# 1 The Neural Basis of General Recognition Theory

F. Gregory Ashby[1] and Fabian A. Soto[2]

[1]University of California, Santa Barbara
[2]Florida International University

## Abstract

General recognition theory (GRT; Ashby & Townsend, 1986) is a multivariate extension of signal detection theory to cases in which there is more than one perceptual dimension. GRT has traditionally only been concerned with behavioral data, but during the past decade or two, much has been learned about the architecture and functioning of the neural circuits that implement the perceptual and decision processes hypothesized by GRT. This chapter reviews this neuroscience literature, with a focus on three separate questions. First, what does the neuroscience literature say about the validity of GRT? Second, how can we use results from the neuroscience literature to improve GRT applications? Finally, how can GRT analyses be extended to neuroscience data?

## 1.1 INTRODUCTION

General recognition theory (GRT; Ashby & Townsend, 1986) is a multivariate extension of signal detection theory to cases in which there is more than one perceptual dimension. GRT has all the advantages of univariate signal detection theory (i.e., it separates perceptual and decision processes) but it also offers the best existing method for examining interactions among perceptual dimensions (or components). Since its inception, hundreds of articles have applied GRT to a wide variety of phenomena, including categorization (e.g., Ashby & Gott, 1988; Maddox & Ashby, 1993), similarity judgment (Ashby & Perrin, 1988), face perception (Blaha, Silbert, & Townsend, 2011; Soto, Vucovich, Musgrave, & Ashby, 2014; R. D. Thomas, 2001; Wenger & Ingvalson, 2002), recognition and source memory

(Banks, 2000; Rotello, Macmillan, & Reeder, 2004), source monitoring (DeCarlo, 2003), attention (Maddox, Ashby, & Waldron, 2002), object recognition (Cohen, 1997; Demeyer, Zaenen, & Wagemans, 2007), feature binding (Ashby, Prinzmetal, Ivry, & Maddox, 1996), perception/action interactions (Amazeen & DaSilva, 2005), auditory and speech perception (Silbert, 2012; Silbert, Townsend, & Lentz, 2009), haptic perception (Giordano et al., 2012; Louw, Kappers, & Koenderink, 2002), and the perception of sexual interest (Farris, Viken, & Treat, 2010). Townsend has been at the forefront of this movement, coauthoring the article that introduced and named GRT and advancing the theory with many subsequent contributions.

Of course, the perceptual and cognitive processes modeled by GRT are mediated by circuits in the brain. During the past decade or two, much has been learned about the architecture and functioning of these circuits. This chapter reviews this neuroscience literature, with a focus on three separate questions. First, what does the neuroscience literature say about the validity of GRT? In other words, do the recent discoveries in neuroscience support or disconfirm the fundamental assumptions of GRT? Second, how can we use results from the neuroscience literature to improve GRT applications? For example, are there experimental conditions that improve the validity of GRT analyses? Third, how can GRT analyses be extended to neuroscience data, and especially to data from neuroimaging experiments?

## 1.2 SUPPORTING EVIDENCE FO THE NEURAL FEASIBILITY OF GRT

GRT is an extremely general model of perception and decision making that has been applied to a great variety of different tasks and behaviors. Nevertheless, it makes some core assumptions that are assumed to hold in all applications. In particular, it assumes two separate stages of processing – a sensory/perceptual stage that must generally precede a second decision stage. It also assumes that all sensory representations are inherently noisy, that every behavior, no matter how trivial, requires a decision, and that decision processes can be modeled via a decision bound. All but the last of these assumptions also define univariate signal detection theory.

When GRT was first proposed nearly 30 years ago, the only one of these assumptions with any independent support was that all sensory representations are noisy. The other assumptions were justified almost exclusively on the basis of intuitive appeal. During the intervening three decades however, an explosion of new neuroscience knowledge has provided strong tests of all GRT assumptions. This section reviews these assumptions and the relevant neuroscience data. As we will see, for the most part, neuroscience has solidified the foundations of GRT.

### 1.2.1 Separate Sensory and Decision Processes

GRT and signal detection theory both assume that decision processes act on a percept that may depend on the nature of the task, but does not depend on the actual response that is made. GRT can be applied to virtually any task. However, the most common applications are to tasks where the stimuli vary on two stimulus components or dimensions. Call these A and B. Then a common practice is to let $A_iB_j$ denote the stimulus in which component A is at level $i$ and component B is at level $j$. GRT models the sensory or perceptual effects of

stimulus $A_iB_j$ via the joint probability density function (pdf) $f_{ij}(x_1, x_2)$. On any particular trial when stimulus $A_iB_j$ is presented, GRT assumes that the subject's percept can be modeled as a random sample from this joint pdf, and that the subject uses these values to select a response. Thus, the pdf $f_{ij}(x_1, x_2)$ is the same on every trial that stimulus $A_iB_j$ is presented, regardless of what response was made, or in other words, GRT assumes that a sensory representation is formed first, and then decision processes use this representation to select a response. A possible neural implementation of the theory would require that the neural networks used to represent the perceptual properties of stimuli are relatively separated from the networks used to set criteria for decision making (which determine response biases).

Even 30 years ago, it was known that the flow of information from sensory receptors up through the brain, pass through sensory cortical regions before reaching motor areas that initiate behaviors. In the case of vision, for example, it was known that retinal ganglion cells project from the retina to the lateral geniculate nucleus of the thalamus, which projects to V1 (primary visual cortex), then V2, V4, and many other regions before the M1 (primary motor cortex) neurons are stimulated that cause the subject to press one response key or the other. So the general neuroanatomy seemed consistent with the GRT assumption of separate sensory and decision processes. Even so, there was virtually nothing known about whether visual cortex plays a significant role in the decision process. For example, the 1986 neuroanatomy was also consistent with a theory in which the response was actually selected as the representation moved up through higher levels of visual cortex, and that the main goal of the processing that occurred in many psychophysical tasks in later non-visual areas (e.g., premotor cortex and M1) was simply to serve as a relay between visual areas and the effectors that would execute the selected behavior. This type of intermingling would violate the GRT assumption that sensory/perceptual and decision processes are relatively separate.

In fact, the evidence is now good that decisions are not mediated within visual cortex. For a while, however, evidence against the GRT assumption of separate sensory and decision processes seemed strong. The most damning evidence came from reports of a variety of category-specific agnosias that result from lesions in inferotemporal cortex (IT) and other high-level visual areas. Category-specific agnosia refers to the ability to perceive or categorize most visual stimuli normally but a reduced ability to recognize exemplars from some specific category, such as inanimate objects (e.g., tools or fruits). The most widely known of such deficits, which occur with human faces (i.e., prosopagnosia), are associated with lesions to the fusiform gyrus in IT. In GRT, a category is defined by a response region, not by a perceptual distribution. So the association of category-specific agnosias to lesions in visual cortex seemed to suggest that the visual areas were also learning the decision bounds that defined the categories.

Of course, a category-specific agnosia that results from an IT lesion does not logically imply that category representations are stored in IT. For example, although such agnosias are consistent with the hypothesis that category learning occurs in IT, they are also generally consistent with the hypothesis that visually similar objects are represented in nearby areas of visual cortex. In particular, it is well known that neighboring neurons in IT tend to fire to similar stimuli.

Take the example of the most anterior IT region in the monkey brain: area TE. This is

the final stage of purely visual processing in the primate brain; thus, if high-level categorical representations were stored in visual cortex, TE would be a likely place for their storage. Research indicates that most neurons in this area are maximally activated by moderately complex shapes or object parts (for reviews, see Tanaka, 1996, 2004); that is, by features that are more complex than simple edges or textures, but not complex enough to represent a whole natural object or object category. Because neurons in TE are selective to partial object features, the representation of a whole object requires the combined activation of at least several of these neurons. In other words, anterior IT seems to code for objects in a sparse distributed manner (Rolls, 2009; E. Thomas, Van Hulle, & Vogels, 2001), which is confirmed by analyses showing that the way in which information about a stimulus increases with the number of IT neurons that are sampled is in line with a sparse distributed code (Abbott, Rolls, & Tovee, 1996; Hung, Kreiman, Poggio, & DiCarlo, 2005; Rolls, Treves, & Tovee, 1997). It appears that TE neurons that code for similar features cluster in columns (Fujita, Tanaka, Ito, & Cheng, 1992), that a single object activates neurons in several columns (Wang, Tanifuji, & Tanaka, 1998; Yamane, Tsunoda, Matsumoto, Phillips, & Tanifuji, 2006), and that the columns that are activated by two similar objects represent features that are common to both (Tsunoda, Yamane, Nishizaki, & Tanifuji, 2001). Thus, damage to some contiguous region of IT (or any other visual cortical area) is likely to lead to perception deficits within a class of similar stimuli, due to their shared perceptual features.

In fact, there is also now strong evidence that decision processes are not implemented within visual cortex. For example, Rolls, Judge, and Sanghera (1977) recorded from neurons in IT of monkeys. In these experiments, one visual stimulus was associated with reward and one with a mildly aversive taste. After training, the rewards were switched. Thus, in effect, the animals were taught two simple categories (i.e., "good" and "bad"), and then the category assignments were switched. If the categorical decision was represented in the visual cortex, then the firing properties of visual cortical neurons should have changed when the category memberships were switched. However, Rolls et al. found no change in the response of any of these cortical neurons, although other similar studies have found changes in the responses of neurons in other downstream brain areas (e.g., orbitofrontal cortex).

More recent studies have found similar null results with more traditional categorization tasks (Freedman, Riesenhuber, Poggio, & Miller, 2003; Op de Beeck, Wagemans, & Vogels, 2001; Sigala, 2004; E. Thomas et al., 2001; Vogels, 1999). In each of these studies, monkeys were taught to classify visual objects into one of two categories (e.g., tree versus non-tree, two categories of arbitrary complex shapes). Single-cell recordings showed that the firing properties of IT neurons did not change with learning. In particular, IT neurons showed sensitivity to specific visual images, but category training did not make them more likely to respond to other stimuli in the same category or less likely to respond to stimuli belonging to the contrasting category.

