



Mapping spikes to sensations

Maik C. Stüttgen^{1*}, Cornelius Schwarz^{2,3} and Frank Jäkel⁴

¹ Department of Biopsychology, University of Bochum, Bochum, Germany

² Systems Neurophysiology, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany

³ Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany

⁴ Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Edited by:

Tobias Kalenscher, Heinrich-Heine University Duesseldorf, Germany

Reviewed by:

Floris P. De Lange, Radboud University Nijmegen, Netherlands
Matthijs (Matt) Van Der Meer, University of Waterloo, Canada

*Correspondence:

Maik C. Stüttgen, Department of Biopsychology, University of Bochum, Bochum, Germany.
e-mail: maik.stuettggen@rub.de

Single-unit recordings conducted during perceptual decision-making tasks have yielded tremendous insights into the neural coding of sensory stimuli. In such experiments, detection or discrimination behavior (the psychometric data) is observed in parallel with spike trains in sensory neurons (the neurometric data). Frequently, candidate neural codes for information read-out are pitted against each other by transforming the neurometric data in some way and asking which code's performance most closely approximates the psychometric performance. The code that matches the psychometric performance best is retained as a viable candidate and the others are rejected. In following this strategy, psychometric data is often considered to provide an unbiased measure of perceptual sensitivity. It is rarely acknowledged that psychometric data result from a complex interplay of sensory and non-sensory processes and that neglect of these processes may result in misestimating psychophysical sensitivity. This again may lead to erroneous conclusions regarding the adequacy of candidate neural codes. In this review, we first discuss requirements on the neural data for a subsequent neurometric-psychometric comparison. We then focus on different psychophysical tasks for the assessment of detection and discrimination performance and the cognitive processes that may underlie their execution. We discuss further factors that may compromise psychometric performance and how they can be detected or avoided. We believe that these considerations point to shortcomings in our understanding of the processes underlying perceptual decisions, and therefore offer potential for future research.

Keywords: psychophysics, perception, signal detection theory, psychometric, neurometric, receiver operating characteristic, psychophysical task, single-unit electrophysiology

INTRODUCTION

Gustav Theodor Fechner is best known as the founding father of psychophysics. It is perhaps less well known that Fechner distinguished between what he called “outer psychophysics”, the relationship between physical stimuli and sensation, and “inner psychophysics”, the relationship between (neuro-) physiological activity and sensation. While being successful at making outer psychophysics the cornerstone of the evolving science of psychology, physiological methods at his time were not developed enough to allow direct investigation of inner psychophysics, and Fechner was well aware of this limitation (Fechner, 1860; Baird and Noma, 1978).

This situation has changed dramatically in the meantime, mainly with the advent of the awake behaving monkey preparation (Evarts, 1966), which allows for the simultaneous assessment of psychophysical measurements (psychometric data, e.g., percent correct responses) and spikes from (mostly cortical) single neurons in sensory areas of the brain (neurometric data; e.g., Newsome et al., 1989; Mountcastle et al., 1990; Vogels and Orban, 1990). These seminal studies, as well as a multitude of studies published since then, have centered around neurometric-psychometric (NP) comparisons in the sense that some measure of performance quality (such as a detection threshold or a difference limen) is extracted

from the neuro- and the psychometric data for direct comparison on the same scale (reviewed in Parker and Newsome, 1998). In concert with the application of signal detection theory (SDT, Green and Swets, 1966) to neurometric data, these studies have provided striking evidence that the stimulus detection and discrimination capacity of single sensory neurons can be close to or even exceed the capacity of the entire organism. These findings are in agreement with Barlow's (1961) notion of redundancy reduction: Barlow postulated that the neuronal representation of stimulus information (i.e., the representation relevant for Fechner's inner psychophysics) must be efficient (Barlow, 1961, 1972); in other words, as few spikes as possible in as few neurons as possible should be used to encode a sensory stimulus, a tenet quite different to Sherrington's (1940) idea of the brain as “a million-fold democracy”, in which each citizen (neuron) counts for little.

The NP comparison has realized Fechner's dream of “inner psychophysics” – relating neurophysiological activity to sensation. However, the precise nature of this relationship is still far from clear. There are a variety of unresolved questions, among them:

- (1) What is the role of single neurons in the representation of information? This question is closely related to the discriminability of a given set of stimuli by single neurons' responses.

Whether a neuron's discriminability is to be considered high or low can be most meaningfully assessed if viewed relative to psychophysical performance (Stüttgen, 2010). High discriminability of single neurons is a prerequisite for sparse coding; thus, it can constrain theories on how information is represented in a given brain area (such as response pooling or the lower envelope principle; Parker and Newsome, 1998). This, in turn, relates closely to Barlow's (1972) notions about the efficiency of neuronal representations.

- (2) What neural code is used for stimulus representation? NP comparisons can be used to compare psychometric to neurometric performance based on different candidate codes. For instance, it has been found that two candidate codes, firing rate and firing periodicity, both carry ample information about vibrotactile stimuli (Hernandez et al., 2000; Arabzadeh et al., 2006). The code with a neurometric performance that best matches the performance of the observers is then typically assumed to be the one that is used by the brain (e.g. Salinas et al., 2000; Luna et al., 2005). More systematic approaches try to assess "complete" neuronal populations (see below) and pit candidate codes against each other. Whenever some code's performance, computed in a statistically optimal fashion, falls short of the subjects' psychometric performance, that neuronal code can be rejected (Jacobs et al., 2009).
- (3) How is sensory information exploited for perceptual decision-making? Aside from the question of sensory processing, perceptual decision-making encompasses the problem of how sensory information is put to use for adaptive action (Gold and Shadlen, 2001, 2007). For example, monkeys do not make use of all stimulus information available to them, but rather commit to action prior to stimulus termination, thereby ignoring useful information (Roitman and Shadlen, 2002; see also Resulaj et al., 2009). Also, psychophysical performance is frequently not solely determined by sensory processes but by a range of biasing factors, among them recent stimulus and reward history (Boneau et al., 1965; Busse et al., 2011).

Importantly, the validity of claims about coding schemes on the single-neuron and population level hinges crucially on the precise assessment of the sensory limits of the observer, i.e., psychophysical sensitivity. While great effort has been devoted to the study of neural coding at both the level of individual neurons and neural populations (e.g., Bialek et al., 1991; Shadlen et al., 1996; de Ruyter van Steveninck et al., 1997; Riehle et al., 1997; Gold and Shadlen, 2007; Jacobs et al., 2009), we believe that research aimed at the understanding of the cognitive processes underlying performance in a given psychophysical task has been comparatively neglected by the community. As we will argue below, this could have led to a systematic underestimation of psychophysical sensitivity of animal subjects and, consequently, to an overestimation of neurometric relative to psychometric sensitivity.

In the remainder of the article, we will first review problems in acquiring neurometric data suitable for NP comparisons, a prime focus of research in the last 20 years. Then, turning toward the problem of measuring the psychometric function, we will introduce signal detection theoretical "process models", i.e., models of the sequence of cognitive steps underlying performance in these

tasks. We will discuss additional factors affecting psychophysical performance not accounted for by such process models. We argue that, in order to gain further insight into the physiology of perception, the entire cascade of cognitive processes underlying perceptual decision-making tasks has to be explored.

ESTIMATING NEUROMETRIC SENSITIVITY

Conducting an NP comparison poses two distinct problems: the assessment of neurometric sensitivity and the assessment of psychometric sensitivity. In this section, we will first briefly introduce what is meant by an NP comparison. Then, we will discuss problems that arise when attempting to determine psychometric sensitivity.

