual input and e_H is haptic input. Holding e_V fixed, we may view $p(e_V \mid s)$ as a function of s. Viewed that way, it is called a *likelihood function*. Let $\hat{s}_V$ be the value of s that maximizes the likelihood function $p(e_V \mid s)$. Define $\hat{s}_H$ analogously. Assume that the prior likelihoods are Gaussians, i.e. normal distributions. Assume also that the perceptual system seeks to maximize the posterior probability $p(s \mid e_V, e_H)$. Given a few additional assumptions (Landy, Banks, and Knill, 2011, p. 9), the posterior $p(s \mid e_V, e_H)$ attains its maximum when s has the value $\hat{s}_{VH}$ given by

[2] One notable idealization: current models typically employ an uncountable hypothesis space. Taken at face value, any such model presupposes mental capacities to represent uncountably many distinct distal properties.

(1) $\hat{s}_{VH} = w_V \hat{s}_V + w_H \hat{s}_H$.

Here w_V and w_H are weights that reflect the reliability of visual and haptic input --- more technically, the variances of the likelihood functions. Thus, $\hat{s}_{VH}$ is a weighted average of the individual size-estimates $\hat{s}_V$ and $\hat{s}_H$. The average is weighted towards the more reliable cue. Visual input regarding shape is usually more reliable than haptic input, so size-estimates are usually weighted towards vision.[3]

To test the weighted average model, Ernst and Banks (2002) instructed subjects to estimate the height of a ridge given visual input, haptic input, or both. Using virtual reality technology, Ernst and Banks brought visual and haptic cues into conflict. They also parametrically manipulated the reliability of vision by introducing noise into the visual stimulus. Size-estimates conformed quite closely to (1). When visual input was relatively noise-free, vision dominated the estimate. As visual input grew noisier, haptic input exerted more influence.[4]

Researchers have extended the weighted average model to many other cases of cue combination between and within modalities, including visual and haptic cues to shape (Helbig and Ernst, 2008), visual and auditory cues to location (Alais and Burr, 2004), visual and proprioceptive cues to hand location (van Beers, Wolpert, and Haggard, 2002), disparity and

---

[3] The derivation of (1) from Bayesian first principles presupposes that individual visual and haptic estimates are unbiased and that visual and haptic noise are independent. Ernst (2012) discusses normative and empirical ramifications of lifting these and other idealizing assumptions.

[4] The weighted average model predicts that $\sigma_{VH}{}^2$, the variance of the posterior, is given by $\sigma_{VH}{}^2 = \frac{\sigma_V{}^2 \sigma_H{}^2}{\sigma_V{}^2 + \sigma_H{}^2}$, where $\sigma_V{}^2$ and $\sigma_H{}^2$ are variances of the individual likelihood functions. A corollary is that the posterior will have lower variance than the individual likelihoods. Intuitively: combining separate information channels reduces the uncertainty from each channel. The variance prediction is crucial for testing the weighted average model (Rohde, van Dam, and Ernst, 2016, p. 7). A system that switches between $\hat{s}_V$ and $\hat{s}_H$ in proportion to the weights w_V and w_V will look on average as if it conforms to (1), even though it does not compute estimates in accord with (1). To confirm that a system is really computing the posterior given by the weighted average model, one must confirm both (1) and the predicted posterior variance.

texture cues to slant (Knill and Saunders, 2003), and motion and texture cues to depth (Jacobs, 1999). In each case, experimental data fit the weighted average model. Consider a laboratory version of ventriloquism: a light flashes while an auditory stimulus (such as a click) occurs at a slightly different location. The auditory stimulus is perceived as located where the light flashes. This is because vision is a far more reliable location cue than audition, so that the weighted average assigns almost all weight to the visual cue. Alais and Burr (2004) induced "reverse ventriloquism" by blurring the visual cue. A blurred cue is an unreliable guide to location. Accordingly, perceived location shifted almost entirely towards the auditory stimulus.

The weighted average model is just a first step towards understanding sensory cue combination (Ernst, 2012). One drawback of the model is that it only applies when the perceptual system has decided to integrate distinct sensory signals into a unified perceptual estimate. There are many cases where integration of sensory signals does not occur and would not be advisable. For example, if you see a sleeping dog and hear another dog barking far away, then you should not integrate visual and auditory information to form a unified location-estimate. Körding et al. (2007) handle this sort of case through a Bayesian *causal inference* model that generalizes the weighted average model. According to the causal inference model, the perceptual system estimates whether two sensory signals derive from the same distal source, and it decides on that basis whether (or to what extent) to integrate the signals into a single unified estimate. The causal inference model is merely one example of how researchers have generalized the weighted average model in order to fit a wider range of cases.[5]

[5] Another drawback of the weighted average model is that it predicts a weighted average no matter how large the conflict between cues. The prediction fails when conflicts are quite large. In such cases, the perceptual system does not compute a weighted average but instead relies upon a single favored cue. This phenomenon is called *robustness*. If we generalize the weighted average model by allowing likelihoods to be mixtures of Gaussians rather than simply Gaussians, then we can model robust estimation quite successfully in Bayesian terms (Girshick and Banks, 2009).

§2.2 The coupling prior

When the perceptual system combines sensory signals from different modalities into a single unified estimate, it does not typically discard information gleaned from the individual signals. Instead, it retains unimodal estimates based upon the individual signals. Depending on the requirements of the perceptual task, it can then access either the combined estimate or the unimodal estimates. For example, there is strong evidence that the perceptual system typically maintains distinct visual size-estimates and haptic size-estimates in addition to a unified visual-haptic size-estimate (Hillis et al., 2002). Intuitively: an object can look to have a certain size even while it feels like it has a different size.

Notably, unimodal estimates are influenced by information from other modalities. When computing a visual size-estimate, the perceptual system does not simply ignore haptic input. It does not compute the visual size-estimate that it would have produced absent any haptic feedback. Instead, haptic input biases the visual size-estimate. Even when experimenters instruct subjects to ignore haptic input and produce a purely visual size-estimate, haptic input significantly biases the visual size-estimate. A similar cross-modal effect arises for visual and auditory cues (Roach, Heron, and McGraw, 2006).

Ernst (2006) explains these cross-modal effects by positing a *coupling prior*, which correlates perceptual estimates associated with different cues. The coupling prior for visual-haptic size-estimation has the form $p(s_V, s_H)$, where s_V is a visual size-estimate and s_H is a haptic size-estimate. Intuitively, the coupling prior encodes the probability that a "visual size" and "haptic size" co-occur. The perceptual system uses the coupling prior to compute a posterior $p(s_V, s_H \mid e_V, e_H)$. It selects privileged visual and haptic size-estimates that maximize the posterior. Figure 1 illustrates with three possible coupling priors. In the first column, the coupling prior is

flat. As a result, the two cues are handled independently: the visual size-estimate is $\hat{s}_V$, the value that maximizes the likelihood $p(s_V \mid e_V)$; and the haptic size-estimate is $\hat{s}_H$, the value that maximizes the likelihood $p(s_V \mid e_V)$. In the third column, the coupling prior encodes complete certainty as to the correlation between visual size and haptic size. All probability mass is concentrated on the diagonal line $s_V = s_H$ in the two-dimensional space of visual and haptic size-estimates. Accordingly, I say that the coupling prior is *concentrated*. A concentrated coupling induces *complete fusion*: the visual estimate and haptic estimate must agree. As Ernst (2007, p. 4) puts it:

> If the system knows that the visually measured size and the haptically measured size of an object are perfectly correlated and it knows the mapping between the measurements, then it can infer that they have to be "identical"; hence, it makes no sense to allow for two separate "percepts of size" --- one visual and one haptic. The two measurements of size should be fused to one percept of object size.