Similar results have been found in neurobiological studies of visual perceptual learning. The standard model of perceptual learning includes an early stage of sensory processing that is separate from a later stage of decision making (Amitay, Zhang, Jones, & Moore, 2014; Law & Gold, 2010). Theories proposing changes in the later decision stage of processing have been particularly successful in accounting for the available data (Amitay et al., 2014), including findings of heightened behavioral sensitivity that are not associated with changes in early sensory areas, but instead with the way that sensory information is used to form

a decision variable at later stages of processing (e.g., Kahnt, Grueschow, Speck, & Haynes, 2011; Law & Gold, 2008).

On the other hand, under certain conditions, categorization training can change the firing properties of IT neurons. Sigala and Logothetis (2002; see also De Baene, Ons, Wagemans, & Vogels, 2008; Sigala, 2004) trained two monkeys to classify faces into one of two categories and then in a separate condition to classify fish. In both conditions, some stimulus features were diagnostic and some were irrelevant to the categorization response. After categorization training, many neurons in IT showed enhanced sensitivity to the diagnostic features compared to the irrelevant features. Such changes are consistent with the widely held view that category learning is often associated with changes in the allocation of perceptual attention (Nosofsky, 1986). Accounting for such shifts in perceptual attention is straightforward in GRT. The typical approach is to assume that increases in the amount of attention allocated to a perceptual dimension reduces perceptual variance on that dimension (Maddox et al., 2002; Soto et al., 2014).

Changes in the selectivity of IT neurons after categorization training are consistent with the hypothesis that category learning is mediated outside the visual system, and that the attentional effects of categorization training are propagated back to visual areas through feedback projections (see Gilbert & Sigman, 2007; Kastner & Ungerleider, 2000). In support of this hypothesis, the effect of category learning on neural responses is stronger in non-visual areas, such as the striatum and prefrontal cortex (PFC), than in IT (De Baene et al., 2008; Seger & Miller, 2010). Simultaneous recordings from PFC and IT neurons during category learning show that, although IT neurons change their firing after learning, the changes are weaker than in PFC and several results suggest that the IT changes might be driven by PFC input (Freedman et al., 2003; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008).

Even so, the presence of feedback projections from higher-level areas (e.g., PFC) to visual cortex (and between visual areas, see Salin & Bullier, 1995) could be taken as evidence that decision processes might influence perceptual representations. As suggested earlier, however, it seems likely that most such influences could be explained by a GRT model that allows attentional modulation of perceptual noise. Furthermore, the fast behavioral and neural responses in visual tasks suggest that information conveyed by feedforward processing is often sufficient for object recognition and classification (e.g., Hung et al., 2005; Thorpe & Fabre-Thorpe, 2001). This supports a simple feedforward two-stage model of perceptual decision making.

Another prediction that follows from the GRT assumption of separate sensory and decision processes is that only decision-related processes should be immediately modifiable by variables such as verbal instruction and reward history. Whenever studies have tested this prediction, they have consistently found that decision tasks, such as setting a response criterion, predominantly recruit motor and frontal areas, rather than sensory association areas. For example, the development of decision biases in people is correlated with activity in the same areas that mediate reward processing during goal-directed action, including the ventral striatum (Chen, Jimura, White, Maddox, & Poldrack, 2015). Other examples come from studies examining the neural basis of the speed-accuracy tradeoff (e.g., Bogacz, Hu, Holmes, & Cohen, 2010; Heitz, 2014), which is usually attributed to changes in response criteria during perceptual decision making. High levels of activity in the striatum

and pre-supplementary motor area (preSMA) are found during decisions that are biased towards speed, both in people (Forstmann et al., 2008) and monkeys (Lauwereyns, Watanabe, Coe, & Hikosaka, 2002), and individual variability in levels of response caution are correlated with changes in striatal activity (Forstmann et al., 2008) and with the strength of white matter tracts between preSMA and the striatum (Forstmann et al., 2010). The development of perceptual representations in the visual system might include interactive processing, top-down influences and even categorical representations. Even so, the evidence clearly suggests that different processes are involved in criterion setting during decision making, as proposed by GRT. As indicated earlier, the evidence also clearly suggests that feedback-based category learning, which also occurs over short timescales, has substrates in associative areas such as the PFC and the basal ganglia.

In summary, we believe that the current evidence provides strong support for the GRT assumption of separate sensory and decision processes. Even so, the presence of feedback connections to visual areas (Salin & Bullier, 1995), top-down influences in vision (e.g., Kveraga, Ghuman, & Bar, 2007; Gilbert & Sigman, 2007), and category-specific agnosias suggests that this debate is far from resolved and therefore that much more research on this important issue is needed. In the interim, however, we believe that the best current evidence suggests that decisions about what response to make are not learned or formed in visual cortex. Instead, the evidence supports the GRT assumption that visual cortex builds the perceptual representation of the visual stimulus as activation moves up from V1 to IT. According to this view, IT could be seen as the terminal stage of the perceptual system. If so, then it follows that decision processes should be mediated by structures receiving direct projections from IT. There are three obvious candidates: the PFC, the medial temporal lobes (i.e., the hippocampal system), and the basal ganglia. As it happens, many relevant studies have examined each of these targets. A review of this large literature, however, is well beyond the scope of this chapter.

### 1.2.1 All Sensory Representations Are Noisy

A second fundamental assumption of GRT (and signal detection theory) is that there is trial-by-trial variability in the perceptual information obtained from every object or event (Ashby & Lee, 1993). The evidence supporting this assumption is overwhelming (for recent reviews, see Faisal, 2012; Faisal, Selen, & Wolpert, 2008), and was overwhelming even in 1986 when GRT was first proposed.

First, physical stimuli are intrinsically variable. For example, the number of photons emitted by a constant light source is approximately Poisson distributed (e.g., Geisler, 1989; Wyszecki & Stiles, 1982). In a Poisson distribution, the mean equals the variance, so the standard deviation of the number of photons reaching the cornea increases as the square root of the stimulus luminance. Therefore, intense stimuli are more variable than threshold-level stimuli.

Second, variability occurs after the stimulus enters the sensory system and before transduction (called perireceptor noise). For example, in vision, variable amounts of light are scattered or absorbed while it passes through the lens and the aqueous and vitreous humors. In fact, Barlow (1977) estimated that somewhere between 67% and 89% of the photons that strike the cornea are never absorbed by a photoreceptor.

Third, virtually all neurons in the brain exhibit spontaneous activity. For example, spontaneous isomerization of photopigment occurs frequently enough to be called "dark light" (Barlow, 1957) and retinal ganglion neurons can have spontaneous firing rates as high as 100 Hz (Robson, 1975).

Finally, there is also variability in the chemical events that occur at every neural synapse and across the neuronal membrane. There are many causes, including variability in the number of neurotransmitter molecules released presynaptically, variability in the time it takes neurotransmitter to diffuse across the synapse, variability in the number of neurotransmitter molecules that bind to a post-synaptic receptor, etc. Collectively, these stochastic processes produce significant variability in the postsynaptic response to identical presynaptic stimulation (e.g., Kleppe & Robinson, 2006). Furthermore, due to thermodynamic effects, voltage-gated ion channels open and close randomly, producing random variability in electrical currents (White, Rubinstein, & Kay, 2000). This "channel noise" might be one of the most important sources of biochemical noise at the neuronal membrane (Faisal, 2012).

One might propose that such randomness at the biophysical and biochemical levels is averaged out at the level of neurons and neural circuits. This is not necessarily true, as operations in neurons are highly non-linear and might act to amplify small effects, and key structures are small enough to be responsive to randomness at small scales (Faisal, 2012). Indeed, it has been shown that channel noise should affect the reliability of action potential propagation in thin axons, such as those found in mammalian cortex (Faisal & Laughlin, 2007).

All these sources of noise should cause us to expect variability in the timing of action potentials within and across trials, even under the same conditions of stimulation. Such variability has been observed (Shadlen & Newsome, 1998). Again, one might propose that the variability observed at the level of the responses of single neurons might average out at larger scales and not be relevant for behavioral performance, in contrast to the predictions of GRT. However, there is plenty of evidence showing that variability in the responses of sensory neurons is behaviorally relevant in perceptual discrimination tasks. For example, several studies have found a relation between the trial-by-trial variability in the activity of neurons and in the responses given by participants (e.g., Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Purushothaman & Bradley, 2005). These results suggest that the random activity from such neurons influences behavioral decisions (Gold & Ding, 2013).

Furthermore, noise in sensory representations can explain observed phenomena involving "discrimination" of identical stimuli. Perceptual learning has been found after training in a discrimination task involving identical stimuli (Amitay, Irwin, & Moore, 2006) and a signal-detection analysis showed that such effects can be explained as resulting from variability in internal sensory representations (Micheyl, McDermott, & Oxenham, 2009). Confirming this hypothesis, Amitay et al. (2013) found that identical auditory stimuli evoked different EEG responses as early as 100 ms after stimulus onset, depending on whether the stimuli were perceived to be same or different.

### 1.2.3 Every Task Requires a Decision

There is a wealth of evidence (reviewed by Gold & Shadlen, 2007) supporting the assump-

tion that even the simplest discrimination tasks recruit separate neural representations for sensory and decision processes. For example, in a vibrotactile frequency discrimination task, participants are exposed to a vibrating stimulus on their skin and are required to report what frequency of stimulation they perceived. Electrophysiological studies in monkeys show that neurons in somatosensory cortex respond in this task as if they encode a sensory representation, with increasing rate of responding as frequency of stimulation increases, but without reflecting the final decision (Hernández, Zainos, & Romo, 2000). On the other hand, neurons in PFC (e.g., Brody, Hernández, Zainos, & Romo, 2003) and premotor cortex (Romo, Hernández, & Zainos, 2004) respond in this task as if they encode the actual judgment leading to a behavioral decision.