Figure 1 illustrates a simple NP comparison for a yes/no detection task. Several stimuli whose intensities are distributed around the presumed absolute threshold of detectability are presented

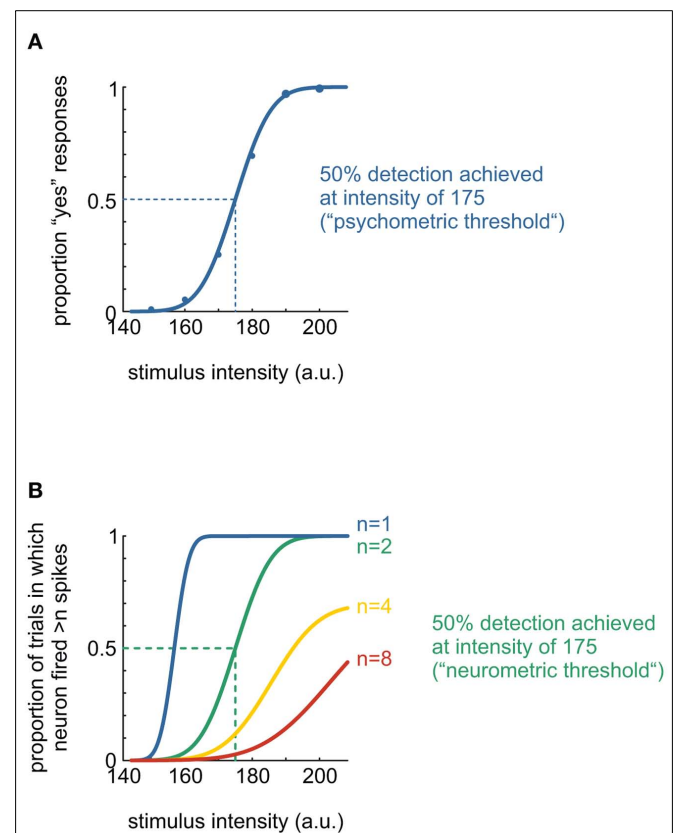


FIGURE 1 | Example illustration of a neurometric-psychometric comparison. (A) A typical psychometric curve from a yes/no detection experiment with six stimuli of varying intensity (see **Box 1**). Smooth line indicates fit of a cumulative Gaussian to the data points. Dotted line indicates the stimulus value at which the psychometric curve reaches 50% of its final height. This value is commonly taken as the psychophysical threshold. **(B)** Some typical neurometric curves: a single neuron's spikes were counted during stimulus presentation. The neuron was assumed to "detect" the event when it fired in excess of n spikes during stimulus presentation, where n here encompasses 1, 2, 4, and 8 spikes. The neurometric curve for $n=2$ matches the psychometric curve best, as assessed by their common threshold of 175 (arbitrary units). Thus, the NP ratio here is $175/175 = 1$.

many times to an observer, whose task is to respond “yes” if he perceives the stimulus and “no” otherwise. As stimulus intensity increases, the proportion of “yes” responses increases as well. The pattern of responses can be fitted by a sigmoidal function. The resulting psychophysical curve is characterized by at least two parameters, the threshold (the point on the abscissa corresponding to 50% detection performance) and the slope of the curve. The term “psychophysical sensitivity” refers to the reciprocal of threshold; thus, the lower the threshold, the higher the sensitivity.

Imagine that, while the observer was performing the task, a sensory neuron of his is recorded, and its spike responses during stimulus presentation are counted. Neurometric curves can be constructed in a simple way by determining the proportion of stimulus presentation trials on which the neuron fired more than n spikes (i.e., the neuron is deemed to detect the event when more than n spikes are fired during stimulus presentation). **Figure 1B** displays the results of this exercise for n comprising 1, 2, 4, and 8 spikes. It is readily visible that the curve constructed with a criterion of $n = 2$ resembles the psychometric curve best in terms of both threshold and slope. In fact, psychometric and neurometric threshold (for $n = 2$) are identical in this example, and the NP ratio therefore equals 1. Of course, there are considerably more ways to construct neurometric curves, perhaps most notably receiver-operating characteristic (ROC) analysis (Vogels and Orban, 1990; Britten et al., 1992), which compares distributions of spike counts for pairs of stimuli and returns the maximum classification performance. There exists a variety of population coding schemes beyond simple spike counts, involving spike timing, spike correlations within (Jacobs et al., 2009), and between spike trains of different neurons (Zohary et al., 1994; Shadlen et al., 1996; Schneidman et al., 2006; Shlens et al., 2006, 2009; Ohiorhenuan et al., 2010).

Obviously, NP comparisons require two different kinds of measurement – estimating psychometric and neurometric sensitivity. We will have more to say about psychometric sensitivity below. For now, we focus on two important preconditions for the assessment of neurometric sensitivity: causality and completeness. Naturally, in order to make meaningful NP comparisons, the sensory neurons under investigation must be involved in the psychophysical task at hand, i.e., neural activity of these neurons must be causally related to psychophysical performance. The neurometric signals entering the NP comparison should be necessary and, ideally, also sufficient to explain the sensory-driven aspects of the behavior. This requires to record from an “informational bottleneck”, i.e., a neural structure through which all relevant signals pass and which does not receive feedback signals from downstream structures. This way, a clear identification of cause and effect is possible. Unfortunately, such informational bottlenecks are rarely to be found, one notable exception being retinal ganglion cells (Jacobs et al., 2009). In the central nervous system, the direction of signal flow in ascending sensory pathways is ambiguous. With very few exceptions, subsequent levels of processing are interconnected in a bidirectional way. Also, circular connectivity often bypasses stations on the ascending pathways (e.g., cortical feedback of brain stem centers bypassing thalamic stations; Furuta et al., 2010). These problems exacerbate in the

neocortex where neurons are intricately interconnected – to their neighbors as well as to a multitude of distant neurons residing in other areas. Thus, a close look at the detailed connectivity of sensory systems blurs the notion of an “ascending pathway.” Instead, sensory systems seem to be described better as complex networks, which receive signals at one point and output signals at another with reverberating signal flow in-between. In conclusion, informational bottlenecks, furnishing completeness, and causality, are difficult to define in the central nervous system. Alternatively, a causal contribution of a neuronal structure can be demonstrated by showing that the behavior is blocked/evoked by lesions/electrical stimulation of the structure in question (Parker and Newsome, 1998). A cautionary note is in order, however, as psychophysical performance can readily be impaired by blocking structures located downstream from the ones actually performing the critical (sensory) computation; in this case, performance degradation may be due to, e.g., response confusion rather than abolition of the sensory function proper. Furthermore, abolishment of a function may result from disrupting non-specific modulatory structures with no contribution to encode or compute the concrete signals under observation. Also, parallel processing of sensory information along a different sensory pathway cannot be ruled out by this strategy; in the latter case, psychometric performance would depend on two or more structures, and the relation between neurons and sensation cannot be pinned down since the relative importance of the structures is not known.

Yet another strategy is to artificially “create” informational bottlenecks by presenting point-like stimuli in both space and time (Hecht et al., 1942; Barlow, 1961; Sakitt, 1972; Johansson and Vallbo, 1979; Vallbo et al., 1984). Point-like stimuli are attractive for studies of the physiology of perception as they reduce the number of neurons engaged in the task and the time window in which neuronal responses need to be monitored to a well specified minimum. Sakitt (1972), for instance, carried this strategy to the extreme as she studied the difference in visual detection performance evoked by just one photon more. One photon will interact just with one molecule of rhodopsin located in just one photoreceptor. This approach therefore ingeniously related the concept of informational bottleneck to just one cell in the layer of photoreceptors. Using sophisticated psychophysical techniques in humans, Sakitt successfully related a stimulus of “one more photon” to a difference in the subject’s performance, and was thus able to conclude that the action of a single photoreceptor has a significant contribution to perception. A related concept is the attempt to electrically stimulate a single neuron, which has been first realized using electrical stimulation of individual primary afferent fibers in humans (Ochoa and Torebjork, 1983; Vallbo et al., 1984). These authors found that subjects perceived the activation of individual tactile nerve fibers in three out of the four classes of fiber types investigated. Some rapidly adapting fibers seem to give rise to a perceptual change with a difference of just one evoked spike. More recently, the technical advent of juxtacellular stimulation made this approach available for the study of the central nervous system (Houweling and Brecht, 2008; Voigt et al., 2008). Injecting just about 15 spikes in one neuron in primary somatosensory cortex evoked a measurable difference in detection performance of

a rat showing that even single cortical neurons can have an effect on perception. This method bears great promise to be used for systematic mapping of behavioral effects in different stations of a sensory pathway.

Unfortunately, single neuron responses (and thus artificial bottlenecks) cannot be realistically obtained with natural sensory stimuli evolving in space and time, at least in mammalian brains. Even limiting the stimulus in space and time and applying near-threshold intensities, not one but many neurons in primary sensory cortices will get activated (de Lafuente and Romo, 2005; Stüttgen and Schwarz, 2008, 2010). Thus, neither the assessment of a complete informational bottleneck (aside of retinal ganglion cells), nor the creation of an artificial one constituted by a single cell seems attainable. One study in the whisker-related primary somatosensory (barrel) cortex of rats has provided a quantitative hint on the number of neurons engaged vs. the number of neurons needed to match the perceptual performance. Using transient single-whisker deflections at psychophysical threshold, around a third of the stimulations at psychophysical threshold intensity were responded by neurons in the principal barrel cortical column of the stimulated whisker (Stüttgen and Schwarz, 2008). Thus, alone in the barrel column receiving strongest input of the stimulated whisker (the principal barrel column) around 3000 neurons are active with minimal stimulation. The number of all cells engaged in primary somatosensory cortex is surely far higher because cells in adjacent barrel columns respond as well to single whisker stimuli. On the other hand, the same study found that five of the most sensitive neurons carry sufficient information to explain the psychometric performance. In case the read-out mechanism is less selective, around 16 barrel cortex neurons might be sufficient. In fact, this discrepancy of thousands of cells engaged by the stimulus vs. a few needed to do the job has become a common theme in all studies trying to compare the performance of single neurons to the one of the subject. Since the pioneering studies of Newsome, Movshon, and coworkers in the late 1980s, the single neuron neurometric performance has been found with few exceptions to be close and somewhat lower compared to that of the observer (Tolhurst et al., 1983; Newsome et al., 1989; Britten et al., 1992; Geisler and Albrecht, 1997; Uka and DeAngelis, 2003; Purushothaman and Bradley, 2005; Stüttgen and Schwarz, 2008; Cohen and Newsome, 2009).