The second column is intermediate between flat and concentrated, encoding a fuzzy correlation between visual and haptic size. Visual and haptic size-estimates fall between estimates induced by the flat and the concentrated coupling priors. This is called *partial fusion*.[6]

INSERT FIGURE 1 ABOUT HERE

[6] Each panel in Figure 1 depicts a function over ordered pairs of real numbers. The real numbers specify visual size-estimates and haptic size-estimates. Our choice of real numbers reflects some choice of measurement units. Figure 1 uses units such that a visual estimate of size *s* corresponds to the same real number as a haptic estimate of the same size *s*. Given this choice of measurement units, the diagonal line $s_V = s_H$ models scenarios where the visual estimate agrees with the haptic estimate. We could have chosen different measurement units, so that those scenarios would be modeled by a different diagonal line. What matters for psychological explanation are the underlying size-estimates, not the real numbers through which we specify those estimates. The crucial fact about the concentrated coupling prior from Figure 1 is that all probability mass is concentrated on a correlation that matches each visual size-estimate with a haptic size-estimate of the same distal size. It is not explanatorily important that we have chosen measurement units so that this correlation corresponds to the line $s_V = s_H$. What matters is sameness of estimated distal conditions, not sameness of the real numbers through which we theorists specify distal estimates. For more on Bayesian perceptual modeling and arbitrary measurement units, see (Rescorla, 2015).

In practice, intermodal cues to a single distal variable virtually always induce partial rather than complete fusion.[7] Intramodal cues typically induce complete fusion. For example, the perceptual system does not retain separate estimates of slant based upon disparity and texture signals (Hillis et al., 2002). Instead, it fuses the signals into a single estimate. The coupling prior framework provides a principled basis for explaining the degree to which sensory cues fuse (van Dam, Parise, and Ernst, 2014): degree of fusion depends upon where the coupling prior falls in the continuum from flat to concentrated. In this spirit, we may posit a fuzzy coupling prior over visual and haptic size-estimates, biasing visual and haptic estimation towards the diagonal but not all the way. The result is partial fusion, which matches the experimental data for visual and haptic size-estimation. Similarly, we may posit a concentrated or near-concentrated coupling prior over disparity-based slant-estimates and texture-based slant-estimates, resulting in complete or near-complete fusion for disparity and texture cues to slant.

In a striking application of the coupling prior framework, Ernst (2007) studied two distal variables that are normally uncorrelated: luminance and stiffness. He exposed subjects to deviant stimuli where luminance and stiffness were correlated, e.g. a stiffer object was likely to be brighter. After an hour-long training session with deviant stimuli, the coupling prior over luminance and stiffness changed. It began flat (as in Figure 2's first column), reflecting the fact that luminance and stiffness are normally uncorrelated. By the end of the training session, the coupling prior had become fuzzy (as in Figure 2's second column), inducing partial fusion of luminance and stiffness. For example, stiffer objects looked brighter. Ernst's experiment illustrates how the coupling prior rapidly evolves to reflect environmental statistics. A less artificial illustration along similar lines is the recent demonstration by Adams, Kerrigan, and Graf (2016) that, in typical perceivers, haptic cues influence perceived gloss: objects that feel

[7] An exception: visual and vestibular cues to self-motion completely fuse (Prsa, Gale, and Blanke, 2012).

slippery look shinier. Adams, Kerrigan, and Graf (2016) explain this cross-modal effect by positing a fuzzy coupling prior over friction-estimates and gloss-estimates. Intuitively: the coupling prior treats slippery objects (objects with low friction) as likely to be shinier.

Luminance and stiffness are distinct distal variables. Friction and gloss are distinct distal variables. In contrast, an object's size is the same distal variable whether one estimates it visually or haptically. Nevertheless, the three cases are not as different as they might seem. From the perceptual system's standpoint, it may be quite unclear whether two cues are cues to a single distal variable or distinct distal variables. *You and I* know that an object's "visual size" and its "haptic size" are one and the same. *To the perceptual system*, the identity may not be initially apparent. It is not *a priori* obvious that the distal cause of certain retinal stimulations is the same property as the distal cause of certain haptic stimulations. Ernst puts the point as follows (2007, p. 2): "the felt and seen size of an object are two totally different sensory signals: One is derived from photons on the retina, and the other one is derived from sensors detecting the fingers' position given some force when in contact with the object." The perceptual system must somehow bring "visual size" into alignment with "haptic size." It must determine how visual size-estimates correlate with haptic size-estimates. More generally, the perceptual system must correlate estimates based upon one cue with estimates based upon another cue, even when they are estimates of a single distal variable. Learning the requisite correlation is a non-trivial task.

In this connection, let us consider how the perceptual system develops during childhood. I already mentioned that the adult perceptual system typically employs a concentrated or near-concentrated prior for disparity and texture cues, resulting in complete or near-complete fusion. Nardini, Bedford, and Mareschal (2010) found that 6 year olds do not fuse disparity and textures cues to slant. Instead, they maintain separate disparity-based and texture-based slant-estimates.

The estimates can diverge considerably. Thus, developing perceptual systems handle intramodal sensory cues differently than adult perceptual systems. The reason: developing perceptual systems have not learned how the distal cause revealed by one cue correlates with the distal cause revealed by the other cue. More formally: the coupling prior employed by developing perceptual systems is much less concentrated than that employed by adult perceptual systems. Similarly, the coupling prior over visual and haptic size-estimates begins relatively flat and becomes more concentrated with experience (van Dam, Parise, and Ernst, 2014, p. 222).

When deciding how to handle disparate sensory signals, the perceptual system faces two questions regarding the etiology of the signals:

(1) Are the signals caused by the same *object* (e.g. one dog versus two dogs)?

(2) Are the signals caused by the same *property* (e.g. "visual size" versus "haptic size")?