Perhaps the most widely used task to study the role of decision processes in simple sensory discriminations is the random-dot motion (RDM) discrimination task. This task presents participants with a cloud of moving dots, some of them moving in a coherent direction and others moving in a random direction. Participants must report the direction of motion of the coherently moving dots, whose proportion is determined by the experimenter. Electrophysiological studies with monkeys have shown that neurons in medial temporal visual cortex (area MT) respond in this task as if they encode sensory evidence for the direction of motion. Their average response after stimulus presentation is stable, reflecting constant differences in firing rate as a function of motion coherence (Gold & Shadlen, 2007). Furthermore, the effect of microstimulation of MT neurons on behavioral choices is similar to what is observed by changes in the actual motion coherence of the stimulus (Hanks, Ditterich, & Shadlen, 2006).

On the other hand, neurons in a number of areas – including lateral intraparietal cortex (LIP), dorsolateral PFC (dlPFC), and superior colliculus – seem to respond as if they encode a decision variable that is closely related to the actual action ultimately chosen by the monkeys. The average response of these neurons to a motion stimulus increases with time from stimulus onset, and the speed at which firing ramps-up increases with the motion coherence of the stimulus. More importantly, the average firing reaches the same threshold for all levels of motion coherence right before the behavioral response is made (e.g., Roitman & Shadlen, 2002, and shows a steep drop afterwards). There is also evidence suggesting that sensory stimulation presented after this threshold has been reached is ignored in the final choice (Kiani, Hanks, & Shadlen, 2008). Finally, the effect of microstimulation of LIP neurons on behavioral choices is smaller than the effect of microstimulation of MT neurons, especially in choice accuracy. This is in line with the idea that MT encodes sensory evidence that is later accumulated by LIP (Hanks et al., 2006). For example, a simple model of these experiments is that microstimulation of MT neurons adds a constant to the sensory evidence, which means that the decision variable increases by this constant amount at every time step. In contrast, microstimulation of LIP neurons would add the same constant to the decision variable itself, and therefore not affect the temporal integration.

Signals resembling evidence accumulation have also been found in human experiments using fMRI (Heekeren, Marrett, Bandettini, & Ungerleider, 2004) and EEG (e.g., Kelly & O'Connell, 2013; O'Connell, Dockree, & Kelly, 2012; Van Vugt, Simen, Nystrom, Holmes, & Cohen, 2012). Note that the presence of such signals is not only evidence for the assumption that decision processes are implemented in the brain in even the simplest perceptual discrimination tasks, but also for the assumption, discussed in the previous section, of sep-

arate sensory and decision processes.

### 1.2.4 Decision Processes can be Modeled with Decision Bounds

There is now overwhelming evidence that humans have multiple learning systems that for the most part are neuroanatomically and functionally distinct (Ashby & Maddox, 2005; Eichenbaum & Cohen, 2001; Squire, 1992). Interestingly, the evidence is good that the default decision strategy of one of these systems can be modeled by decision bounds that satisfy decisional separability (DS), but the decision strategy of the other system likely to contribute to the learning of tasks where GRT is relevant is incompatible with decision bounds.

The most complete description of two of the most important learning systems is arguably provided by the COVIS theory (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby, Paul, & Maddox, 2011). COVIS assumes separate explicit reasoning and procedural-learning systems that compete for access to response production. The explicit system uses executive attention and working memory to select and test simple verbalizable hypotheses about which response is correct. The procedural system gradually associates responses with regions of perceptual space via reinforcement learning. COVIS assumes that explicit decision making is mediated by a broad neural network that includes the PFC, anterior cingulate, head of the caudate nucleus, and the hippocampus, whereas the key structures in the procedural-learning system are the striatum and premotor cortex.

COVIS has been applied most frequently to categorization tasks that are defined by a many-to-one stimulus-to-response mapping, whereas GRT is applied most frequently to identification tasks in which the stimulus-to-response mapping is one-to-one. Even so, COVIS has been applied to identification tasks (Ashby, Waldron, Lee, & Berkman, 2001), GRT is frequently applied to categorization tasks (Ashby & Maddox, 1994), and there are long-standing proposals that similar perceptual and decision processes mediate both identification and categorization (Nosofsky, 1986). For these reasons, evidence about decision processes that underlie COVIS should be directly applicable to GRT.

The evidence is good that the explicit system is constrained to make independent decisions about single stimulus dimensions and that these decisions can be combined in ways that follow the rules of Boolean algebra. Singling out a stimulus dimension requires selective attention – an ability that has been associated with dimensions that are perceptually separable (Ashby & Townsend, 1986; Garner, 1974; Lockhead, 1966; Maddox, 1992; Shepard, 1964). Thus, in any experiment that uses perceptually separable stimulus dimensions, DS is likely to hold – at least locally – if subjects perform the task explicitly. Later in this chapter, we describe experimental conditions that promote explicit processing.

Note that the qualifier "locally" plays a critical role in this hypothesis. If subjects use explicit strategies then the resulting decision bounds should be constructed from piecewise vertical and horizontal line segments, but this does not guarantee that DS will hold overall. For example, consider the GRT model of a hypothetical $2 \times 2$ identification experiment shown in Figure 1. Here, $a_i b_j$ denotes the response region associated with stimulus $A_i B_j$. Note that perceptual separability (PS) and DS hold on stimulus dimension $X_2$ but not on stimulus dimension $X_1$. Because the bounds are all constructed from vertical and horizontal line segments, this decision model is compatible with an explicit strategy (assuming selective
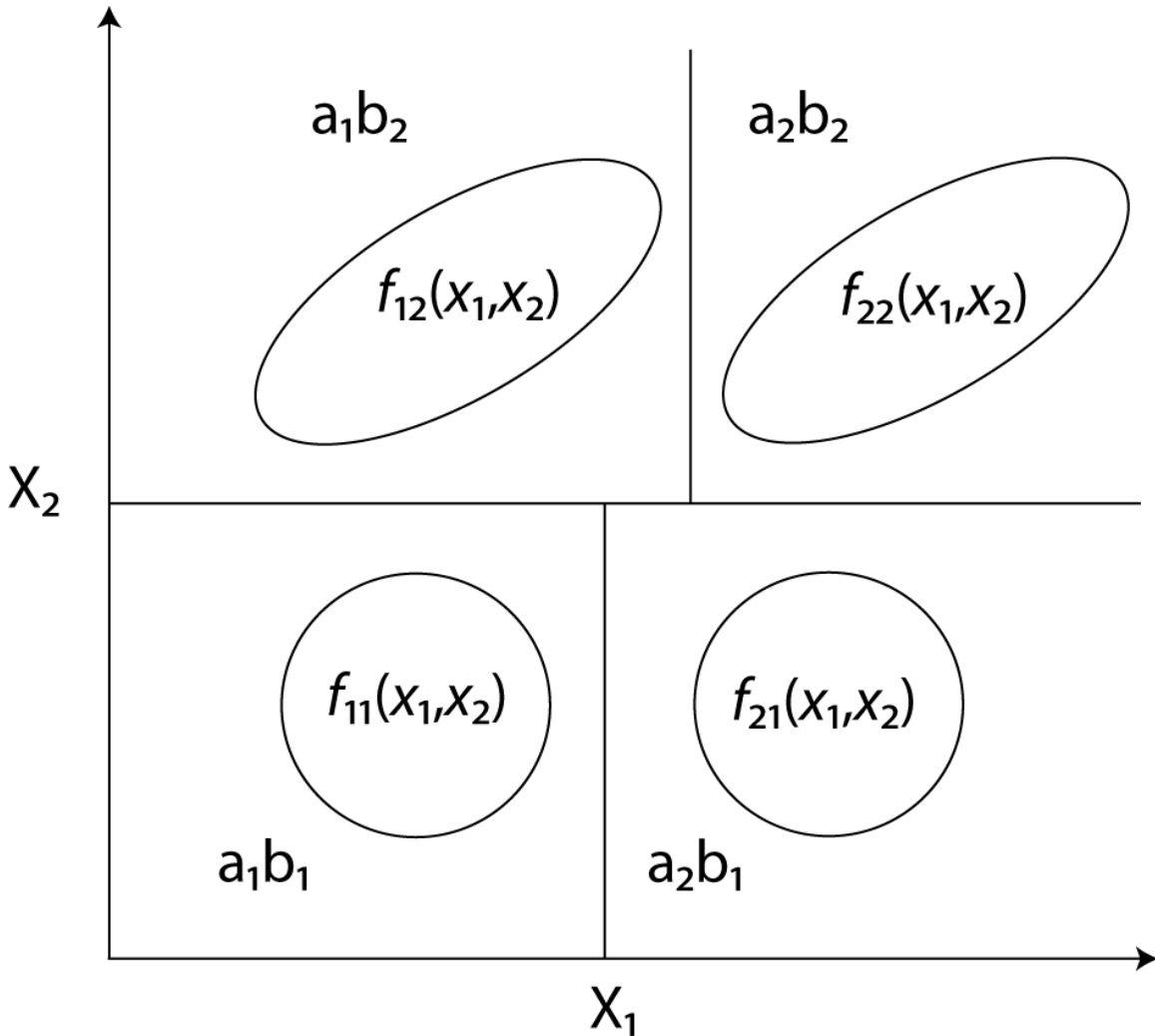
*Figure 1*. A GRT model that satisfies perceptual and decisional separability on dimension $X_2$ but not on dimension $X_1$.

attention to the stimulus dimensions is possible).