An outlier result were the findings of the pioneering studies (Britten et al., 1992; Celebrini and Newsome, 1994), as neurometric sensitivity was judged up to 10 times higher than the psychometric one. However, this estimate of neurometric sensitivity has been recently adjusted downward by showing that monkeys use only the first few hundred milliseconds of a stimulus, while the neurometric integration time in the original study extended to the full stimulus presentation of 2 s. Thus, the neurons were unfairly favored in the earlier study (Cohen and Newsome, 2009). In addition, it needs to be pointed out that these pioneers actually based their estimate of the neurometric sensitivity on *two* neurons, and not (as is often falsely understood) on a single neuron. The neurometric sensitivity was calculated from the measured neuron (selected to display high directionality), combined with a virtual one, the “antineuron”, with opposite direction selectivity but otherwise identical response properties. In fact, the task to discriminate

two stimulus directions does not fit well the properties of a single MT neuron, which, due to its high directional selectivity, is limited to convey information about the presence of a stimulus in a single preferred direction, and largely ignores the presence of stimuli in other directions. As a consequence, single MT neuron discriminability to two opposite directions is presumably far lower than claimed in the original study (Britten et al., 1992). Notably, other studies with the neurometric analysis strategy of postulating antineurons also found considerable fractions of neurons whose sensitivity exceeded that of the observer (MT: Uka and DeAngelis, 2003; MST: Heuer and Britten, 2004). A perhaps more suitable psychophysical task to probe the sensitivity of MT neurons would be the detection of movement direction along the neuron's preferred axis vs. zero net motion, which to our knowledge has not been tried so far. These points hardly diminish the impact of these landmark studies, but they suggest the view that single MT cells are likely to fall in line with neurons in many other sensory areas investigated since then, showing neurometric sensitivities to be somewhat lower than psychometric sensitivity.

With the common finding of NP sensitivity ratios of close to 1 but not exceeding it, it is typically very easy to combine neurometric performance of far fewer neurons than the number suspected (or known) to be engaged in the task to exceed the performance of the subject. Popular responses to this problem have been to postulate (i) sources of noise in downstream processing, (ii) detrimental effects introduced by neuronal correlations, or (iii) intricacies of read-out mechanisms. Despite their value as testable hypotheses, these possibilities must be deemed highly under-constrained without a direct assessment of the complete neuronal population, even though some of them – such as neuronal correlations – have received experimental support (Zohary et al., 1994; Cohen and Newsome, 2008).

In conclusion, future NP comparisons are likely to go beyond measurements of spike counts from single units with the aim to identify neuronal population codes. To do this, informational bottlenecks must be studied. The only attainable complete bottleneck in the central nervous system is the population of retinal ganglion cells, which should be further exploited for this purpose. In the peripheral nervous system, somatosensory afferents are equally attractive. Artificial bottlenecks must be extended to activity that evolves in space and time to identify neuronal activity leading to more complex perception. Juxtacellular stimulation of single neurons can be employed to systematically test activity varying over time (Houweling et al., 2010). Optogenetic approaches, which already allow to interfere with a genetically targeted population of cells, are a promising new tool to achieve this goal with amazing spatiotemporal precision (Yizhar et al., 2011), even in non-human primates (Diester et al., 2011).

ESTIMATING PSYCHOMETRIC SENSITIVITY

Sensations are not directly observable in the laboratory. Instead, the subject in a psychophysical experiment (observer) is asked to produce different responses contingent on particular aspects of his sensations. This could be the mere presence or absence of a sensation (stimulus detection), whether two stimuli are perceived to be different (stimulus discrimination), whether a stimulus is a specific one in a set of candidates (identification), or whether

a given stimulus belongs to a specific category of stimuli (categorization). In the early days of psychophysics, the behavioral response was simply seen as the effect of a stimulus once its intensity exceeded a sensory threshold. The threshold could be estimated by varying the stimulus intensity and measuring the percentage of correct responses (**Figure 1A**). However, psychophysical findings varied considerably both across tasks and across laboratories, prompting psychophysicists to develop more reproducible methods (Blackwell, 1952; Swets, 1961a,b; Swets et al., 1961). The decision-theoretic stance of SDT (Green and Swets, 1966) alerted experimenters to the fact that psychophysical measurements hinge crucially on non-sensory factors, among them prior probabilities, payoffs, and task strategy, thus recognizing the active role of the observer. Importantly, SDT's main index of psychophysical sensitivity (d') promised improved replicability of results across both tasks and laboratories. The ensuing success of SDT yielded a sharp increase in the use of its concepts, such as ROC analysis, across various fields of research (Swets, 1973). The power of these ideas is reflected also in the measurement of neurometric data. When comparing them to psychometric data, neurometric discriminability is often measured by ROC analysis (Vogels and Orban, 1990; Britten et al., 1992, 1996; Parker and Newsome, 1998). If decisional factors play a role for the behavioral response in psychophysical tasks, it is reasonable to deploy their manipulation for the study of perceptual processes. As a consequence, the focus in the last decade has shifted away from neuronal sensitivity toward the study of perceptual decision-making. It is now asked how and to what degree sensory representations also reflect the behavioral choice (Britten et al., 1996; Romo and Salinas, 2003; Gold and Shadlen, 2007; Nienborg and Cumming, 2009, 2010). A more recent development has been to go beyond varying stimulus parameters and explicitly vary payoffs and/or the frequency of the stimuli to study directly how representations of stimulus and choice correspond and interact (Feng et al., 2009; Rorie et al., 2010; Teichert and Ferrara, 2010; Stüttgen et al., 2011).

Despite these developments, the psychophysical task – at least when used to measure neuronal sensitivities – has, by and large, been considered merely a means to measure responses. Typically, the minimization of extra-sensory factors is considered as a given. Against the backdrop of the insights gained by SDT half a century ago about the psychological nature of even the simplest sensory detection tasks, it gives cause of concern how little possible effects of extra-sensory factors on the psychometric curve are discussed. Our goal in this review is to remind the reader that all parameters of the psychometric curve depend on the detailed procedure and will, thus, significantly affect the estimation of psychometric sensitivity and thereby the NP ratio. Each psychophysical task comes with different memory requirements, constraints on information processing, and effects on motivation and bias that limit the use of sensory information. In order to study how sensory information processing works – even at the sensor level – ultimately these more “psychological” factors have to be taken into account. We will start with a brief review of SDT and an analysis of cognitive processes underlying performance in commonly employed psychophysical tasks. Then, we will discuss additional non-stimulus factors outside the SDT framework that may significantly influence the subject's responses.

SDT ANALYSIS OF THE YES/NO TASK

The vast majority of researchers undertaking the NP comparison employ Go–NoGo (GNG), yes/no (YN), or Forced-Choice (FC) tasks (see **Box 1**, **Table 1**, and **Figure 2** for brief descriptions of these and some other psychophysical tasks). SDT offers a broad conceptual framework for the analysis of different psychophysical tasks. Here, we will illustrate SDT concepts mainly with YN and FC. The interested reader is referred to MacMillan and Creelman (2005) for further paradigms and discussion.

Signal detection theory starts with the assumption that each presentation of a signal yields a variable internal representation on a hypothetical decision axis. Similarly, even in the absence of sensory input, the system generates a non-zero, somewhat variable response. In the simplest and most widely used case, the distributions of the internal representation of both stimulus (S) and noise (N) are assumed to be normal and their variances identical (**Figure 3**).