The causal inference model mentioned in §2.1 nicely handles type (1) questions. However, the model implicitly assumes that signals deriving from the same object also derive from the same property, so it cannot handle situations where the perceptual system determines that two signals derive from the same object but remains uncertain whether they derive from the same property. (Cf. Ernst and Di Luca, 2011, p. 239.) For this reason, the model is not well-suited to address type (2) questions.[8]

The coupling prior framework easily addresses type (2) questions. It can also address type (1) questions, by building implicit expectations about causal structure into the coupling prior (Ernst and Di Luca, 2011). However, the coupling prior framework does not explicitly

[8] Hospedales and Vijayakumar (2009) analyze the experimental results from (Hillis et al., 2002) using a causal inference model. Their analysis hinges upon a very specific feature of the (Hillis et al., 2002) experimental setup: namely, that the experiment involves an oddity-detection task. It is unclear how if at all one might generalize the analysis to other experimental setups, such as the discrimination task studied in (Roach, Heron, and McGraw, 2006). Adams (2016) compares how the coupling prior approach and the causal inference approach handle cross-modal effects in a visual-auditory estimation task. She concludes that the coupling prior approach fits the data much better.

model perceptual estimation of causal structure. For example, it does not explicitly model the difference between estimating that visual and auditory input derive from a single dog versus estimating that they derive from two different dogs. The causal inference model arguably provides a better setting for addressing type (1) questions. In the present paper, I am primarily concerned with type (2) questions, which is why I focus on the coupling prior framework. Obviously, a complete treatment of cue combination must address questions of both types.

§3. Co-referring perceptual representations

Any Bayesian perceptual model features *probabilities* attached to *hypotheses*. What are the "probabilities"? And what are the "hypotheses" to which they attach?

Following conventional wisdom, my answer to the first question is that the probabilities are *subjective*. They register the degree of confidence that the perceptual system reposes in a hypothesis. Thus, they reflect aspects of the perceiver's psychology, rather than objective chances in the distal environment.

My answer to the second question is that the hypotheses are *mental representations*. A mental representation is a mental item that represents. More explicitly: (a) it is the kind of thing that can be instantiated by mental states, events, or processes; and (b) it has representational properties. Hypotheses posited within Bayesian perceptual psychology satisfy (a): perceptual activity instantiates hypotheses by maintaining priors over them, by reallocating probabilities over them in rough conformity to Bayesian norms, and by selecting a particular hypothesis as a privileged estimate of distal conditions. Hypotheses also satisfy (b). They represent specific distal properties, including shapes, sizes, and colors (Rescorla, 2015). For example, a Bayesian model of size perception posits hypotheses that represent possible sizes, a Bayesian model of

slant perception posits hypotheses that represent possible slants, and so on. Any Bayesian perceptual model posits an array of hypotheses that represent distal properties. We may aptly call the hypotheses *perceptual representations*.

Say that two mental representations *co-refer* when they represent the same object or property. Current scientific theories of cue combination presuppose that, in many cases, the perceptual system employs co-referring perceptual representations. To illustrate, consider size-estimation based on visual and haptic cues. As we saw in §2, the perceptual system maintains a coupling prior $p(s_V, s_H)$. s_V is a vision-based hypothesis that represents some distal size. s_H is a touch-based hypothesis that represents some distal size. In my terminology, s_V and s_H are perceptual representations that represent distal sizes. Depending on s_V and s_H, they may represent either the same size or different sizes. Even when s_V and s_H represent the same size, they are distinct perceptual representations. They must be distinct, because otherwise there would not be a two-dimensional space of visual size-estimates and haptic size-estimates. There would only be a one-dimensional space of size-estimates, and no coupling prior would be possible.

Compare luminance and stiffness. There is an array of representations $l_1, l_2, l_3, \ldots$ that represent possible luminance values, and there is a second array of representations $s_1, s_2, s_3, \ldots$ that represent possible degrees of stiffness. Each representation l_i is distinct from each representation s_k. In the Ernst (2007) experiment, the perceptual system acquires a probabilistic correlation (encapsulated by the coupling prior) between the arrays $l_1, l_2, l_3, \ldots$ and $s_1, s_2, s_3, \ldots$. It can only encode such a correlation if there are luminance-representations and distinct stiffness-representations to correlate. Similarly, the developing perceptual system acquires a probabilistic correlation between size *as represented visually* and size *as represented haptically*. It can only encode such a correlation if there are vision-based size-representations $s_{V_1}, s_{V_2}, s_{V_3}, \ldots$ and distinct

touch-based size-representations $s_{H_1}, s_{H_2}, s_{H_2}, \ldots$ to correlate. We theorists know that luminance is not stiffness but that a certain size *as represented visually* is identical to that same size *as represented haptically*. The developing perceptual system has no such knowledge at its initial disposal. It must discover for itself how size *as represented visually* correlates with size *as represented haptically*. To do so, it requires separate vision-based and touch-based size-representations.

A similar analysis applies to intramodal cue combination. Consider slant-estimation based on disparity and texture cues. As we saw in §2, the developing perceptual system computes quite distinct slant-estimates based upon disparity and texture cues, whereas the adult perceptual system approaches complete fusion. We can explain the contrast by positing a change in the coupling prior over disparity-based slant-estimates and texture-based slant-estimates. The prior begins relatively flat, then becomes concentrated or near-concentrated. Our explanation presupposes a two-dimensional space of disparity-based slant-estimates and texture-based slant-estimates. It presupposes an array of disparity-based representations that represent possible slants and a separate array of texture-based representations that represent possible slants. Even when a disparity-based representation co-refers with a texture-based representation, they are distinct representations that occupy distinct roles in perceptual processing.

That co-referring perceptual representations occupy distinct psychological roles is especially evident in the intermodal case. As noted above, sensory signals from different modalities typically lead the perceptual system to form separate estimates of a single distal variable (Hillis et al., 2002). The separate estimates are given by distinct mental representations. In this manner, the intermodal case vividly illustrates how the perceptual system computes over distinct arrays of mental representations representing possible values of a single distal variable.

The intramodal case does not afford such a vivid illustration, at least for the adult perceptual system where complete fusion occurs. The adult perceptual system does not typically form separate estimates corresponding to distinct cues within a single modality. Hence, there is perhaps not as much intuitive pressure for us to postulate distinct arrays of perceptual representations. However, the *developing* perceptual system does not completely fuse (Nardini et al. 2010), so the intuitive pressure is still present for developing perceivers. In any event, the coupling prior framework requires us to posit co-referring perceptual representations for all these cases. Even when the coupling prior is concentrated, it is defined over a two-dimensional space formed by distinct arrays of mental representations.

§4. Berkeley on perceptual ideas

My position has historical routes tracing back at least to Berkeley's *An Essay Towards a New Theory of Vision* (1709/1948). As was standard at the time, Berkeley spoke of "ideas" rather than "mental representations." His core thesis in the *Essay* is that visual ideas are distinct from haptic ideas. He writes: "it is plain the Objects of Sight and Touch make, if I may so say, two Sets of Ideas, which are widely different from each other" (CXI), and "*The Extension, Figures, and Motions, perceived by Sight are specifically distinct from the Ideas of Touch, called by the same Names, nor is there any such thing as one Idea, or kind of Idea common to both Senses*" (CXXVII). Berkeley holds that the perceptual system must learn through experience how visual and haptic ideas correlate (XLV, CIV). He defends a similar position regarding visual and auditory ideas (XLVI-XLVIII, CXXX). He does not address whether different cues within a single modality are associated with distinct ideas.