Selective attention is thought to depend on the PFC (e.g., Miller & Cohen, 2001. Thus, the ability to quickly form decision bounds that satisfy local DS might be unique to species with a well-developed PFC. Some evidence supports this prediction. First, of course, many studies have shown that healthy humans learn categories that can be separated by a one-dimensional decision bound (that satisfies DS) far more quickly than categories that can be separated by a diagonal bound, even when the two category structures are exactly equated on all category-separation statistics (e.g., Ashby & Maddox, 2005). Second, pigeons learn these two types of category structures equally well and at exactly the same rate (Smith et al., 2011). Although pigeons do possess brain regions thought to be analogous to PFC, such structures are not homologous, and it seems clear that pigeons lack at least some of the cognitive mechanisms recruited during rule-based learning (see Soto & Wasserman, 2014). Third, macaque and capuchin monkeys, who both have substantial PFCs, show similar

performance advantages as humans on the one-dimensional category structures. This result suggests that the human and monkey advantage for DS is probably not language-based, even though rules that satisfy DS are typically easy to describe verbally. Although much more data are needed to settle this issue, a tentative first hypothesis is that selective attention is the critical attribute that makes one-dimensional rules so easy to learn.

Ashby and Waldron (1999) tested whether or not the procedural system learns decision bounds. Subjects in this experiment learned two categories of stimuli that varied on two stimulus dimensions. The stimuli were sampled randomly from non-normal distributions. In the first experiment, the variance-covariance matrices from the two categories were identical, but the optimal category bound was quadratic, whereas in the second experiment the opposite was true – that is, the two variance-covariance matrices were different, but the optimal bound was linear. Previous research had shown that when the optimal bound was linear, people responded as if using linear bounds, and when the optimal bound was quadratic, people responded as if using quadratic bounds. Ashby and Waldron (1999) argued that the only information signaling which type of bound to use that could be reliably estimated within a few hundred learning trials was in the category variances and covariances. Thus, if subjects learned decision bounds, they should have responded suboptimally in both experiments – with a linear bound in the first experiment and a quadratic bound in the second. However, in both experiments, subjects responded optimally, suggesting that they were not using decision bounds.

More evidence suggesting that the human procedural system does not use decision bounds was reported by Casale, Roeder, and Ashby (2012). This article reported results from a series of experiments in which subjects were first trained on categories that were separated either by a bound that satisfied DS or by a diagonal bound. After subjects learned the categories, they were transferred to novel categories in which perfect performance was possible with the original training bound. If subjects had learned decision bounds during the original training, then transfer performance should be good because there was nothing new to learn. In fact, transfer was almost perfect in the DS conditions, but there was no evidence of any transfer in the diagonal bound conditions, suggesting that the procedural system does not learn a global decision rule, but instead only learns local stimulus-response associations.

Much evidence suggests that procedural learning is mediated largely within the striatum, and is facilitated by a dopamine (DA) mediated reinforcement learning signal (Ashby & Ennis, 2006; Badgaiyan, Fischman, & Alpert, 2007; Grafton, Hazeltine, & Ivry, 1995; Jackson & Houghton, 1995; Knopman & Nissen, 1991). The well-accepted theory is that positive feedback that follows successful behaviors increases phasic DA levels in the striatum, which has the effect of strengthening recently active synapses, whereas negative feedback causes DA levels to fall below baseline, which has the effect of weakening recently active synapses (Houk, Adams, & Barto, 1995). The critical site of learning within the striatum is thought to be at synapses between striatal medium spiny neurons and pyramidal neurons projecting from visual association areas of cortex (in the case of GRT experiments with visual stimuli). The idea is that each medium spiny neuron becomes associated with a motor goal (e.g., press the button on the left), so the procedural system learns associations between each presented stimulus (via activation of visual cortical neurons) and the relevant motor response (Ashby, Ennis, & Spiering, 2007; Ashby & Waldron, 1999). These cortical-striatal projections are

characterized by massive convergence – the ratio is somewhere around 10,000-to-1 (Wilson, 1995).

Based on these anatomical considerations, Ashby and Waldron (1999; Ashby et al., 2011) proposed the striatal pattern classifier (SPC) as a model of decision processes in the procedural learning system. The SPC assumes a high-resolution perceptual space that maps to a low-resolution decision space. Each perceptual unit is assumed to represent a different neuron (or group of neurons) in visual cortex and each decision unit is assumed to represent a different medium spiny neuron in the striatum. Typical applications might include 10,000 perceptual units and only a few decision units. Each perceptual unit has a different preferred stimulus, but presentation of a stimulus causes a graded pattern of perceptual activation because the tuning of each perceptual unit is modeled via a radial basis function. Each decision unit is associated with a response. Initially, the perceptual units are fully connected to the decision units, but reinforcement learning changes the strengths of these connections. Computationally, the SPC is similar to the covering map version of ALCOVE (Kruschke, 1992) and to Anderson's rational model (Anderson, 1991).

## 1.3 USING NEUROSCIENCE RESULTS TO IMPROVE GRT APPLICATIONS

The preceding sections suggest that the neuroscience literature provides considerable support for the assumptions that underlie GRT. The single caveat concerns the assumption that decision processes can be modeled with (piecewise) linear or quadratic decision bounds. We saw that the evidence is good that this assumption is valid when subjects are using PFC-mediated explicit strategies. In fact, under these conditions, bounds that satisfy local DS should be expected. However, if subjects use striatal-mediated procedural strategies then the decision bound assumption of GRT is problematic. Thus, one approach that could be used to maximize the validity of GRT is to adopt experimental methods that encourage subjects to use explicit decision strategies. If successful, an added bonus is that DS is likely to hold, at least locally. DS greatly simplifies data analyses, both because it makes numerical integration under bivariate normal distributions easier, but also because it allows stronger inferences about PS and independence (e.g., Ashby & Soto, 2015; Silbert & Thomas, 2013).

It is important to note that encouraging subjects to use decision strategies that satisfy DS when they might instead naturally use some procedural strategy could possibly affect the underlying perceptual representations. For example, Soto and Ashby (2015) reported that categorization training with a DS bound increased the PS of novel stimulus dimensions. The extent of such decisional/perceptual interactions is unknown and much more work is needed, but for the present purposes, several remarks are in order. First, the pre-training required to increase PS was fairly extensive, so changes in PS are unlikely in a one-session experiment (Soto & Ashby, 2015). Second, such training is unlikely to affect the PS of familiar stimulus dimensions, and third, many GRT analyses, especially those based on summary statistics, assume DS holds and therefore are invalid if subjects are using some procedural strategy. For these reasons, researchers might often want to encourage their subjects to use DS strategies.

In fact, a large literature establishes conditions that favor explicit decision strategies over procedural strategies. Critical features include the nature of the optimal decision bound, the instructions given the subjects, and the nature and timing of the feedback, to name just a

few (e.g., Ashby & Maddox, 2005, 2010). For example, Ashby, Waldron, Lee, and Berkman (2001) fit the full GRT identification model to data from two experiments. In both, 9 similar stimuli were constructed by factorially combining 3 levels of the same 2 stimulus components. Thus, in stimulus space, the 9 stimuli had the same $3 \times 3$ grid configuration in both experiments. In the first experiment however, subjects were shown this configuration beforehand and the response keypad had the same $3 \times 3$ grid as the stimuli. In the second experiment, the subjects were not told that the stimuli fell into a grid. Instead, the 9 stimuli were randomly assigned responses from the first 9 letters of the alphabet. In the first experiment, where subjects knew about the grid structure, the best-fitting GRT model assumed DS on both stimulus dimensions. In the second experiment, where subjects lacked this knowledge, the decision bounds of the best-fitting GRT model violated DS. Thus, one interpretation of these results is that the instructions biased subjects to use explicit strategies in the first experiment and procedural strategies in the second experiment.

Therefore, to encourage the use of explicit strategies, subjects should be told about the factorial nature of the stimuli and the response device should map onto this factorial structure in a natural way. Furthermore, several other design principles should be followed. First, the intertrial interval should be long enough so that subjects have sufficient time to process the meaning of the feedback. Maddox, Ashby, Ing, and Pickering (2004) showed that a short delay interferes with explicit but not procedural learning. If subjects are testing explicit hypotheses about the correct decision strategy then after error feedback for example, they might decide to reject their current hypothesis, and then select and implement some new strategy. These processes require time, attention, and effort. In contrast, evidence suggests that feedback processing in the procedural-learning system is essentially automatic.

Second, working memory demands should be minimized (e.g., avoid dual tasking) to ensure that working memory capacity is available for explicit hypothesis testing. This is critical because several studies have shown that a simultaneous dual task interferes with explicit learning much more than with procedural learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006).

## 1.4 EXTENDING GRT TO NEUROIMAGING DATA

The perceptual distributions and decision bounds in typical GRT applications are latent variables in the sense that they are hypothetical and unobservable. Modern neuroimaging technology however, makes it possible to observe brain activity with a precision that seemed impossible when GRT was first developed. For example, the newest fMRI scanners record an indirect measure of neural activation (i.e., the BOLD response) in as many as 500,000 separate brain locations at a temporal resolution of better than one second. The GRT perceptual and decision processes are presumed to provide summary descriptions of the neural activation that is mediating the behavior under study. Therefore, at least theoretically, fMRI might allow the possibility of observing this neural activation almost directly. Thus, an exciting area of future research is to extend GRT applications to neuroimaging data.

There are at least two qualitatively different ways that GRT could contribute to neuroimaging data analysis. One is that a variety of different GRT models could be used to make predictions about how neural activation should change across stimuli or conditions. The version that provides the best account of the neuroimaging data would then be

supported relative to its competitors. Within the neuroimaging literature, mathematical models used in this way are known as *encoding models* (e.g., Naselaris, Kay, Nishimoto, & Gallant, 2011). The encoding approach is similar to typical applications of GRT, except neuroimaging data are used to test the models as well as behavioral data. Unlike traditional behavioral applications however, multiple GRT analyses could be possible with neuroimaging data. For example, fMRI provides enough data that different GRT models could be fit to data from each separate neural region-of-interest (ROI). For example, suppose that the best fitting models satisfy PS in early visual cortical areas, but that in some higher-level visual association area, the best-fitting model violates PS. Such a discovery would greatly advance our understanding of human object perception because it could begin to address the question of why PS fails on behavioral data.