The task can be conceptualized as a statistical decision problem. The observer is assumed to partition the decision axis into the discrete response options that are available to him: “yes”, a signal was present, and “no”, no signal was present (a similar logic applies to discrimination tasks). On each trial, there are four possible outcomes: (1) a signal is presented, and the observer responds “signal” (hit), (2) a signal is presented, and the observer responds “no signal” (miss), (3) no signal is presented, and the observer responds “signal” (false alarm), and (4) no signal is presented, and the observer responds “no signal” (correct rejection). Cases 1 and 4 are correct responses; cases 2 and 3 are false. Given this experimental setup, a payoff matrix assigns a value to each of the four possible outcomes. Usually, correct responses are equally likely to yield reinforcement, and reinforcers are of the same magnitude for cases 1 and 4. Incorrect responses are usually punished, and again punishments are of the same magnitude for cases 2 and 3. If this is the case, and the stimuli are equally likely to occur, the observer's optimal (in the sense of maximizing accuracy, and therefore expected payoff) decision criterion is located right in the middle between the two stimulus distributions. Thus, the probability of hits equals that of correct rejections, and the probability of false alarms equals that of misses. The discriminability of the two stimuli, N and S, is given by the difference of means of the two stimulus distributions on the decision variable, divided by the common standard deviation (SD) of the distributions. This measure is called d' . SDT separates sensory discriminability (indexed by d') and response bias, which is the distance of the decision criterion from a neutral position (a measure called c), and therefore, at least in theory, provides a measure of sensitivity untainted by response bias. This separation is of great value because the usual index of performance in psychophysics, percent correct, is known to be highly susceptible to variations in task structure and response bias (Green and Swets, 1966).

SDT ANALYSIS OF THE TWO-INTERVAL FORCED CHOICE TASK

A classic example for how SDT can help relate different psychophysical tasks is the relationship between YN and two-interval forced choice (2-IFC; the same applies to spatial two-alternative forced choice, 2-AFC). In a simple instantiation of 2-IFC, the

Box 1 | Description of psychophysical tasks

Go–NoGo (GNG)

The observer has one response option R available (e.g., pressing or releasing a lever, performing a nose poke, or pecking a response key) and is required to respond when a stimulus of class A is presented and not to respond when a stimulus of class B is presented. The outline of a typical trial is depicted in **Figure 2A**. After an inter-trial interval (ITI), a stimulus is presented. If the subject responds within a given time frame (the “response window”) after target onset, reward is delivered; in case of a response after a non-target stimulus, punishment is delivered.

The biggest advantage of the GNG method is its simplicity. Animals are easily trained on GNG using intense, suprathreshold stimuli. Consequently, the intensity difference between the stimuli in classes A and B is gradually reduced, until no further improvement is possible (see Schwarz et al., 2010, for a methods review). Then, presenting a pseudorandom sequence of several stimuli (“method of constant stimuli”), the response probability for each stimulus is recorded, and a psychometric curve can be constructed (detection: Stüttgen et al., 2006, discrimination: Gerdjikov et al., 2010).

Note that this description refers to an instantiation of the yes/no task (a single stimulus per trial) in the form of a Go–NoGo paradigm. Of course, GNG can also be conducted with two stimuli per trial as in two-interval forced choice (see below).

Yes/no (YN; also known as A Not-A or single-interval forced choice)

The observer has two response options R_A and R_B available and is required to respond with “ R_A ” when stimulus A is presented and with “ R_B ” when stimulus B is presented. Rather than two individual stimuli, A and B can be classes of stimuli, e.g., leftward and rightward motion of various strengths. The outline of a typical trial is depicted in **Figure 2B**. After an inter-trial interval (ITI), a stimulus taken from either class A or class B is presented. If the observer emits the appropriate response within a given time frame (response window), reward is delivered; if the incorrect response is emitted, the subject is punished, usually by a brief time-out. In many monkey studies, the response consists in making a saccade to one of two choice targets. The term “yes/no” derives from main usage of the paradigm in studies of stimulus detection in the early days of psychophysics. However, use of the YN method is not limited to detection but is employed in studies of discrimination performance as well. Notably, many neuroscience papers list this paradigm as a “forced-choice task” (e.g., Britten et al., 1992), or “two-choice task” (Kepecs et al., 2008). Sometimes the YN-method is referred to as “single-interval forced choice.” Although this terminology has some conceptual appeal, we will avoid this term lest we add to the terminological confusion. Note that, in signal detection theoretical contexts, the YN method is typically understood to employ only two stimuli per block of trials (the consequences of departure from this rule are discussed in the main text). In addition, psychophysicists sometimes refer to a yes/no task with more than two stimuli as the “method of single stimuli.” Here we will use the term “yes/no task” for any task in which a single stimulus is presented per trial and in which the subject has two response options available, regardless of the total number of stimuli in the stimulus set.

Yes/no with reference (YNR)

This method is similar to the yes/no task described above with two different stimuli per trial. On each trial, a reference stimulus is presented first; then, a second stimulus (target) is presented. The subject’s task is to judge whether the target stimulus is more or less intense than the reference stimulus along some sensory continuum. The rationale in using YNR is to avoid decrements in performance due to bad recall of the reference stimulus’ features (stimulus uncertainty; see, e.g., Hautus et al., 2009).

Identification

The subject is presented with one of m stimuli in a single interval and has to emit one of m possible responses. Hence, the yes/no method with two stimuli is a special case of an identification task with only two responses. In cases where there are two responses which are not thought of as literally “yes” and “no,” such as leftward vs. rightward motion, identification might be a better term than YN.

Same-different

The observer is presented with two stimuli, either simultaneously or in succession, and has to judge whether they are the same or different. The position or the sequence of the two stimuli in a pair is randomized. Unlike YNR, the first stimulus in this task is not identical across trials.

Forced Choice

This task can take many forms. In the most common application (the “ n -interval forced choice task,” n -IFC), there are n stimuli on each trial, and the observer has to choose a target out of $n - 1$ distractor stimuli. In tactile psychophysics, a common implementation is the two-interval forced choice task (2-IFC, e.g., Luna et al., 2005; **Figure 2C**). Here, a stimulus is presented for a brief interval of time (e.g., 1 s), which is followed by a short inter-stimulus interval, and the presentation of a second stimulus. The subject has to decide which of the two stimuli the target is (e.g., which stimulus is of larger intensity or higher frequency). If 2-IFC is used to assess detection performance, one of the stimuli is the null stimulus, the other one is the target. Another implementation of forced choice is the spatial n -alternative forced-choice method (n -AFC, **Figure 2D**): on each trial, n stimuli are presented on a screen in front of the subject, who has to pick the target stimulus (e.g., Jacobs et al., 2009; see also Jäkel and Wichmann, 2006).

Incidentally, FC can also be instantiated as a GNG task, e.g., by asking the subject to respond when it believes the first stimulus to be the target, and to withhold responding when it believes otherwise.

A note on terminology

It is important to note the discordant uses of psychophysical terms in the animal neuro-psychophysics and the psychological literature. In forced choice methods (psychological use), the observer is always presented with multiple stimuli per trial, either in temporal succession (n -IFC) or simultaneously (e.g., at different spatial locations, n -AFC). An inconsistency of this terminology is that YN tasks are not commonly called FC, although they do feature a forced choice component (they require the observer to emit one of two responses on each trial). This is probably the reason why animal studies often call YN tasks FC, bearing the danger that characteristics of the different tasks that critically relate to the comparison of neurometric and psychometric data slip out of focus and get neglected (see main text). Here we adopt the psychological terminology (which is consistent with signal detection theory, see also Section 2.3.5 in Kingdom and Prins, 2010), despite the mentioned inconsistency; accordingly, many of the paradigms called “forced choice” in the neuroscience literature are referred to as yes/no method in the present review.