Berkeley supplements his core thesis with numerous doctrines that I reject. Some of the doctrines are characteristic of his era. Some are idiosyncratically his own. A few highlights:

(i) Like many early modern philosophers, Berkeley holds that ideas are the immediate "objects" of perception (XLIX, CXI, CXXIX, CLXVII). If you see a tree, what you most directly see are your visual ideas of the tree. If you touch the tree, what you most directly feel are your haptic ideas of the tree.

(ii) Berkeley holds that visual ideas acquire determinate representational import only by virtue of correlations with haptic ideas (LXII-LXIV, LXIX, CXLIII-CXLVII). Visual ideas come to represent distal properties (such as shapes, sizes, or distances) only when the developing perceptual system learns how they correlate with haptic ideas.

(iii) In other writings (1710/1949), although perhaps not explicitly in the *Essay*, Berkeley espouses the radical *idealist* position that material things are composed of ideas.

I reject all three doctrines:

(i) Like most contemporary philosophers, I reject the early modern prejudice that mental representations, rather than the worldly objects and properties represented by mental representations, are the immediate objects of perception. We do not literally perceive mental representations. We perceive distal objects and their distal properties. We do so by way of instantiating perceptual representations, but we do not thereby perceive the perceptual representations themselves.

(ii) I deny that visual perceptual representations depend for their representational import on correlations with haptic perceptual representations. The scientific work surveyed in §2 posits Bayesian inference over visual and haptic representations that represent distal properties *quite independently of any particular correlation instituted by the*

perceptual system between visual and haptic representations. This work undermines Berkeley's claim that vision is parasitic upon touch for its representational import.

(iii) Needless to say, I reject all versions of idealism.

There are many other aspects of Berkeley's discussion that I reject but that I will not mention. Of course, his discussion is scientifically outdated. Nevertheless, I agree with his core thesis. I agree that visual cues and haptic cues generate distinct perceptual representations ("ideas"), and similarly for visual and auditory cues. My discussion extends Berkeley's core thesis by associating distinct *intramodal* cues with distinct perceptual representations.

Berkeley defends his core thesis partly through discussion of *Molyneux's question*. The question is whether someone born blind would, upon gaining sight, immediately recognize how visually perceived shape correlates with haptically perceived shape. Berkeley answers the question negatively (CXXXII). He argues on this basis that visual ideas of shape are distinct from haptic ideas of shape and that one must learn through experience how the ideas correlate (CXXXIII-CXXXV). He pursues a similar argumentative strategy regarding visual and haptic perception of distance (XL-XLII). Since Berkeley's time, there has been considerable psychological research on Molyneux's question, including recent work that arguably supports a negative answer (Sinha, Wulff, and Held, 2014). However, the currently available experimental data is not decisive (Schwenkler, 2012), (Van Cleve, 2014). I have therefore opted not to invoke Molyneux's question in developing my own argument. I have instead emphasized the explanatory structure instantiated by Bayesian models of cue combination.

§5. Perceptual modes of presentation

I now elucidate perceptual co-reference using some tools introduced by Frege.

In a seminal discussion, Frege (1892/1997) adduced cases where a thinker does not recognize an entity as the same because she represents it in different ways. One may not realize *that Hesperus is Phosphorus*, even though Hesperus and Phosphorus are one and the same entity (Venus). One may not realize *that mercury is quicksilver*, or *that cilantro is coriander*, or *that groundhogs are woodchucks*, or that *Obamacare is the Affordable Care Act*. These are now usually called "Frege cases." The moral Frege drew from Frege cases is that one can think about a single entity in different ways. He argued that a good theory should posit *ways of representing* entities, which he called *modes of presentation*. Frege said relatively little about what "modes of presentations" are. Partly as a result, critics often condemn his approach as overly obscure. As Fodor (2008) argues, though, modes of presentation do not seem particularly obscure once we accept that mental activity involves computation over mental representations. Having endorsed mental representations, we have committed ourselves to mental items that "present" entities to thought. We can then gloss *modes of presentation* as *mental representations*. For example, we can postulate two mental representations that denote mercury in different ways: the first represents it *as mercury*, while the second represents it *as quicksilver*.

Frege focused on modes of presentation as they arise in high-level cognition. Recently, several authors have suggested that we should generalize by positing *perceptual modes of presentation* (Burge, 2010), (Chalmers, 2004), (Kulvicki, 2007), (Peacocke, 1989), (Thompson, 2010). In effect, §3 constitutes an argument for perceptual modes of presentation. As the argument highlights, there are different ways of perceptually representing a single distal property. The perceptual system may represent size in a vision-based way or a touch-based way. It may represent slant in a disparity-based way or a texture-based way. Following Fodor, I have

glossed the "ways" as mental representations. Distinct perceptual representations of a single distal property "present" the property differently for purposes of perceptual computation.

Philosophers who defend a broadly Fregean approach to perception usually (e.g. Chalmers, 2004; Kulvicki, 2007; Thompson, 2010), although not invariably (e.g. Burge 2010), emphasize *phenomenology*. They adduce phenomenological differences between perceptual experiences --- differences in "what it is like" to have the experiences. They infer that the experiences involve distinct perceptual modes of presentation. However, any argument along these lines embodies a contestable picture of the relation between phenomenological and representational aspects of experience. For example, the undeniable phenomenological differences between a visual percept of size and a haptic percept of size do not immediately entail any difference among mental representations. One might hold that a single mental representation of size is associated with one phenomenology when deployed by vision and a different phenomenology when employed by touch. Whether such a position is tenable hinges upon a controversial question: whether phenomenological properties supervene upon representational properties.

My own argument does not rely upon phenomenological considerations. Instead, I emphasize the role that perceptual representations play within perceptual computation. I claim that our current best computational theories of sensory cue combination presuppose co-referring perceptual representations corresponding to distinct sensory cues. One advantage of my argumentative strategy is that it generalizes readily from intermodal cue combination to intramodal cue combination. Phenomenological arguments do not so generalize.

To strengthen the connection with Frege's discussion, let us consider *probabilistic Frege cases*. As Chalmers (2011) emphasizes, we can construct probabilistic analogues to Frege's examples. For instance, we can imagine a thinker who harbors the probabilities

p(*Mercury is in the beaker*) = .9

p(*Quicksilver is in the beaker*) = .2

and the conditional probabilities

p(*Mercury is in the beaker* | *Mercury is in the beaker*) = 1

p(*Mercury is in the beaker* | *Quicksilver is in the beaker*) = .2.

The thinker is uncertain as to whether mercury is quicksilver. As a result, her *mercury* mode of presentation figures in different unconditional and conditional probabilities than her *quicksilver* mode of presentation. §3's argument for co-referring perceptual representations hinges upon probabilistic perceptual Frege cases. A non-concentrated coupling prior $p(s_V, s_H)$ encodes uncertainty regarding the correlation between "visual size" and "haptic size." The perceptual system is uncertain as to whether a distal size *as represented in visual terms* is identical to a distal size *as represented in haptic terms*. As a result, vision-based size-representations figure in different unconditional and conditional probabilities than touch-based size-representations. In both cognitive and perceptual cases, we posit distinct co-referring modes of presentation so as to capture the system's subjective probabilities and probabilistic inferences.