A second approach to a GRT-based analysis of neuroimaging data would be to extract the raw perceptual distributions and/or decision bounds directly from the neuroimaging data. Models used in this way are known as *decoding models* in the neuroimaging literature (e.g., Naselaris et al., 2011). Of course, decoding of this type is impossible if only behavioral data are used, but especially with fMRI, the extra observability makes decoding the GRT perceptual distributions, for example, a reasonable goal. One advantage of decoding methods is that they can be distribution free, and therefore make fewer assumptions than traditional GRT approaches. For example, it might be possible to estimate the GRT perceptual distributions using a nonparametric estimator. This would allow for example, nonparametric tests of PS. The next two sections describe encoding and decoding approaches to neuroimaging data analysis in more detail.

### 1.4.1 Neural Encoding Models

An encoding model is a formal representation of the relation between sensory stimuli and the response of a single neuron or a group of neurons. Because neuronal responses are inherently stochastic, the encoding problem requires specifying $p(r|s)$, or the probability that neuronal response $r$ occurs on trials when stimulus $s$ is presented (Pillow, 2007). An encoding model usually includes a set of parameters, which we denote via the vector $\boldsymbol{\theta}$, which summarize the information in the stimulus that drives the neural activity. These parameters are estimated from the data in such a way that $p(r|s, \boldsymbol{\theta})$ becomes a good approximation to $p(r|s)$. The data required to fit an encoding model are obtained from experiments presenting a variety of stimuli to participants (typical choices are white noise stimuli and natural images) and recording some measure of neural activity in response to the stimuli (e.g., the fMRI BOLD response). In the case of GRT, one approach might be to construct two different types of neuronal response models. In sensory models, $p(r|s, \boldsymbol{\theta})$ would be constructed from the GRT perceptual distribution associated with the presented stimulus, whereas in decision models, $p(r|s, \boldsymbol{\theta})$ would be constructed from the decision bounds (e.g., from the discriminant functions). One could then apply both model types to each ROI to determine whether the ROI was a sensory or decision region.

The standard encoding model is constructed from one or more channels, as shown in Figure 2. Each channel is sensitive to some information about the stimulus, which can be as simple as the luminance value in a single pixel (e.g., models of neuronal receptive fields in the early visual system; for a recent review, see Sharpee, 2013) or as complex
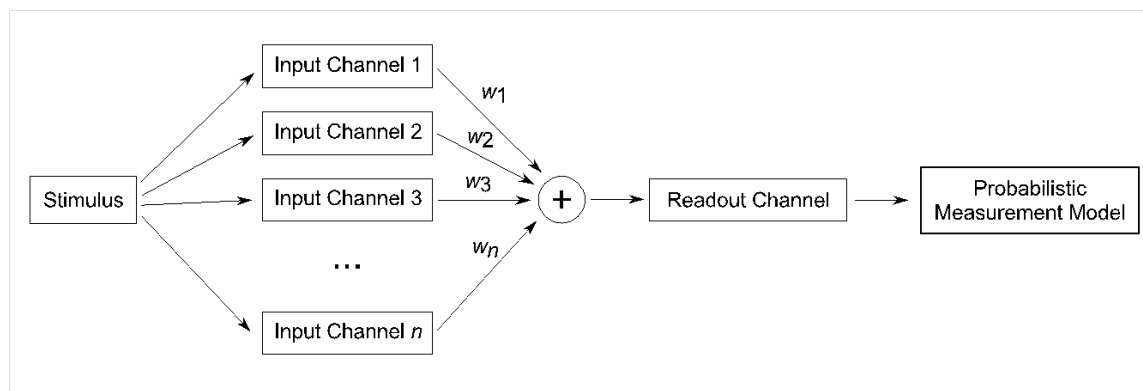
*Figure 2*. A schematic representation of a standard encoding model. Figure by F. Soto, shared under creative commons license at http://figshare.com/articles/Standard-encoding-model/1385405

as semantic information about an object presented in the stimulus (Huth, Nishimoto, Vu, & Gallant, 2012). Usually, the responses from several channels are pooled to compute the response of a readout channel in the next step of processing. This hierarchical scheme can be extended to include as many levels and channels as necessary, but the most common choice is the simple structure shown in Figure 2, where the response of several input channels is known and described by a fixed, often nonlinear transformation, and the response of a single readout channel is a linear function of the activity in the input channels. Models of this type are known as linearizing encoding models (Naselaris et al., 2011), because the free parameters (e.g., channel weights) are restricted to the linear readout portion of the model. For example, in an encoding model for a voxel located in primary visual cortex, each channel in Figure 2 could be replaced by a single Gabor wavelet. The response of the channel to an image is determined by filtering the image using the Gabot wavelet, which implements a nonlinear mapping from the stimulus to the channel's response (see Naselaris et al., 2011). In studies of object representation, it is possible to include input channels that encode for even more complex information, such as the semantic categories to which an object belongs (e.g., Huth et al., 2012; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009).

The contribution of each input channel to the response of the readout channel, represented by a weight $w$, is usually unknown. These weights and any other free parameters in the model must therefore be estimated from data using standard statistical procedures. To do this, a statistical model is included that makes it possible to compute the probability of a neural observation given the output of the readout channel, which describes the probability of observing a measure of neural activity – such as neural spikes, firing rate, local field potentials (LFPs), BOLD responses, etc. – as a function of the activity in the readout channel and possibly other parameters. The statistical model is sometimes explicitly stated, which allows maximum likelihood estimation of the model parameters (Pillow, 2007), but in many cases it is only implicitly assumed to be described by a normal distribution with mean equal to the activity of the output channel. This assumption of normality, which is common in models of fMRI and EEG data, is useful to estimate the value of the free parameters $w$ through linear regression (e.g., Garcia, Srinivasan, & Serences, 2013; Kay, 2012; Kay, Naselaris, Prenger, & Gallant, 2008).
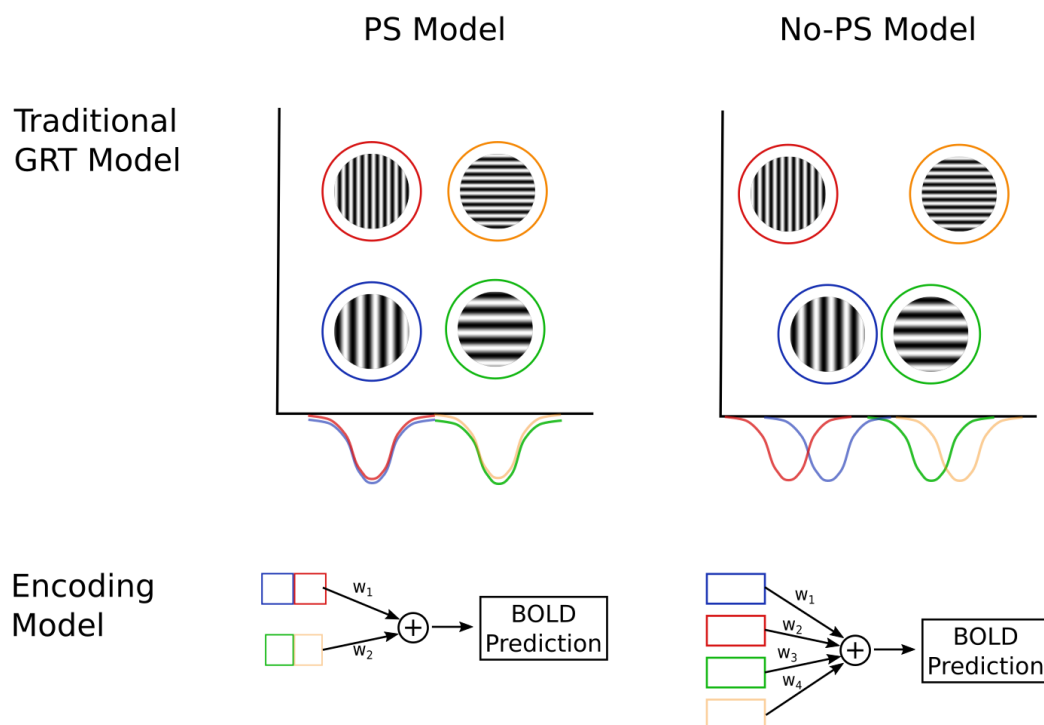
*Figure 3*.    GRT models implementing perceptual separability (PS Model, left col-
umn) and failure of perceptual separability (No-PS Model, right column).    The
top row represents traditional GRT models and the bottom row represents encod-
ing GRT models.    Figure by F. Soto, shared under creative commons license at
http://figshare.com/articles/GRT_encoding_models/1547914

A standard encoding model assumes that a small population of neurons encodes each
stimulus dimension (Pouget, Dayan, & Zemel, 2003). According to this view, a dimension
is represented by a group of neurons, each with a tuning function that determines its firing
rate to different stimuli, with peak firing at the value of the preferred stimulus and decreased
firing as stimuli become more dissimilar to the preferred stimulus.

In the Figure 2 model, each input channel represents one neuron or group of identical
neurons within this population code, with different mean firing rates to each of the stimuli
included in the task. Implementing a signal detection theory model using this formalism is
quite simple, only requiring the addition of noise in the neurons' firing rate. This is usually
modeled by a Gaussian distribution, making this a Gaussian population model (Ma, 2010),
but other distributions are possible. Thus, each one of the input channels in Figure 2 would
actually be a noisy channel, with responses influenced by some random component.