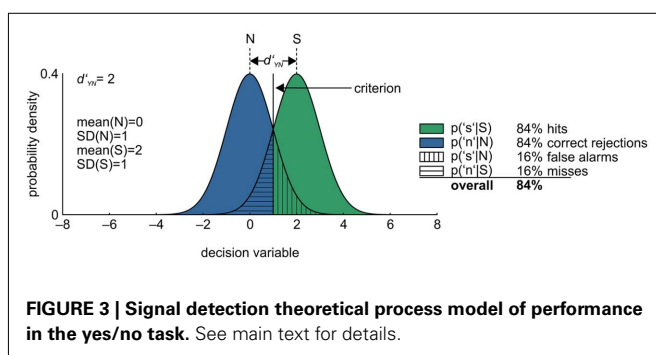
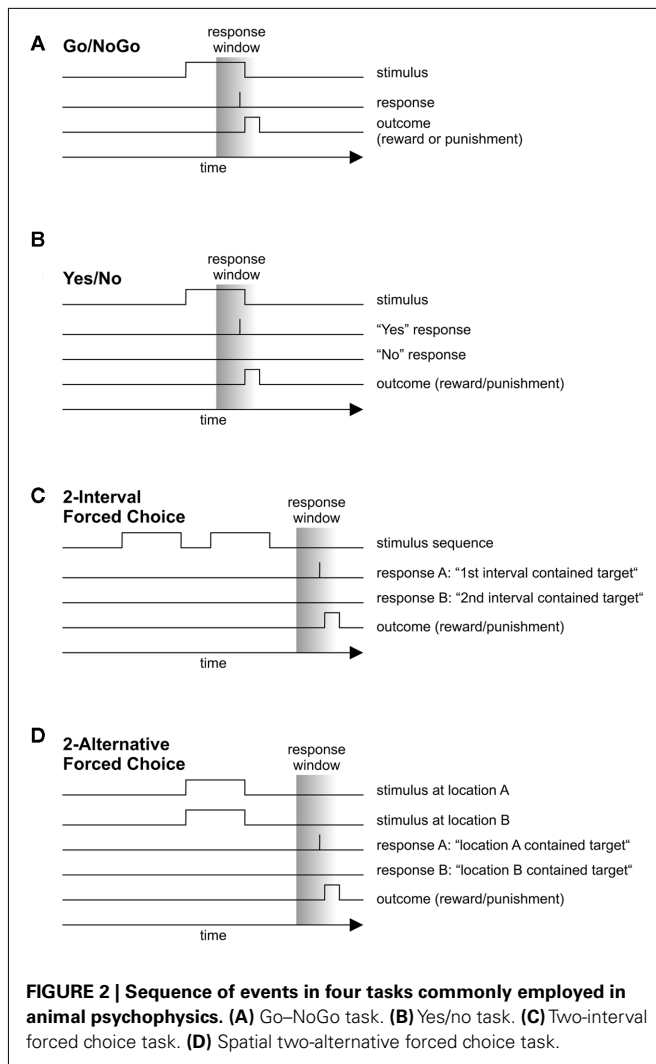
Table 1 | Overview over the most frequently used tasks in animal psychophysics and their properties.

	Number of stimuli per trial	Number of response options	Working memory for stimulus required?	Susceptibility to bias			Example for a putative process model	Example studies
				Motor	Interval	Motivation		
Go–NoGo	1 (but may vary)	1	No	++	n/a	++	Encode S – compare to C stored in LTM – decide whether to respond	Cook and Maunsell (2002), Stüttgen et al. (2006), Mehta et al. (2007), Palmer et al. (2007), Stüttgen and Schwarz (2008, 2010), Gerdjikov et al. (2010), O'Connor et al. (2010a,b), Frederick et al. (2011)
Yes/no	1	2	No	+	n/a	0	Encode S – compare to C stored in LTM – decide which response to emit	Britten et al. (1992, 1996), Krupa et al. (2001), de Lafuente and Romo (2005), Frederick et al. (2011), von Heimendahl et al. (2007)
Yes/no with reference	2	2	Depends on strategy	+	n/a	0	Encode R – transfer to WM – encode S – compare S&R – decide	Mountcastle et al. (1990), Purushothaman and Bradley (2005), Qin et al. (2009), also see Lee et al. (2007), Hautus et al. (2009)
Same-different	2	2	Depends on strategy	+	+	0	Encode S1 – transfer to WM – encode S2 – compare – decide	Vogels and Orban (1990)
<i>m</i> -Interval forced choice	<i>m</i>	<i>m</i>	Yes	+	+	0	For <i>m</i> = 2: encode S1 – transfer to WM – encode S2 – compare – decide	Hernandez et al. (1997), Luna et al. (2005)
Spatial <i>m</i> -alternative forced choice	<i>m</i>	<i>m</i>	No	+	n/a	0	For <i>m</i> = 2: encode S1 and S2 (simultaneously or sequentially?) – compare – decide	Mentzer (1966), Knutsen et al. (2006), Jacobs et al. (2009), Adibi and Arabzadeh (2011), Busse et al. (2011)

S, stimulus; *S1*, first stimulus; *S2*, second stimulus; *RM*, reminder; *C*, criterion; *WM*, working memory; *LTM*, long-term memory; ++, strong; +, moderate, 0, weak, n/a, not applicable.

observer is confronted with two different stimuli per trial; let us assume these are the same two stimuli that have been used previously in the YN task. The observer is presented with both stimuli on each trial, but each stimulus is assigned randomly to one of two successive temporal intervals. The observer's task is to designate which interval contained the target. Hence, contrary to the YN task, where the subject observes a sample from either the signal or the noise distribution on each trial, here the subject gets one sample from each without knowing which one is presented in which of the two intervals. The optimal strategy in this case is to take the difference between the two values and base the decision on the sign of the difference. **Figure 4** shows the two

distributions that arise when the decision is based on the samples' difference. They represent the cases when (a) the first interval contained the target ($S \rightarrow N$) and (b) the second interval contained the target ($N \rightarrow S$). In the first case, the distribution of the differences will be centered on the mean of the *S* distribution minus the mean of the *N* distribution – thus, the mean difference will be d' as would be obtained in a yes/no task (henceforth referred to as d'_{YN}), with a variance of $\sqrt{2}$. In the second case, the distribution of the differences will be centered on the mean of *N* minus the mean of *S*, again with a variance of $\sqrt{2}$, but with a mean of $-d'_{YN}$. Consequently, the distance of the two distributions is $2*d'_{YN}$. However, because of the increased variance,



d' in the 2-IFC task (d'_{FC}) is $\frac{2 \cdot d'_{YN}}{\sqrt{2}} = d'_{YN} \cdot \sqrt{2}$. Translated to correct performance, if chance performance equals 50%, this corresponds to an increase from 84% to 92% correct if $d'_{YN} = 1$. Thus, SDT predicts that an observer (if he adopts the optimal strategy) will have a $\sqrt{2}$ times higher discriminability (indexed by d'_{YN}) in a 2-IFC task than in a YN task with the very same two stimuli, i.e., $d'_{FC} = \sqrt{2} \cdot d'_{YN}$. Indeed, this prediction was

approximately confirmed in some studies (Swets, 1959) but not in others (Yeshurun et al., 2008).

THE STIMULUS SET AND PRIOR PROBABILITIES

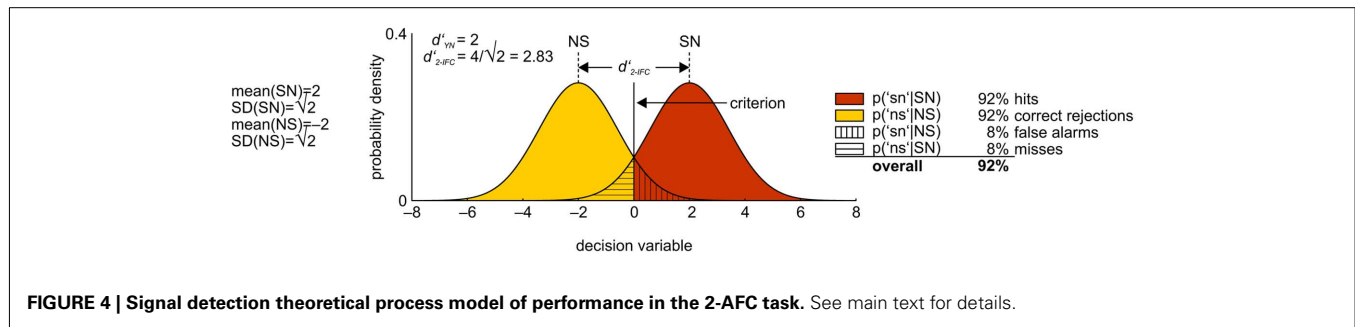
In most neurophysiological experiments, animals are presented with more than two stimuli varying in their discriminability. Each pair of stimuli has a specific d' that can be measured in an YN task as described above. Hence, each stimulus is assumed to give rise to a Gaussian distribution on the decision axis. How should the subject respond if we showed all stimuli randomized in one and the same block of an experiment? In **Figure 5A**, an observer is confronted with two stimulus categories, S1 and S2. S1 consists of a single stimulus (the noise-only stimulus N), S2 consists of five stimuli (each with a different d' compared to N). All six stimuli occur with equal probability (1/6) and the subject's task is to "detect" any stimulus that is greater than S1 in a simple YN task. The rightmost panel illustrates overall proportion of correct responses as a function of criterion placement. The resulting psychometric curve is depicted in **Figure 5E** (magenta), an example detection study which used such a stimulus set is Stüttgen et al. (2006).

Now consider a somewhat different situation: the stimuli are identical to those described above, but presentation of S1 is as likely as presentation of all stimuli in S2 taken together – thus $p(S1) = 0.5$, $p(S2) = 0.5$, and $p(S2_i) = 0.5/5 = 0.1$ for stimulus i , where $i \in \{1,2,3,4,5\}$. The optimal decision criterion has shifted considerably, and overall accuracy has dropped by 10% (see **Figure 5B**). The resulting psychometric function is shown in **Figure 5E** (blue), illustrating a marked reduction in the proportion of "S2" responses across all stimuli (for an example study, see Gerdjikov et al., 2010).