In cognitive Frege cases, the thinker can explicitly represent identity. She can recognize *that Hesperus is Phosphorus*, or *that mercury is quicksilver*. It is much less evident that the perceptual system can explicitly represent identity, especially identity relations among properties. Nevertheless, the perceptual system has available a probabilistic analogue of identity judgments. A concentrated coupling prior encodes certainty in a particular correlation between visual size-estimates and haptic size-estimates. For many purposes, a concentrated coupling prior

serves as an analogue to the explicit judgment *that Hesperus is Phosphorus*. Viewed in this light, the transition from a flat coupling prior to a concentrated coupling prior looks analogous to the transition from not knowing *that Hesperus is Phosphorus* to knowing *that Hesperus is Phosphorus*.

§5. Explaining perception

An important moral emerges: explanation within perceptual psychology must sometimes consider mode of presentation and not simply denotation. When we seek to explain how the perceptual system estimates a distal variable, we must adduce the *way* that the perceptual system represents the variable's values. Mode of presentation crucially informs how perceptual computation proceeds. For example, the perceptual system can represent size in a vision-based way or a touch-based way, and it can represent slant in a texture-based way or a disparity-based way. Those differences influence the course of perceptual inference. The influence becomes apparent within the coupling prior framework, which posits a prior probability defined over estimates differentiated not just by their denotations but also by the way the perceptual system represents the denotations. Good explanation of intermodal and intramodal cue combination must attend to these fine-grained differences.

Over the past few decades, philosophers have intensely debated how we should taxonomize mental states for purposes of psychological explanation. The debate concerns how fine-grained a taxonomization we should employ. Fregeans advocate a relatively fine-grained taxonomization that takes mode of presentation into account. The literature offers various opposing coarse-grained taxonomic schemes, one of the most widely discussed being the *Russellian* scheme. So-called due to its origin in Russell's (1903) work, the Russellian scheme

eschews modes of presentation. It classifies mental states by citing denotations and representational properties determined by denotations. On a Russellian approach, we should not postulate modes of presentation over and above denotations. Various philosophers have developed the Russellian approach, usually focusing on high-level cognition (Soames, 2002) but sometimes applying it to perception as well (Thau, 2002).

While the Russellian approach may be useful for certain purposes, §§3-5 cast doubt upon whether it provides an adequate foundation for perceptual psychology. Good explanation of sensory cue combination apparently requires a finer-grained Fregean taxonomic scheme that cites modes of presentation over and above denotations.

The philosophical literature offers various strategies through which a committed Russellian might try to circumvent modes of presentation. Russellians typically insist that an agent represents some denotation by representing properties that distinguish the denotation from all other possible denotations. They then explain Frege cases by citing different represented properties that single out the same denotation. For example, a Russellian might seek to differentiate Hesperus-thoughts from Phosphorus-thoughts not by invoking different modes of presentation but rather by invoking different distinguishing properties represented by a thinker: a Hesperus-thought represents the heavenly body as appearing at certain positions at certain times, while a Phosphorus-thought represents it as appearing at certain other positions at certain other times. Russellian maneuvers along these lines have been extensively debated in the literature. However compelling they may be for the case of high-level cognition (and I myself do not find them compelling), they seem quite implausible when applied to perception.

To illustrate, suppose a Russellian tries to distinguish visual and haptic size-estimates along the following lines: the visual estimate represents a given size as *the cause (or typical*

cause) of certain visual stimulations or sensations and the haptic estimate represents that same size as *the cause (or typical cause) of certain haptic stimulations or sensations*. This Russellian maneuver differentiates visual and haptic size-estimates by citing properties represented by the perceptual system, without invoking modes of presentation beyond such represented properties. The maneuver is problematic, because it hinges upon the unsupported and implausible claim that perception represents distal sizes as causes of sensory states. When I perceive an object as having a certain size, I do not perceive its size *as* causing me to have certain stimulations or sensations. I may upon reflection represent such causal relations *within high-level thought*, but I do not in any natural sense *perceive* the causal relations. I perceptually represent the object's size, not the causal influence that the object's size exerts upon my sensory apparatus. Common sense and perceptual psychology both reject any suggestion that, when I perceive some distal property, I thereby perceive the distal property as causally influencing my own sensory states. We may safely set the Russellian maneuver aside.[9]

§6. Instantiating a perceptual representation

I have argued that Bayesian perceptual psychology posits co-referring mental representations. My position enshrines a fine-grained Fregean conception of psychological explanation, as opposed to a comparatively coarse-grained Russellian conception. The question remains just *how* fine-grained a taxonomic scheme we require. Under what conditions does mental activity instantiate a given perceptual representation? How finely should we differentiate among perceptual representations? For example, under what conditions does mental activity

[9] Burge (1991) critiques a proposal, due to Searle (1983), according to which each visual experience represents causal relations between distal conditions and that very visual experience. Much of Burge's critique readily extends to the Russellian proposal that perception represents distal properties as causally influencing sensory states.

instantiate vision-based size-representation s_{V_i}? What changes to perceptual processing entail that s_{V_i} is no longer being instantiated?

I will not provide systematic answers to these questions. But I will address two phenomena that good answers should take into account: *perceptual adaptation* and *perceptual constancies*.

§5.1 Perceptual adaptation

Perceptual computation constantly evolves in response to a changing environment or a changing interface between perceiver and environment. *Perceptual adaption* is "a semipermanent change in perception or perceptual-motor coordination which serves to effectively reduce or eliminate an apparent discrepancy between or within sensory modalities or the errors introduced by this discrepancy" (Welch, 1978, p. 8). Examples:

- *Luminance-stiffness prior*. As discussed in §2, Ernst (2007) exposed subjects to deviant stimuli that altered the coupling prior over luminance and stiffness. As a result, perceptual estimation of luminance and stiffness changed.
- *Shape from shading*. A concave stimulus lit from overhead generates the same retinal shading as a convex stimulus lit from below. To infer shape from the ambiguous shading cue, the perceptual system deploys a prior over possible lighting directions. The prior assigns higher probability to overhead lighting directions. Adams, Graf, and Ernst (2004) experimentally manipulated the prior by exposing subjects to deviant visual-haptic stimuli indicating a non-standard lighting direction. The experimental manipulation shifted the prior towards the non-standard lighting direction, yielding a significant change in visual shape-estimation.

- *Ventriloquism aftereffect*. Suppose we repeatedly expose a perceiver to a ventriloquism illusion: a visual stimulus and an auditory stimulus with fixed spatial separation. The perceptual system will then change how it estimates location based solely upon auditory cues. Perceived location of the auditory stimulus *without any accompanying visual stimulus* shifts markedly along the spatial separation. This is the *ventriloquism aftereffect*. Sato, Toyoizumi, and Aihara (2007) argue that the aftereffect reflects a change in the prior likelihood $p(e_A \mid l_A)$ relating audition-based location-estimates l_A and auditory input e_A. Intuitively: when sustained ventriloquism occurs, the perceptual system changes its expectations regarding which auditory stimulations will result from a sound at a given location.