The role of the readout channel in this model is to compute a decision variable, which
can then be used as the basis for a final behavioral choice. Gold and Shadlen (2001) showed
that a simplified version of this framework (with only two neurons encoding each dimension)
is able to compute a log-likelihood ratio test for a binary perceptual discrimination, as in
traditional signal detection theory.
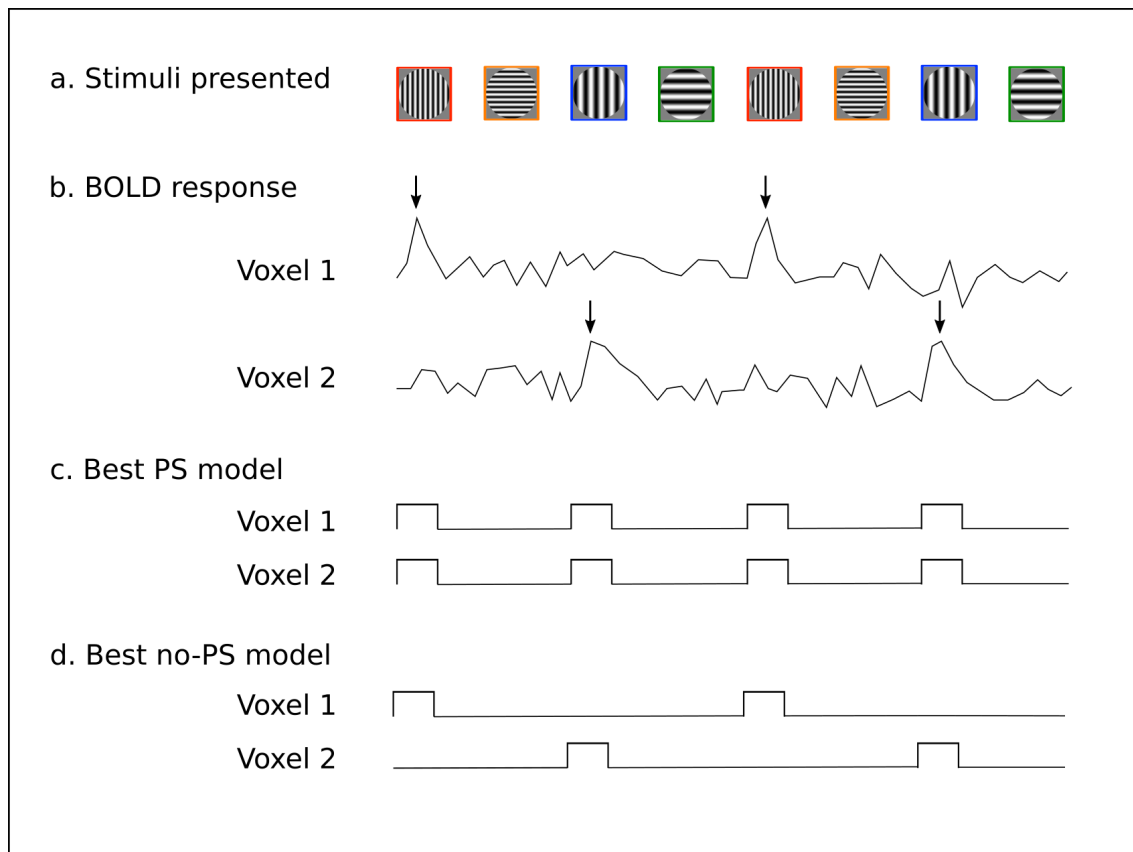
*Figure 4*. Hypothetical fMRI experiment testing perceptual separability of orientation from width; Panel (a) shows the sequence of stimuli shown and (b) shows the BOLD response measured in two voxels of interest; Panels (c) and (d) show the predictions of encoding models implementing perceptual separability and failure of perceptual separability, respectively. Figure by F. Soto, shared under creative commons license at http://figshare.com/articles/Fitting_data_from_an_fMRI_experiment_with_GRT _encoding_models/1547915

Although extending this framework to full GRT models will require substantial work, very simple modifications allow implementing at least some tests of GRT concepts. Imagine an experiment in which stimuli vary along two dimensions, each with only two possible values. A good example is sets of bars that vary in width (narrow versus wide) and orientation (vertical versus horizontal). The combination of possible values on both dimensions makes a total of four stimuli. The concept of PS from GRT is illustrated in the top row of Figure 3. Each axis in these models represents a particular stimulus dimension, each circle represents the bivariate distribution of perceptual effects produced by a specific stimulus, and the curves at the bottom represent the marginal distribution of perceptual effects along a specific dimension. To the left we see a GRT model in which PS holds for the orientation dimension. In this PS Model, stimuli that have the same orientation have equivalent marginal perceptual distributions on the *x*-axis. To the right we see a GRT model in which

PS does not hold for the orientation dimension. In this No-PS Model, stimuli that have the same orientation do not have equivalent marginal perceptual distributions. Note how the definition of PS in GRT refers only to marginal distributions of perceptual effects. Because these marginal distributions are unidimensional, they can be directly modeled using the probabilistic population coding scheme explained earlier.

An example of simple encoding models for PS and failure of PS is shown in the bottom row of Figure 3. To the left we see an encoding model implementing PS. In this model, there is a single channel that encodes both vertical bar stimuli, completely ignoring information about bar width. Similarly, there is a second single channel that encodes both horizontal bar stimuli and ignores information about bar width. The orientation dimension is thus encoded in this model separately from the width dimension: the representation of orientation is invariant to changes in width. The encoding model shown to the right implements a failure of PS. In this case, there are separate channels encoding wide vertical bars and narrow vertical bars. There are also separate channels encoding wide horizontal bars and narrow horizontal bars. The representation of orientation in this model is not separable from width, but it rather depends on width.

Remember that encoding models are fitted to neural data using traditional statistical procedures. A simple test of the separability of neural representations would require fitting a PS Model and a No-PS Model, and performing parameter estimation and model selection on those models. In our example, we could show participants the four oriented bar stimuli during fMRI scanning. Imagine that we are particularly interested in the representation of orientation in primary visual cortex. Figure 4a shows a sequence of stimuli that could be shown to participants in such a hypothetical experiment. Assume, for dydactical purposes, that the representation of orientation in primary visual cortex is not separable from the representation of width, with neurons in different voxels encoding the same orientation for different widths. This could produce BOLD responses like those observed in Figure 4b. Voxel 1 is clearly responding only to vertical bars that are thin, whereas voxel 2 is responding only to vertical bars that are wide. If both the PS Model and No-PS Model are fitted to these data, then it is likely that the No-PS model would offer a better fit, as this model could explain the data in voxel 1 by having a strong weight for the channel encoding the combination of vertical and thin and low weights for all other channels, and it could explain the data in voxel 2 by having a strong weight for the channel encoding the combination of vertical and thick and low weights for all other channels. The predictions of this model could be similar to those observed in Figure 4d. On the other hand, the PS Model would in both cases assign a strong weight to the channel that encodes "vertical" and a low weight to the channel that encodes "horizontal." As shown in Figure 4c, the model would correctly predict the responses of both voxels in half of the trials, but it would also incorrectly predict responses in the other half of the trials in which no response is observed. Model comparison should thus conclude that the No-PS model provided a better fit to the data.

Throughout this example, we assumed the simplest possible encoding model in which each channel represents either the presence or absence of a feature. Such a model should help to identify violations of PS such as that shown in the top-right of Figure 3, in which the means of marginal distributions are displaced, but it would not identify all possible violations of PS. Better encoding models could be developed if we could get estimates of

the perceptual effects of a stimulus on each trial. One way to do this is by fitting GRT models to response time data (Maddox & Ashby, 1996) and inverting the model to obtain estimates of perceptual effects (under the assumption of DS; for an introduction to these basic concepts from GRT see Ashby & Soto, 2015).
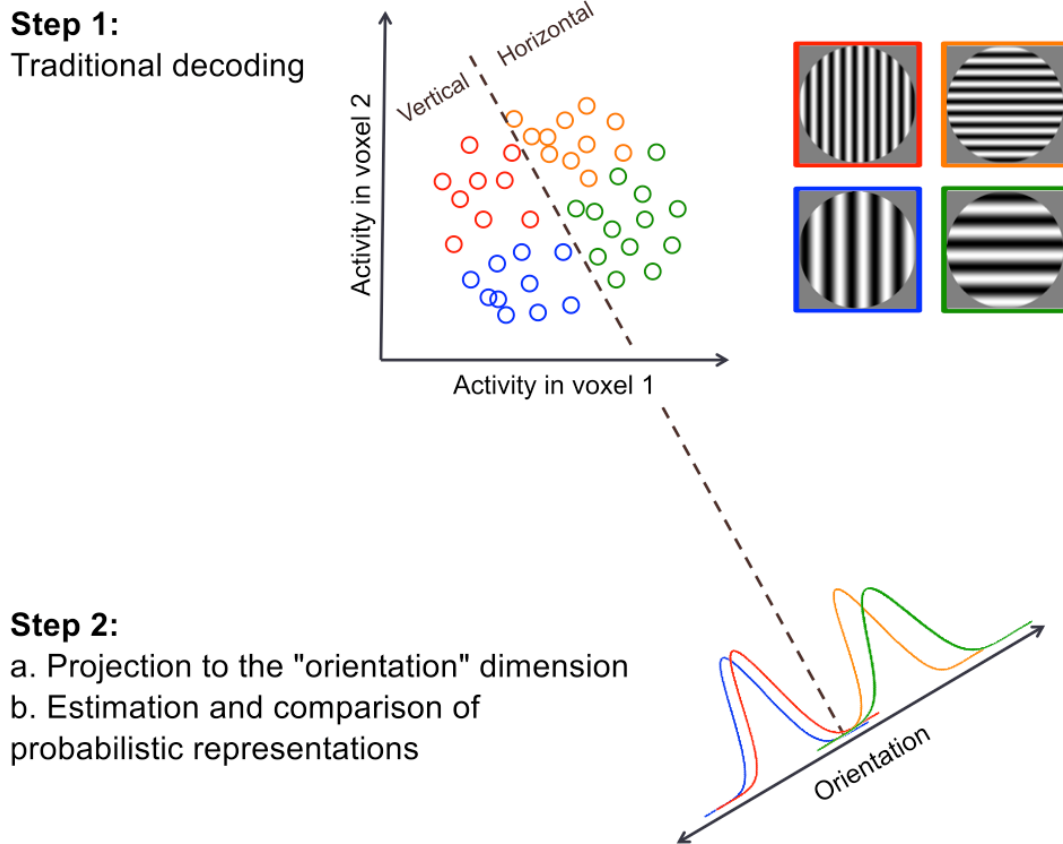
### 1.4.2 Neural Decoding

The term neural decoding is used in the computational neuroscience literature to refer both to a series of methods used by researchers to extract information about the task from neural data (Naselaris et al., 2011; Quiroga & Panzeri, 2009) and to the mechanisms used by readout neurons to extract similar information, which is later used for decision making and other cognitive processes (Gold & Ding, 2013; Pouget et al., 2003). Referring back to Figure 2, we can think of the input channels as implicitly encoding a particular variable, which is then decoded by the readout channel.