Imagine yet another situation: the observer is confronted with only two stimuli per session, S1 and one of the stimuli in category S2, in a series of YN experiments. This case is illustrated for two hardly distinguishable stimuli (**Figure 5C**) and two easily distinguishable stimuli (**Figure 5D**). For each of five possible pairs, percentage of correct responses can be calculated and used to construct a psychometric curve (**Figure 5E**, green). Of course, one could also conduct five consecutive 2-AFC or 2-IFC tasks, yielding somewhat higher performance (**Figure 5E**, red). As outlined in the previous section, 2-AFC/2-IFC performance (red) is consistently higher than YN performance (green) for ideal observers. This exercise illustrates an important point: psychophysical performance, measured in proportion correct responses, can be different under different tasks or even within the same task when identical stimuli occur with different probabilities. Notably, performance across tasks looks identical when transformed into the same unit of sensitivity, such as d'_{YN} .

THE MINIMALLY INFORMED OBSERVER

We can use the previous example to make another point. Regarding the NP comparison, it is crucial that the performance of the neurons is considered under the same constraints as the subject. In NP comparisons, the term "ideal observer" is often used very loosely to describe the optimal performance that an observer could achieve in the task given the neural recordings and some assumptions about the neural code. There is, however, also the question of how much of the task and the stimuli is known to the observer. In



order to distinguish an observer that has all the information available that is also available to the experimenter (and might hence be called ideal) from the situation that the subject is in, Boneau and Cole (1967) coined the term “minimally informed observer” for a model that only uses the information that is available to the subject and nothing more.

Assume the subject is confronted with the situation depicted in **Figure 5B** – an YN task in which the S1 stimulus presentation probability is the same as that of all S2-stimuli together. In parallel, unit recordings from sensory neurons were obtained, and the experimenter wishes to relate the subject’s performance to that of single neurons. The experimenter could, for example, compute ROC curves from the neuronal data for each pair of S1–S2 stimuli, and integrate the area under the ROC. The area under the ROC curve will correspond to the performance in a 2-AFC task if the difference model of SDT is correct and hence requires a correction of $\sqrt{2}$ to be comparable to an YN task. In **Figure 5E**, this amounts to a transition from the red to the green curve. Still, the neuron would be unfairly favored, since the analysis assumes a sequence of YN tasks with only two stimuli per block, while the observer was faced with six stimuli simultaneously. Contrary to the experimenter the observer does not know which stimulus is shown on each trial. Thus, the only possible strategy for the observer is to adopt a single decision criterion, as shown in **Figure 5B**. Optimal performance would accordingly result in the blue psychometric function in **Figure 5E**, and this is the correct analysis to apply to the neuronal data: to find a criterion which maximizes the percentage of correct responses when multiple stimuli can appear. Studies in which this procedure was applied include de Lafuente and Romo (2005) and Palmer et al. (2007).

PROCESS MODELS FOR PSYCHOPHYSICAL PERFORMANCE

Given the success of SDT in fitting psychophysical data, it is tempting to think of the calculations involved as actual cognitive processes. For the YN task, the sequence of steps can be conceptualized as follows: (1) encode the stimulus into the decision variable, (2) compare current value of the decision variable to the decision criterion retrieved from long-term memory, (3) decide on a response (**Figure 3**; see also Tanner, 1961). A process model for the GNG task with one stimulus per trial would be identical to that for YN, the difference being that, in YN, the observer has two response options (aside from non-task behavior), while in GNG, the observer has only one. Gomez et al. (2007) tested formal models of GNG and conclude that core processes of GNG and YN may be identical under some circumstances.

Two-interval forced choice is more complicated because there exist more than one process model for appropriate (but also sub-optimal) behavior. (a) The observer could ignore the stimulus in the first interval altogether and treat this task as an YN task, basing his decisions only on sensory evidence gathered in the second interval; (b) he could do the converse and base his decisions only on the first interval. In these two cases, we would expect that he performs just the same way with two stimuli as in the yes/no task. Another strategy (c) that will give the same performance with regard to percent correct would be to perform two times YN in succession. If the stimulus is detected in neither interval, or if it is falsely detected in both, a random response is produced. Otherwise the interval in which a stimulus was detected is chosen. Yet, there is a fourth, the optimal strategy, (d) that was already discussed above (illustrated in **Figure 4**).

The important thing to note here is that, for a given psychophysical task, there may be more than one decision strategy to follow. It is often a convenient assumption that subjects follow the optimal strategy, but we must not forget that in most studies that conduct NP comparisons it is only an assumption. Consequently, both the processing required to yield a decision variable and the resulting performance may differ from subject to subject. Even if the sensory front end as an input to the system is fixed, subjects may use the available information in various, potentially sub-optimal, ways. The 2-IFC task can however be adapted to force the animal to pay attention to both stimuli; see Romo and Salinas (2003).

A similar caveat as for 2-IFC applies to the use of the yes/no task with a reference stimulus (YNR, see **Box 1**), as for example used in Mountcastle et al. (1990); Purushothaman and Bradley (2005), Qin et al. (2009), and Bizley et al., 2010; also see Lee et al., 2007; Hautus et al., 2009). An analysis of the YNR task is depicted in **Figure 6**. Here, two stimuli per trial are presented. However, unlike 2-AFC/2-IFC, the first stimulus (reference, R) is identical for each trial, and the subject has to decide whether the second stimulus is more or less intense than R on some stimulus dimension. Again, the task is ambiguous as to its decision strategy. One strategy (the optimal one) is to ignore R completely and concentrate only on the second stimulus for decision-making. That way, YNR reduces to YN (**Figure 6A**). This assumes, of course, that all stimuli are known exactly to the subject. R is, however, only introduced because the experimenter thinks that this is not the case. One suboptimal strategy that seems likely is hence to (1) encode R, (2) encode the second stimulus, (3) take their difference, and (4) decide according to the sign of the difference; if

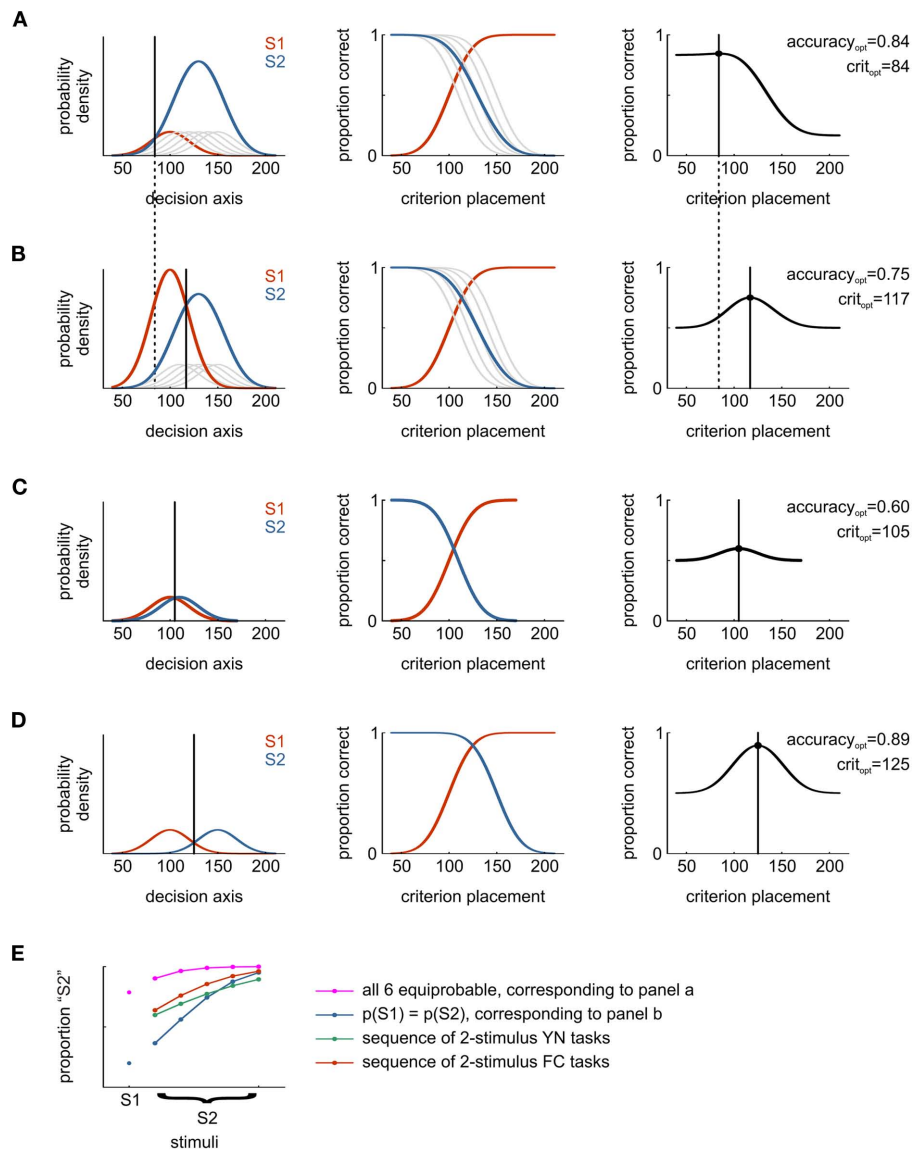


FIGURE 5 | Illustration how different stimulus presentation probabilities and different ROC-analysis strategies may yield disparate estimates of sensory performance. (A)