In each example, the prior probabilities or the prior likelihoods change so that they more closely match changing environmental statistics.

Do perceptual representations change whenever the priors change? Or can a perceptual representation persist as the priors change?

My answer is that a perceptual representation can persist even as the priors change. In each example of perceptual adaptation (luminance-stiffness, shape from shading, and the ventriloquism aftereffect), the perceptual system responds to the changing environment by reallocating probabilities over a hypothesis space containing perceptual representations. The hypothesis space, and the perceptual representations contained therein, remain fixed. Otherwise, the perceptual system would not be *reallocating* probabilities over the hypothesis space. It would instead be replacing one hypothesis space with another. Consider the ventriloquism aftereffect, which involves a shift in the prior likelihood $p(e_A \mid l_A)$ relating location-estimate l_A and auditory input e_A. What changes is the conditional probability $p(e_A \mid l_A)$ assigned to e_A given l_A. In order

for $p(e_A \mid l_A)$ to shift, l_A must remain fixed. Persistence of l_A through adaptation is built into the adaptation model offered by Sato, Toyoizumi, and Aihara (2007, p. 3341). The model contains an explicit rule governing how the old conditional probability of e_A given l_A is replaced by a new conditional probability of *the same* e_A given *the same* l_A. The rule presupposes that a persisting perceptual representation l_A figures in both the old and the new prior likelihood.

A similar diagnosis applies to other Bayesian models of perceptual adaptation, such as the models found in (Burge, Ernst, and Banks, 2008), (Ernst and Di Luca, 2011), (Stocker and Simoncelli, 2006). Each model presupposes a fixed hypothesis space and describes how unconditional or conditional probabilities redistribute over the space due to environmental changes. The same perceptual representations figure in both the old and the new probability assignments. Assuming that these models are on the right track, we must recognize that a single fixed perceptual representation can participate in a range of priors. The perceptual representation persists even as the priors change.

The upshot: a perceptual representation can persist through significant changes in perceptual processing.

§5.2 Perceptual constancies

A *perceptual constancy* is a capacity to represent some fixed distal property despite radical variation in proximal sensory input. For example, very different retinal angles *a* and *b* can cause a perceptual estimate of size *s*, so long as the perceptual system estimates that the object causing angle *a* is located at a suitably different depth than the object causing angle *b*. Similarly, the perceptual system may represent a surface as square despite considerable variation in the

shape that the surface casts upon the retina. Human perception has constancies for shape, size, color, location, depth, and many other distal aspects of the environment.

Burge (2010) invokes perceptual constancies to defend a very fine-grained conception of perceptual representation. He posits items that he calls *perceptual attributives*: "A perceptual attributive is an aspect of perceptual representational content that functions to indicate a repeatable type and to group or characterize purported particulars as being of that type" (2010, p. 380). There are perceptual attributives that represent distal sizes, shapes, colors, and so on. Burge argues that significant differences in proximal input yield different perceptual attributives. For example, different attributives occur when retinal angle *a* triggers a perceptual estimate of distal size *s* and when distinct retinal angle *b* triggers a perceptual estimate of that same size *s*. Burge glosses his approach in Fregean terms: the attributives represent a single referent *s*, but they represent it in different ways (2010, pp. 40-41). Similarly for other cases where different proximal stimulations trigger perceptual estimates of a single distal property.

In (Rescorla, 2014), I questioned Burge's fine-grained conception. I asked why we should say that different proximal stimulations trigger different perceptual attributives. To illustrate, suppose that retinal angles *a* and *b* trigger perceptual estimates of a single size *s*. Why posit different size-representations corresponding to *a* and *b*, rather than saying that different retinal angles can trigger the same perceptual size-representation? Burge (2014) replies to my critique and offers new arguments for his fine-grained conception. For reasons of space, I will not address the new arguments. Instead, I will explore how Burge's fine-grained conception relates to Bayesian modeling of perception.

The main claim I wish to defend is that Bayesian perceptual psychology does not enshrine Burge's fine-grained conception. Consider Ernst's (2006) model of visual-haptic size-

estimation. The model employs a two-dimensional space of vision-based and touch-based estimates. For each distal size *s*, the model presupposes a *unique* vision-based representation s_{V_i} that represents *s* and a *unique* touch-based representation s_{H_k} that represents *s*. The model does not differentiate between vision-based size-representations triggered by one retinal angle and vision-based representations triggered by a different retinal angle. In general, Bayesian models do not differentiate among perceptual representations based upon triggering proximal stimulation. Prior probabilities and prior likelihoods are defined over perceptual representations that lack privileged ties to specific patterns of proximal input. A perceptual representation *h* participates in many conditional probabilities $p(e_1 \mid h)$, $p(e_2 \mid h)$, …, $p(e_n \mid h)$, …, where each e_n is a different proximal input. Depending on the model's details, many different proximal inputs e_n may trigger the same final perceptual estimate $\hat{h}$. The perceptual representation $\hat{h}$ is not tied to any one e_n.

My analysis suggests a relatively coarse-grained conception of perceptual representations, along the following lines. The perceptual system has the capacity to estimate the value of some distal variable (e.g. depth) based upon some sensory cue (e.g. disparity). Estimation deploys priors $p(h)$ and $p(e \mid h)$, where *e* reflects a possible value of the sensory cue and where *h* is a perceptual representation that represents a possible value of the distal variable. Representation *h* is tied to the sensory cue, but it is not tied to any specific values of the cue. Different sensory cues (e.g. disparity versus motion parallax) typically entail different perceptual representations. Different values of a single cue (e.g. different disparities) do not. Radically different inputs *e* may trigger the same final perceptual estimate $\hat{h}$.

My conception *differs* from Burge's, but the conceptions do not necessarily *conflict*. Burge might reply that coarse-grained representations figure in the subpersonal processes that

produce the percept but that the final percept contains a finer-grained perceptual attributive. This reply seems consistent with Bayesian perceptual psychology, although I am not sure how plausible it is or how compelling Burge would find it.

There is another way to reconcile the two conceptions: one might treat them as different but compatible ways of describing the same perceptual states. By analogy, consider utterances of the following sentences:

The police booked the suspect.

John booked a hotel room.

One might hold that the utterances involve a single word "booked." Alternatively, one might hold that the utterances involve two distinct words pronounced the same way. Finally, one might hold that these are both legitimate descriptions, embodying different but legitimate conceptions of *word*: a coarse-grained conception on which relatively many utterances instantiate the same word, or a fine-grained conception on which relatively few utterances instantiate the same word. One might say that the different conceptions yield different but equally legitimate ways of classifying utterances. Similarly, one might hold that my coarse-grained conception of perceptual representations and Burge's finer-grained conception are different but equally legitimate ways of classifying perceptual states. On the coarse-grained conception, perceptual states can instantiate the same perceptual representation of size even though they result from very different retinal angles. On the finer-grained conception, any such perceptual states instantiate distinct perceptual representations of size. Both conceptions are legitimate, although one conception might serve certain explanatory purposes better than the other.