Researchers have decoded many forms of information from neural data, ranging from abstract properties, such as an object's identity or its membership to a particular class (e.g., Hung et al., 2005; Kriegeskorte, Formisano, Sorger, & Goebel, 2007; Li, Cox, Zoccolan, & DiCarlo, 2009; Said, Moore, Engell, Todorov, & Haxby, 2010; Soto, Waldschmidt, Helie, & Ashby, 2013), to the more concrete reconstruction of the actual images presented during an experiment (e.g., Miyawaki et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011; Thirion et al., 2006). The methods used to extract information are also highly variable, and include a variety of statistical learning algorithms (for a review, see Pereira, Mitchell, & Botvinick, 2009) and information theoretic measures (see Quiroga & Panzeri, 2009). Simple linear classification and regression algorithms are the most commonly used decoding methods, a practice that rests on the assumption that the mapping from observed neural activity to stimulus information is linear (Naselaris et al., 2011), in the same fashion as in the encoding models reviewed in the previous section. A theoretical justification for this choice is that real neurons integrate information from upstream neurons through a mechanism akin to a weighted sum with a threshold, in a similar manner to a linear classifier (Li et al., 2009). In contrast, a nonlinear classifier could produce successful classification from activity in a region that does not explicitly encode the features of interest (e.g., the retina; Naselaris et al., 2011). Furthermore, in practice it is usually the case that the decoding accuracy of nonlinear classifiers does not improve substantially over that of linear classifiers (e.g., Hung et al., 2005; Pereira et al., 2009).

Neural decoding using linear models is similar in goal to encoding models that represent information about a single stimulus dimension (Kay, 2012). In both cases, the goal is to show that a particular stimulus dimension is encoded by a neuronal population. Whereas an encoding model uses the dimension of interest to make predictions of neural responses, a linear decoding model uses the measured activity to recover the dimension of interest, and fitting the model amounts to finding a direction in the activity space that can best describe the differences across stimuli on the dimension of interest.

This suggests a method, which is schematized in Figure 5, to extend neural decoding procedures to test for separability of neural representations. To understand how this method works, let us return to the example of the hypothetical experiment described in the previous section, in which participants are scanned in the MRI while being presented with a series of

**Step 1:**
Traditional decoding

Activity in voxel 2

Vertical | Horizontal

Activity in voxel 1

**Step 2:**
a. Projection to the "orientation" dimension
b. Estimation and comparison of
probabilistic representations

Orientation

*Figure 5*. A schematic representation of a test of separability of neural representations, implemented as an extension to traditional linear decoding procedures. Figure by F. A. Soto, shared under creative commons license at http://figshare.com/articles/Test-of-separability-of-neural-representations/1385406

oriented bar stimuli (see Figure 4a), with the goal of understanding whether the representation of orientation is separable from the representation of width in primary visual cortex. Assume that we get estimates of neural activity on each trial from two voxels located in our region of interest. Such estimates can be obtained, for example, by simply estimating the mean BOLD response 4 or 6 s after stimulus onset, or by some other, more sophisticated method (Mumford, Turner, Ashby, & Poldrack, 2012; Turner, Mumford, Poldrack, & Ashby, 2012). The estimates of activity could then be represented in a two-dimensional voxel space, as shown at the top of Figure 5.

Each point here represents activity on a different trial, with different colors representing different stimuli that have been repeatedly presented across the experiment. If our goal was to decode bar orientation from these two voxels, a traditional decoding analysis would use a linear classifier to find the line in this space that best separates trials on which vertical bars were shown from trials on which horizontal bars were shown. This classification bound is represented by the dotted line in Figure 5. The line orthogonal to the classification

bound represents the direction in voxel space that best discriminates one orientation from the other (that is, the line along which the points corresponding to vertical bars are maximally separated from the points corresponding to horizontal bars). Thus, it is reasonable to assume that this is the direction in this specific voxel space along which orientation is encoded. If we take all the observed data points and project them onto this "orientation" dimension, the two distributions of points for a given value of orientation (e.g., "vertical"), each corresponding to a different value of width, can be compared to one another. Under the hypothesis of PS, these two sets of points should come from the same marginal distribution. There are a variety of statistical tests in the literature that allow a test of this hypothesis, with the most popular being the two-sample Kolmogorov-Smirnov test, but with more statistically powerful alternatives also available. We have recently used such a test to study separability of the neural representations of identity and emotional expression of human faces in an fMRI study (Soto, Vucovich, & Ashby, in preparation).

## 1.5 CONCLUSIONS

GRT has been exceptionally successful. But to date, virtually all applications have been purely behavioral. When GRT was first developed in 1986, this behavioral focus was necessary because little was known about the architecture and functioning of the neural circuits that implement the perceptual and decision processes hypothesized by GRT. Thirty years later however, this picture has changed dramatically and now many neuroscience results speak to the validity of GRT and suggest ways to improve its application. In addition, new neuroimaging technologies have created massive and complex data sets that beg for sophisticated new quantitative analyses. FMRI data analysis, for example, has been dominated by the general linear model – a statistical method that was never intended as a model of psychological process. GRT, with its extensive validation as a general model of perceptual and decision processes, therefore offers promising new tools for neuroimaging data analysis.

<div style="text-align:center">References</div>

Abbott, L., Rolls, E. T., & Tovee, M. J. (1996). Representational capacity of face coding in monkeys. *Cerebral Cortex*, *6*(3), 498–505.

Amazeen, E. L., & DaSilva, F. (2005). Psychophysical test for the independence of perception and action. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 170–182.

Amitay, S., Guiraud, J., Sohoglu, E., Zobay, O., Edmonds, B. A., Zhang, Y. X., & Moore, D. R. (2013, July). Human decision making based on variations in internal noise: an EEG study. *PLoS ONE*, *8*(7), e68928. doi: 10.1371/journal.pone.0068928

Amitay, S., Irwin, A., & Moore, D. R. (2006, November). Discrimination learning induced by training with identical stimuli. *Nature neuroscience*, *9*(11), 1446–1448. doi: 10.1038/nn1787

Amitay, S., Zhang, Y. X., Jones, P. R., & Moore, D. R. (2014). Perceptual learning: Top to bottom. *Vision Research*, *99*, 69–77. doi: 10.1016/j.visres.2013.11.006

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*(3), 632–656.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multi-dimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. *Advances in psychology*, *99*, 369–399.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, *38*(4), 423–466.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.

Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). New York: Cambridge University Press.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124–150.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, *103*(1), 165–192.

Ashby, F. G., & Soto, F. A. (2015). Multidimensional signal detection theory. In J. Busemeyer, J. T. Townsend, Z. J. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 13–34). New York: Oxford University Press.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154–179.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363–378.

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130*(1), 77–96.

Badgaiyan, R. D., Fischman, A. J., & Alpert, N. M. (2007). Striatal dopamine release in sequential learning. *Neuroimage*, *38*, 549–556.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*(4), 267–273.

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *The Journal of Physiology*, *136*(3), 469–488.

Barlow, H. B. (1977). Retinal and central factors in human vision limited by noise. In H. B. Barlow & P. Fatt (Eds.), *Vertebrate photoreception* (p. 337-358). New York: Academic Press.

Blaha, L. M., Silbert, N. H., & Townsend, J. T. (2011). A general recognition theory study of race adaptation. *Proceedings of Fechner Day*, *27*(1), 95–100.

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010, May). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology*, *63*(5), 863–891. doi: 10.1080/17470210903091643

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*(01), 87–100.

Brody, C. D., Hernández, A., Zainos, A., & Romo, R. (2003). Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral Cortex*, *13*(11), 1196–1207.

Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition*, *40*(3), 434–449.

Chen, M.-Y., Jimura, K., White, C. N., Maddox, W. T., & Poldrack, R. A. (2015). Multiple brain networks contribute to the acquisition of bias in perceptual decision-making. *Frontiers in Neuroscience*, *9*, 1–13.

Cohen, D. J. (1997). Visual detection and perceptual independence: Assessing color and form. *Perception & Psychophysics*, *59*(4), 623–635.

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, *15*(9), 717–727.

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*(3), 292–303.

Demeyer, M., Zaenen, P., & Wagemans, J. (2007). Low-level correlations between object properties and viewpoint can cause viewpoint-dependent object recognition. *Spatial Vision*, *20*(1), 79–106.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain.* Oxford University Press.