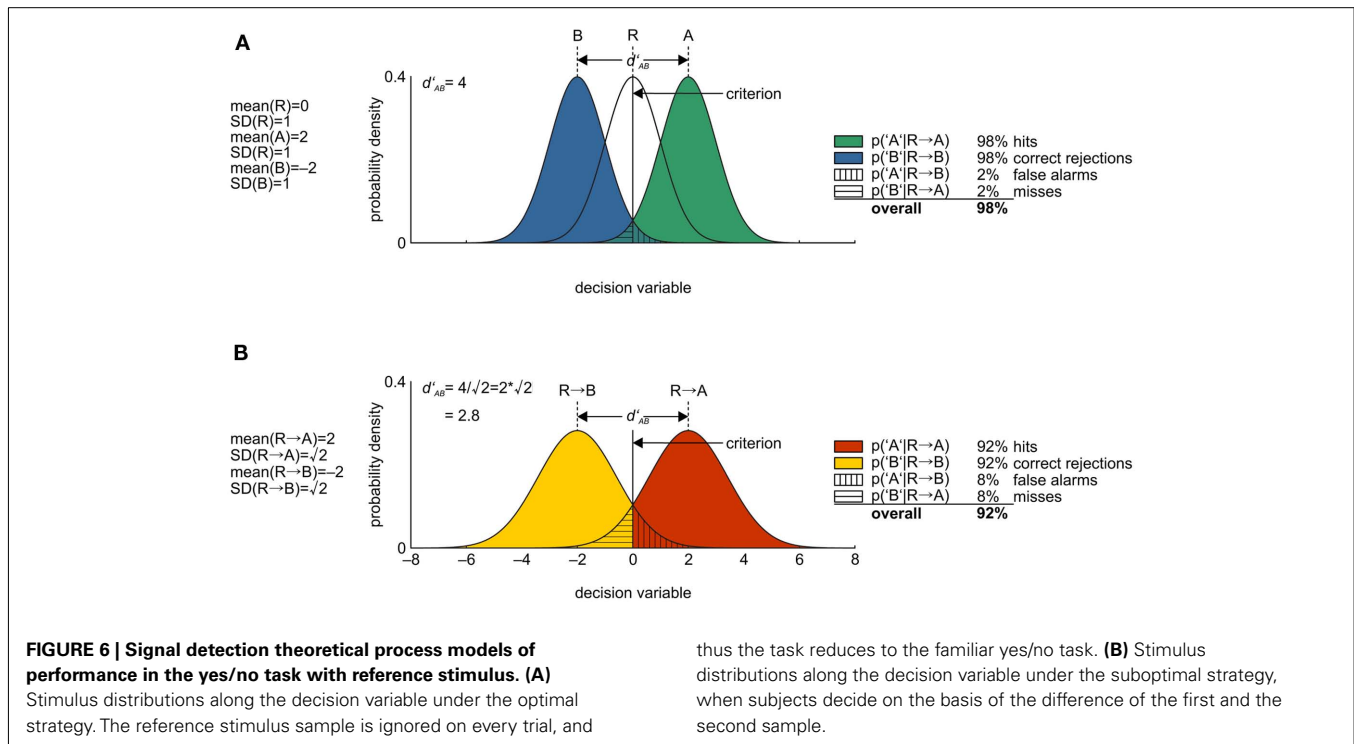
The total stimulus set comprises six different stimuli, five of which correspond to S2 (gray distributions, blue distribution is the sum of five individual ones) and one corresponds to S1 (red). All six stimuli occur with equal probability (means: 100:10:150) and have identical SD (20). Middle panel: depending on the location of the response criterion on the decision axis, different sets of probabilities of a correct response exist. For each possible criterion on the abscissa, the corresponding accuracies for each stimulus can be read off the ordinate. Right panel: overall proportion of correct responses (across all stimuli) as a function of criterion placement. Vertical line indicates optimal criterion placement. **(B)** As in **(A)**, but probability of S1 and S2 are equal (0.5 each);

within the S2 category, all stimuli are equally probable ($p=0.1$). For the same set of stimuli as in a, the optimal criterion is shifted considerably to the right, and the overall proportion of correct responses drops from 0.84 to 0.75. **(C)** As in **(A)**, but showing performance in a two-stimulus yes/no task with S1 and one stimulus out of S2 with the weakest signal strength. **(D)** As in **(A)**, but showing performance in a two-stimulus yes/no task with S1 and one stimulus out of S2 with strongest signal strength. **(E)** Psychometric functions for different task conditions: magenta, task as in **(A)**, blue, task as in **(B)**, green: psychometric curve resulting from a sequence of separate yes/no experiments where stimuli are presented pairwise and in blocks (i.e., S2 vs. S1-S1, S2 vs. S1-S2, S2 vs. S1-S3 etc.), red: psychometric curve resulting from a sequence of 2-AFC experiments where stimuli are presented pairwise and in blocks.

positive, the second stimulus is deemed more intense (**Figure 6B**). This strategy is identical to the fourth strategy discussed in the context of the 2-IFC task; but this time, it yields suboptimal performance, decreasing 98% correct performance to 92% in the example. Furthermore, because of the ambiguity in task execution,

it is unknown which neurometric analysis is most appropriate for this case.

One study actually demonstrated that, in YNR, animals ignore the reference stimulus and thereby follow the optimal strategy. Hernandez et al. (1997) trained monkeys to discriminate between



two vibrotactile stimuli of different frequency. Monkeys were presented with a base stimulus first and a comparison stimulus second, and they had to judge whether the frequency of the comparison stimulus was higher than that of the base stimulus. Importantly, when the reference stimulus was omitted in control experiments, psychophysical performance did not change, suggesting that the reference stimulus has indeed been ignored by the animals. Also, when conditions were changed such that both base and comparison frequency varied randomly from trial to trial, performance dropped to chance levels, indicating that the animals did not perform the subtraction strategy as delineated above (Figure 6B).

While process models inspired by SDT make clear predictions for comparing performance across different psychophysical tasks, data supporting these models as description of an observer's decision strategy is sparse and conflicting. For example, Yeshurun et al. (2008) reexamined several claims about the 2-IFC method. They found, contrary to widespread belief, that the 2-IFC task is not unbiased: observers consistently prefer one of the two intervals, and this preference could not be explained by attentional state, complexity of the stimulus display, interstimulus interval (ISI), or experience of the observers. That 2-IFC is usually not unbiased was also remarked on by Klein (2001) and the topic was recently revisited by Garcia-Perez and Alcalá-Quintana (2011) in a reanalysis of a large number of datasets. Moreover, sensitivity during the two intervals may differ: Yeshurun et al. (2008) provide some experimental evidence that d' in the first interval is larger than d' in the second interval. Similar observations have been reported and commented on by other authors (Nachmias, 2006; Ulrich and Vorberg, 2009; Ulrich, 2010). This asymmetry could be due to memory limitations, i.e., only a portion of the information from the first

interval is retained, or due to perceptual interactions between the two presentation intervals. Importantly, Yeshurun et al. (2008) found no evidence that d' in 2-IFC is $d'_{VN} \sqrt{2}$, as postulated by SDT. Thus, the standard SDT-difference model of 2-IFC performance was rejected, and the authors conclude that “we do not currently know how to model what observers actually do in 2-IFC tasks and that we have no reason to think that models appropriate to one choice of stimuli can be generalized to others.” In a similar vein, Jäkel and Wichmann (2006) compared 2-IFC with spatial 2-AFC and spatial four-alternative forced-choice (4-AFC) in a contrast detection task and found that, surprisingly, 2-IFC with foveal stimulation produced the highest thresholds and 4-AFC with more peripheral stimulation the lowest thresholds in naïve observers, but not in a highly experienced one. In a discrimination task with similar stimuli, 4-AFC did produce higher thresholds than 2-IFC, as expected. Although their data do not allow a clear interpretation of how the psychometric functions from the different tasks relate to each other, the authors speculate that extra-sensory factors, like sensory memory and spatial attention, have different effects in different tasks. It is noteworthy that these extra-sensory effects are ignored in SDT.