Let us grant that, for some purposes, it is fruitful to differentiate between a perceptual size-representation triggered by retinal angle *a* and a perceptual size-representation triggered by

different retinal angle *b*. Let us grant that we should sometimes classify perceptual states in this fine-grained way. Even so, a coarser-grained conception seems more appropriate for many purposes. Bayesian perceptual models embody the coarser-grained conception. The models posit perceptual representations tied to specific sensory cues but not to specific values of those cues.

§6. Conclusion

I have argued that different sensory cues are typically associated with distinct sets of perceptual representations. I have also defended a relatively coarse-grained conception of these representations. A single perceptual representation can persist despite changing priors (§5.1) and despite significant variation in proximal input (§5.2). Thus, although my broadly Fregean treatment is finer-grained than a Russellian treatment, it is not as fine-grained an approach as some Fregeans have advocated.

Many questions remain. Under what conditions does a perceptual state instantiate a given perceptual representation? What is the difference between instantiating vision-based representation s_{V_i} versus co-referring touch-based representation s_{H_k}? Generally speaking, what distinguishes co-referring perceptual representations? Intuitively, the primary difference between a vision-based size-representation and a touch-based size-representation is that the representations have canonical links to different information sources. The vision-based representation is canonically linked to retinal stimulations, including retinal angle and depth cues. The touch-based representation is canonically linked to input from sensors that detect finger position. Similarly, one depth-representation may be canonically linked to disparity cues while another is canonically linked to motion-parallax cues. More generally, canonical links to different aspects of proximal sensory stimulation help determine whether a perceptual state

instantiates a particular perceptual representation. To unpack this idea, one must say what the "canonical links" consist in. Since I have not done so, I do not claim to have provided anything like a complete account. My goal has been to make progress by scrutinizing some well-confirmed models of perceptual computation.

Works Cited

Adams, W. 2016. "The Development of Audio-Visual Integration for Temporal Judgements." *PLoS Computational Biology* 12: e1004865.

Adams, W., Graf, E., and Ernst, M. 2004. "Experience Can Change the 'Light-From-Above' Prior." *Nature Neuroscience* 7: pp. 1057-1058.

Adams, W., Kerrigan, I., and Graf, E. "Touch Influences Perceived Gloss." *Scientific Reports* 6: p. 21866.

Alais, D., and Burr, D. 2004. "The Ventriloquist Effect Results from Near-optimal Bimodal Integration." *Current Biology* 14: pp. 257-262.

Berkeley, G. 1709/1948. *An Essay Towards a New Theory of Vision*. Rpt. in *The Works of George Berkeley, Bishop of Cloyne*, vol. 1, eds. E. E. Luce and T. E. Jessop. Edinburgh: Thomas Nelson.

---. 1710/1949. *A Treatise Concerning the Principles of Human Knowledge*. Rpt. in *The Works of George Berkeley, Bishop of Cloyne*, vol. 2, eds. E. E. Luce and T. E. Jessop. Edinburgh: Thomas Nelson.

Burge, T. 1991. "Vision and Intentional Content." In *John Searle and His Critics*, eds. E. Lepore and R. Van Gulick. Malden: Blackwell.

---. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

---. 2014. "Reply to Rescorla and Peacocke: Perceptual Content in Light of Perceptual Constancies and Biological Constraints." *Philosophy and Phenomenological Research* 88: pp. 485-501.

Burge, J., Ernst, M., and Banks, M. 2008. "The Statistical Determinants of Adaptation Rate in Human Reaching." *Journal of Vision* 8: pp. 1-19.

Chalmers, D. 2004. "The Representational Character of Experience." In *The Future for Philosophy*, ed. B. Leiter. Oxford: Clarendon Press.

---. 2011. "Frege's Puzzle and the Objects of Credence." *Mind* 120: pp. 587-365.

Ernst, M. 2006. "A Bayesian View on Multimodal Cue Integration." In *Human Body Perception From the Inside Out*, eds. G. Knoblich, I. Thornton, M. Grosjean, and M. Shiffrar. Oxford: Oxford University Press.

---. 2007. "Learning to Integrate Arbitrary Signals from Vision and Touch." *Journal of Vision* 7: pp. 1-14.

---. 2012. "Optimal Multisensory Integration: Assumptions and Limits." In *The New Handbook of Multisensory Processes*, ed. B. Stein. Cambridge: MIT Press.

Ernst, M., and Banks, M. 2002. "Humans Integrate Visual and Haptic Information in a

Statistically Optimal Fashion." *Nature* 415: pp. 429-433.
Ernst, M., and Di Luca, M. 2011. "Multisensory Perception: From Integration to Remapping." In *Sensory Cue Integration*, eds. J. Trommershäuser, K. Körding, and M. Landy. Oxford: Oxford University Press.
Fodor, J. 1975. *The Language of Thought*. New York: Thomas Y. Crowell.
---. 2008. *LOT2*. Oxford: Clarendon Press.
Frege, G. 1892/1997. On *Sinn* and *Bedeutung*. Rpt. in *The Frege Reader*, ed. M. Beaney, trans. M. Black. Malden: Blackwell.
Gallistel, C. R. and King, A. 2009. *Memory and the Computational Brain*. Malden: Wiley-Blackwell.
Girshick, A., and Banks, M. 2009. "Probabilistic Combination of Slant Information: Weighted Averaging and Robustness as Optimal Percepts." *Journal of Vision* 9: pp. 1-20.
Helbig, H., and Ernst, M. 2008. "Haptic Perception in Interaction with Other Senses." In *Human Haptic Perception: Basics and Applications*, ed. M. Grunwald. Boston: Birkhäuser Verlag.
Helmholtz, H. von. 1867. *Handbuch der Physiologischen Optik*. Leipzig: Voss.
Hillis, J., Ernst, M., Banks, M., Landy, M. 2002. "Combining Sensory Information: Mandatory Fusion Within, but Not Between, Senses." *Science* 298: pp. 1627-1630.
Hospedales, T., and Vijayakumar, S. 2009. "Multisensory Oddity Detection as Bayesian Inference." *PloS One* 4: e4205.
Jacobs, R.A. 1999. "Optimal Integration of Texture and Motion Cues to Depth." *Vision Research* 39: pp. 3621–3629.
Knill, D., and Richards, W. 1996. *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.
Knill, D., and Saunders, J. 2003. "Do Humans Optimally Integrate Stereo and Texture Information for Judgments of Surface Slant?". *Vision Research* 43: pp. 2539–2558.
Körding, K., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum J., and Shams L. 2007. "Causal Inference in Multisensory Perception." *PLoS One* 2: e943.
Kulvicki, J. 2007. "What Is What It's Like?". *Synthese* 156: pp. 205-229.
Landy, M., Banks, M., and Knill, D. 2011. "Ideal-Observer Models of Cue Integration." In *Sensory Cue Integration*, eds. J. Trommershäuser, K. Körding, and M. Landy. Oxford: Oxford University Press.
Nardini, M., Bedford, R., Mareschal, D. 2010. "Fusion of Visual Cues is Not Mandatory in Children." *Proceedings of the National Academy of Sciences* 107: pp. 17041-17046.
O'Callaghan, C. 2012. "Perception and Multimodality." In *The Oxford Handbook to Philosophy and Cognitive Science*, eds. E. Margolis, R. Samuels, and S. Stich. Oxford: Oxford University Press.
Peacocke, C. 1989. "Perceptual Content." In *Themes from Kaplan*, eds. J. Almog, J. Perry, and H. Wettstein. Oxford: Oxford University Press.
Prsa, M., Gale, S., and Blank, O. 2012. "Self-motion Leads to Mandatory Cue Fusion Across Sensory Modalities." *The Journal of Neurophysiology* 108: pp. 2282-2291.
Pylyshyn, Z. 1984. *Computation and Cognition*. Cambridge: MIT Press.
Rescorla, M. 2014. "Perceptual Constancies and Perceptual Modes of Presentation." *Philosophy and Phenomenological Research* 88: pp. 468-476.
---. 2015. "Bayesian Perceptual Psychology." In *The Oxford Handbook of the Philosophy of Perception*, ed. M. Matthen. Oxford: Oxford University Press.