Faisal, A. A. (2012). Noise in Neurons and Other Constraints. In N. L. Novère (Ed.), *Computational Systems Neurobiology* (pp. 227–257). Dordrecht, Netherlands: Springer Netherlands.

Faisal, A. A., & Laughlin, S. B. (2007, May). Stochastic simulations on the reliability of action potential propagation in thin axons. *PLoS Computational Biology*, *3*(5), e79. doi: 10.1371/journal.pcbi.0030079

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008, April). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. doi: 10.1038/nrn2258

Farris, C., Viken, R. J., & Treat, T. A. (2010). Perceived association between diagnostic and non-diagnostic cues of women's sexual interest: General recognition theory predictors of risk for sexual coercion. *Journal of Mathematical Psychology*, *54*(1), 137–149.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E. J., . . . Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*, *107*(36), 15916–15920.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*(45), 17538–17542.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, *23*(12), 5235–5246.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*(6402), 343–346.

Garcia, J. O., Srinivasan, R., & Serences, J. T. (2013). Near-real-time feature-selective modulations in human cortex. *Current Biology*, *23*(6), 515–522.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314.

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*(5), 677–696.

Giordano, B. L., Visell, Y., Yao, H.-Y., Hayward, V., Cooperstock, J. R., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditionsa). *The Journal of the Acoustical Society of America*, *131*(5), 4002–4012.

Gold, J. I., & Ding, L. (2013, April). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in Neurobiology*, *103*, 98–114. doi: 10.1016/j.pneurobio.2012.05.008

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*(1), 10–16.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, *7*, 497–510.

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006, May). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience*, *9*(5), 682–689. doi: 10.1038/nn1683

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004, October). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*(7010), 859–862. doi: 10.1038/nature02966

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150. doi: 10.3389/fnins.2014.00150

Hernández, A., Zainos, A., & Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proceedings of the National Academy of Sciences*, *97*(11), 6191–6196.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012, December). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224. doi: 10.1016/j.neuron.2012.10.014

Jackson, S., & Houghton, G. (1995). Sensorimotor selection and the basal ganglia: A neural network model. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 337–368). Cambridge, MA: MIT Press.

Kahnt, T., Grueschow, M., Speck, O., & Haynes, J. D. (2011). Perceptual learning and decision-making in human medial frontal cortex. *Neuron*, *70*(3), 549–559.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*(1), 315–341.

Kay, K. N. (2012). Understanding visual representation by developing receptive-field models. In N. Kriegeskorte & G. Kreiman (Eds.), *Visual population codes: Towards a common multivariate framework for cell recording and functional imaging* (p. 133-161). Cambridge, MA: MIT Press.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008, March). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. doi: 10.1038/nature06713

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, *33*(50), 19434–19441.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*(12), 3017–3029.

Kleppe, I. C., & Robinson, H. P. C. (2006). Correlation entropy of synaptic input-output dynamics. *Physical Review E*, *74*(4), 041909.

Knopman, D., & Nissen, M. J. (1991). Procedural learning is impaired in huntington's disease: Evidence from the serial reaction time task. *Neuropsychologia*, *29*(3), 245–254. doi: http://dx.doi.org/10.1016/0028-3932(91)90085-M

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy*

*of Sciences*, *104*(51), 20600–20605.

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145–168.

Lauwereyns, J., Watanabe, K., Coe, B., & Hikosaka, O. (2002, July). A neural correlate of response bias in monkey caudate nucleus. *Nature*, *418*(6896), 413–417. doi: 10.1038/nature00892

Law, C. T., & Gold, J. I. (2008, April). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, *11*(4), 505–513. doi: 10.1038/nn2070

Law, C. T., & Gold, J. I. (2010). Shared mechanisms of perceptual learning and decision making. *Topics in Cognitive Science*, *2*(2), 226–238.

Li, N., Cox, D. D., Zoccolan, D., & DiCarlo, J. J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? *Journal of Neurophysiology*, *102*(1), 360–376.

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, *72*(1), 94–104.

Louw, S., Kappers, A. M., & Koenderink, J. J. (2002). Haptic discrimination of stimuli varying in amplitude and width. *Experimental Brain Research*, *146*(1), 32–37.

Ma, W. J. (2010, October). Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research*, *50*(22), 2308–2319. doi: 10.1016/j.visres.2010.08.035

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification–speeded classification relationship. *Journal of Experimental Psychology: Human perception and performance*, *22*(4), 795.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*(4), 582–591.

Maddox, W. T., Ashby, F. G., & Waldron, E. M. (2002). Multiple attention systems in perceptual categorization. *Memory & Cognition*, *30*(3), 325–339.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419.

Micheyl, C., McDermott, J. H., & Oxenham, A. J. (2009, January). Sensory noise explains auditory frequency discrimination learning induced by training with identical stimuli. *Perception & Psychophysics*, *71*(1), 5–7. doi: 10.3758/APP.71.1.5

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., . . . Kamitani, Y. (2008, December). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, *60*(5), 915–929. doi: 10.1016/j.neuron.2008.11.004

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*, 2636–2643.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage*, *56*(2), 400–410.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–915.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, *21*(19), 1641–1646.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*(12), 1244–1252.

Pereira, F., Mitchell, T., & Botvinick, M. (2009, March). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, *45*(1), S199–S209. doi: 16/j.neuroimage.2008.11.007

Pillow, J. (2007). Likelihood-based approaches to modeling the neural code. In K. Doya, S. Ishii, A. Pouget, & R. P. N. Rao (Eds.), *Bayesian brain: Probabilistic approaches to neural coding* (pp. 53–70). Cambridge, MA: MIT Press.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*(1), 381–410.

Purushothaman, G., & Bradley, D. C. (2005, January). Neural population code for fine perceptual decisions in area MT. *Nature Neuroscience*, *8*(1), 99–106. doi: 10.1038/nn1373

Quiroga, R. Q., & Panzeri, S. (2009, March). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, *10*(3), 173–185. doi: 10.1038/nrn2578

Robson, J. G. (1975). Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. *Handbook of Perception*, *5*, 81–116.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, *22*(21), 9475–9489.

Rolls, E. T. (2009). The neurophysiology and computational mechanisms of object representation. In S. Dickinson, M. J. Tarr, A. Leonardis, & B. Schiele (Eds.), *Object categorization: Computer and human vision perspectives.*

Rolls, E. T., Treves, A., & Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research*, *114*(1), 149–162.

Romo, R., Hernández, A., & Zainos, A. (2004). Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, *41*(1), 165–173.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. *Psychological Review*, *111*(3), 588–616.

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010, May). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*(5). doi: 10.1167/10.5.11

Salin, P.-A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Reviews*, *75*(1), 107–154.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219.

Shadlen, M. N., & Newsome, W. T. (1998, May). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of Neuroscience*, *18*(10), 3870–3896.

Sharpee, T. O. (2013). Computational identification of receptive fields. *Annual Review of Neuroscience*, *36*, 103–120.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87.

Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behavioural Brain Research*, *149*(1), 1–7.

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320.

Silbert, N. H. (2012). Syllable structure and integration of voicing and manner of articulation information in labial consonant identification. *The Journal of the Acoustical Society of America*, *131*(5), 4076–4086.

Silbert, N. H., & Thomas, R. D. (2013). Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review*, *20*(1), 1–20.

Silbert, N. H., Townsend, J. T., & Lentz, J. J. (2009). Independence and separability in the perception of complex nonspeech sounds. *Attention, Perception, & Psychophysics*, *71*(8), 1900–1915.

Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, *18*(2), 414–421.

Soto, F. A., & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, *139*, 105–129.

Soto, F. A., Vucovich, L., & Ashby, F. G. (in preparation). Testing the independence of neural representations of face identity and expression through multidimensional signal detection theory.

Soto, F. A., Vucovich, L., Musgrave, R., & Ashby, F. G. (2014). General recognition theory with individual differences: a new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review*, *22*(1), 88–111.

Soto, F. A., Waldschmidt, J. G., Helie, S., & Ashby, F. G. (2013). Brain activity across the development of automatic categorization: A comparison of categorization tasks using multi-voxel pattern analysis. *Neuroimage*, *71*, 284–897. doi: 10.1016/j.neuroimage.2013.01.008

Soto, F. A., & Wasserman, E. A. (2014). Mechanisms of object recognition: What we have learned from pigeons. *Frontiers in Neural Circuits*, *8*.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*(3), 232–243.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*(1), 109–139.

Tanaka, K. (2004). Inferotemporal response properties. *The visual neurosciences*, 1151.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J. B., Lebihan, D., & Dehaene, S. (2006, December). Inverse retinotopy: Inferring the visual content of images from brain activation patterns. *Neuroimage*, *33*(4), 1104–1116. doi: 10.1016/j.neuroimage.2006.06.062

Thomas, E., Van Hulle, M. M., & Vogels, R. (2001). Encoding of categories by noncategory-specific neurons in the inferior temporal cortex. *Journal of Cognitive Neuroscience*, *13*(2), 190–200.

Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, *63*(4), 625–650.

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science*, *291*, 260–262.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, *4*(8), 832–838.

Turner, B., Mumford, J., Poldrack, R., & Ashby, F. (2012). Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*, *62*, 1429–1438.

Van Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in Neuroscience*, *6*, 106.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. part 2: single-cell study. *European Journal of Neuroscience*, *11*(4), 1239–1255.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168–176.

Wang, G., Tanifuji, M., & Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neuroscience Research*, *32*(1), 33–46.

Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 872–892.

White, J. A., Rubinstein, J. T., & Kay, A. R. (2000). Channel noise in neurons. *Trends in neurosciences*, *23*(3), 131–137.

Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 29–50). Cambridge, MA: Bradford.

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulas.* New York: Wiley.

Yamane, Y., Tsunoda, K., Matsumoto, M., Phillips, A. N., & Tanifuji, M. (2006). Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *Journal of Neurophysiology*, *96*(6), 3147–3156.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398.