Roach, N., Heron, J., and McGraw, C. 2006. "Resolving Multisensory Conflict: A Strategy for Balancing the Costs and Benefits of Audio-Visual Integration." *Proceedings of the Royal Society B* 273: pp. 2159-2168.

Rohde, M., van Dame, L., and Ernst, M. 2016. "Statistically Optimal Multisensory Cue Integration: A Practical Tutorial." *Multisensory Research* 29: pp. 297-317.

Russell, B. 1903. *Principles of Mathematics*. Cambridge: Cambridge University Press.

Sato, Y., Toyoizumi, T., and Aihara, K. 2007. "Bayesian Inference Explains Perception of Unity and Ventriloquism Aftereffect: Identification of Common Sources of Audiovisual Stimuli." *Neural Computation* 19: pp. 3335-3355.

Schwenkler, J. 2012. "On the Matching of Seen and Felt Shape by Newly Sighted Subjects." *i-Perception* 3: pp. 186-188.

Searle, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.

Sinha, P., Wulff, J., and Held, R. 2014. "Establishing Cross-Modal Mappings: Empirical and Computational Investigations." In *Sensory Integration and the Unity of Consciousness*, eds. D. Bennett and C. Hill. Cambridge: MIT Press.

Soames, S. 2002. *Beyond Rigidity*. Oxford: Oxford University Press.

Stocker, A., and Simoncelli, E. 2006. "Sensory Adaptation within a Bayesian Framework for Perception." In *Advances in Neural Information Processing Systems*, vol. 18, eds. Y. Weiss, B. Schölkopf, and J. Platt. Cambridge: MIT Press.

Thau, M. 2002. *Consciousness and Cognition*. Oxford: Oxford University Press.

Thompson, B. 2010. "The Spatial Content of Experience." *Philosophy and Phenomenological Research* 81: pp. 146-184.

van Beers, R., Wolpert, D., and Haggard, P. 2002. "When Feeling Is More Important Than Seeing in Sensorimotor Adaptation." *Current Biology* 12: pp. 834-837.

Van Cleve, J. 2014. "Berkeley, Reid, and Sinha on Molyneux's Question." In *Sensory Integration and the Unity of Consciousness*, eds. D. Bennett and C. Hill. Cambridge: MIT Press.

van Dam, L., Parise, C., and Ernst, M. 2014. "Modeling Multisensory Integration." In *Sensory Integration and the Unity of Consciousness*, eds. D. Bennett and C. Hill. Cambridge: MIT Press.

Welch, R. 1978. *Perceptual Modification: Adapting to Altered Sensory Environments*. New York: Academic Press.

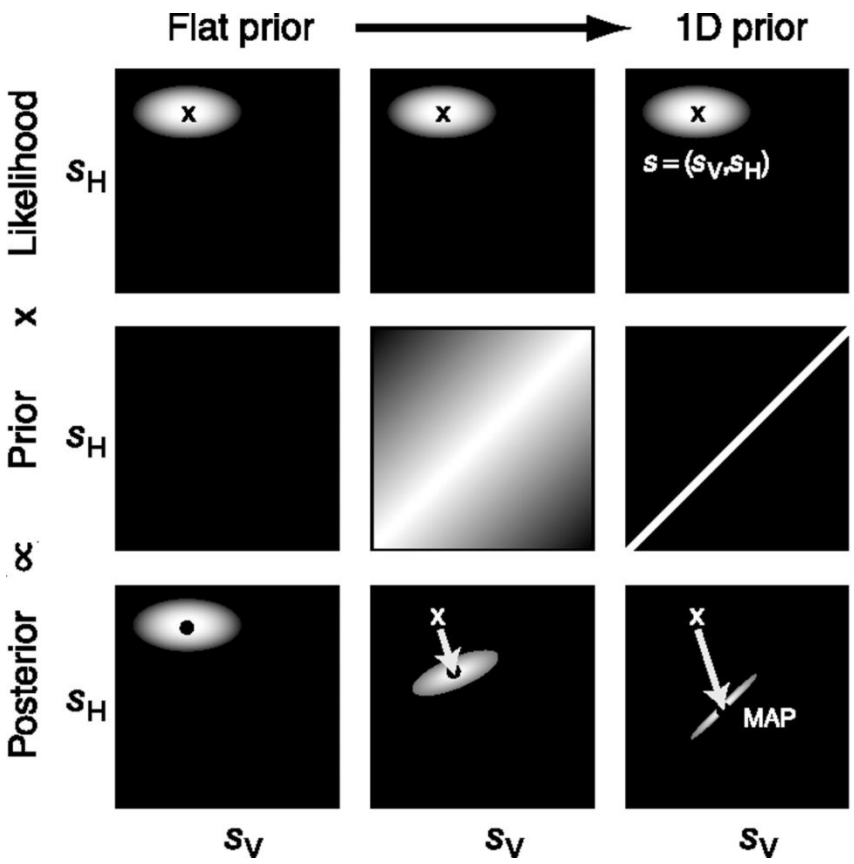

Figure 1. Each panel depicts the two-dimensional space of visual-haptic size-estimates $s = (s_V, s_H)$. The horizontal axis contains visual size-estimates s_V. The vertical axis contains haptic size-estimates s_H. White indicates high probability mass. Black indicates low probability mass. The top row depicts the likelihood function that results from a visual-haptic stimulus x with discrepant visual and haptic cues. Assuming no bias in either sensory channel, $x = (\hat{s}_V, \hat{s}_H)$. The middle row depicts three possible coupling priors, ranging from flat on the left to concentrated on the right. The bottom row depicts the posteriors that result from combining the likelihood with each coupling prior. • is the *maximum a posteriori* (MAP) estimate, i.e. the estimate that maximizes the posterior. The arrow indicates the extent to which the coupling prior biases • away from $(\hat{s}_V, \hat{s}_H)$ and towards the diagonal line $s_V = s_H$. Rpt. from (Ernst, 2007) by permission of the Association for Research in Vision and Ophthalmology.