On the neurometric side it makes sense to calculate sensitivity using the optimal procedure in order to get an upper bound on the performance that an ideal observer could achieve based on the neural data. We usually also assume that the whole observer behaves optimally when calculating psychometric sensitivity. We have to be aware, however, that the actual sensitivity of the observer may be higher than observed, since he may be using the information that is available to him in a suboptimal way. Ideally, obtained psychometric functions should index “true sensitivity” – i.e., measure discrimination performance of a sensory system and

be unaffected by choice of psychophysical method, variations in motivation, response measure, or response topography. The simultaneous measurement of neuronal and behavioral responses is considered the gold standard for conducting the NP comparison, because neuronal responses are not altered by anesthesia, the animal is actively engaged in the task, and stimulus variability across trials affects neurons and observer alike (Parker and Newsome, 1998). That way, important confounds inherent in comparing neurometric and psychometric data from different animals, such as plasticity of sensory representations during learning (Polley et al., 2004) or task-dependent changes in interneuronal correlations (Cohen and Newsome, 2008) are avoided. However, as outlined above, simultaneous acquisition of neurometric and psychometric data is not sufficient for conducting valid NP comparisons, because task-specific (and -unspecific, see below) factors may affect psychophysical performance without affecting neurometric performance. As a consequence, psychophysical performance will frequently fall short of true sensitivity.

ADDITIONAL FACTORS AFFECTING MEASURED DISCRIMINABILITY

Psychometric discrimination and detection performance for identical stimuli have been shown to be affected not only by type of task (see preceding section), but by a variety of other factors as well. SDT explicitly acknowledges the role of prior presentation probability and reinforcement history of the stimuli, but there exists a wide range of factors which, we believe, have been largely ignored in previous work. In the following paragraphs, we will review some non-sensory factors that are known to affect psychophysical performance. A short list of important factors in conducting NP comparisons is provided in **Table 2**.

Learning, motivation, and fatigue

One would expect psychometric functions to change based on learning and this is a good reason to work with highly trained observers and only analyze the responses after their performance does not improve anymore (Fine and Jacobs, 2002). This is of course the case for most animal experiments, especially those involving monkeys, even though some studies employing rats or mice sometimes stop training when an arbitrary performance

criterion of, e.g., 80–85% has been achieved. Nevertheless, even within a session, a highly trained animal may show systematic deviations from stationarity. In order to achieve a high level of motivation, animals in psychophysical studies are usually food- or water deprived. Nienborg and Cumming (2009) used a yes/no task to assess disparity discrimination. They found that the delivery of larger rewards led to increased performance as measured by the slope of the psychometric function.

Hunger and satiety are known to offset response curves in psychophysical GNG tasks. Boneau and Cole (1967) separated response probabilities observed during the first half of an experimental session, when the subject was supposedly most hungry, from the second half of the session, when the animal was arguably less hungry; they observed a substantial decrease in overall response probability from the first to the second half of the session, which showed up at the level of the psychometric function as a shift of threshold. Similar effects are of course to be expected when the subject gets tired. In order to detect such non-stationarities, one possibility is to compute a rank-biserial correlation between trial number and responses (e.g., 1 for correct and 0 for incorrect; see Stüttgen and Schwarz, 2008). Ideally, the correlation should be 0. If the correlation assumes negative values, the number of correct responses is increasing over the duration of the session. As another means to detect such effects, Wichmann and Hill (2001a,b) describe a statistical test that uses the order of the blocks in a constant stimuli design to predict the residuals for the fit of the psychometric function. Freund et al. (2011) assess the severity of such violations on the estimation of psychometric functions and suggest a suitable correction for the resulting confidence intervals.

Attention

Animals are presumably inattentive to the task a significant portion of the time. In principle, this problem is unrelated to the psychophysical task employed, but it may be especially detrimental in GNG. Lapses of attention in GNG will tend to yield fewer responses overall and thereby increase the measured threshold. In GNG, the experimenter has no way to identify whether the absence of a response in a given trial is indeed based on assessment of the sensory evidence in a trial or due to non-sensory factors such as lapse of attention or decreased motivation. However, even in YN and FC, this may cause problems if the animal does not simply refrain from responding on such trials, but instead presses buttons or makes saccades at random. To complicate issues further, it could be that, beyond non-sensory influences on response bias, sensitivity itself could be affected by fluctuations of attention. For example, Treue and Martinez Trujillo (1999) reported that tuning curves of neurons in MT are gain-controlled by attention. Assuming spike count as the relevant code, this could affect performance if neurons tuned to the stimulus increased their firing rate while the firing rate of “comparison neurons” not tuned for the stimulus, remained the same; in SDT terms, the mean of the signal distribution would move away from the mean of the noise distribution, yielding an increase in d' . To control for fluctuations in attention within a session, experimenters can follow strategies as suggested in the previous section on motivation and fatigue.

Table 2 | Overview over the most frequent factors potentially affecting the NP comparison.

Temporal uncertainty (stimulus onset and offset not made explicit in detection task)
Stimulus uncertainty (presenting more than two stimuli per block of trials, or presenting novel stimuli)
Stimuli for neurometric and psychometric data collection differ appreciably
Neurometric data gathered with different animals, under anesthesia or <i>in vitro</i>
Ambiguous task structure (e.g., yes/no with reminder)
Subject not trained to asymptotic performance
Subject lacks motivation
Subject is inattentive
Response bias

Working memory

In GNG and YN tasks, working memory is not required in the sense that sensory information needs to be maintained over a short time span, e.g., a visual or auditory signal (this is not meant to imply that task execution is completely independent of working memory, as the animal needs to recall what task to perform, which lever or button to press under what circumstances etc.). In 2-IFC, SDT assumes perfect retention of the first stimulus, regardless of the ISI. If the sequence of the stimuli is seamless, no working memory is needed and discriminability depends on the temporal contrast of the two stimuli. In this case, the 2-IFC paradigm tests predominantly sensory coding. If, however, the stimuli are separated by a non-zero ISI, storing and retrieving stimulus properties in working memory plays a decisive role. Importantly, performance in 2-IFC is affected by the duration of the ISI. If the ISI is too long, performance decreases (Harris et al., 2001). It is a welcome recent development in neurophysiology that the mechanisms of sensory working memory are under investigation (Romo et al., 1999). The interplay between simple psychophysical paradigms and working memory is certainly a worthwhile field of theoretical and experimental development (Machens et al., 2005). In any case, it is likely that neurometrics in 2-IFC overestimate performance when sensory neuron responses during the first stimulus period are used, rather than the memory trace of the first stimulus as represented by working memory neurons.

Sequential effects

Boneau et al. (1965) showed that, at the level of individual trials, non-rewarded stimuli are more likely to elicit a response when they immediately follow a rewarded trial (in a GNG task). In addition, Busse et al. (2011) report that animals tend to switch sides after each trial, regardless of success or failure. Verplanck et al. (1952) have shown that for human subjects trials in a detection experiment are not independent – contrary to the usual assumption, subsequent trials in a detection experiment were positively correlated. If these effects were just due to higher cognitive effects, like the gambler's fallacy, then perhaps there would be hope that their effects could be minimized by instruction or training. However, there are indications that the sequential effects are a trace of the mechanisms that produce the observed behavior. For identification tasks (such as the YN task) it seems likely that the subject needs to store the ideal stimuli in long-term memory and then compares the stimulus on each trial to the stored representations to decide on a response. However, Stewart et al. (2005) argue that in many cases long-term memory is not necessary and that a simple mechanism that only compares the current stimulus to the last trial can explain many aspects of the data. For detection tasks, Treisman and Williams (1984) have argued that the sequential effects arise through an adaptive setting of the criterion based on previous trials. If this was the case then the fluctuations in the criterion should be taken into account when assessing a subject's sensitivity, e.g., by separating trials according to stimuli presented in each preceding trial.

Stimulus set

The measured psychophysical discriminability of two given stimuli can depend on whether, which and how many other stimuli

constitute a stimulus set in an experimental session (Stüttgen, unpublished data).

In many psychological experiments the stimulus range can influence the behavior that one wants to measure, often with unexpected results (Poulton, 1975). For example, Lages and Treisman (1998) show that a task that suggests comparison of a stimulus to a reference stimulus from long-term memory is actually solved by the subject by taking into account the stimulus range without recourse to the reference stimulus at all (a possible explanation for this can be found in Treisman and Williams, 1984).

A related problem is that, in many neurophysiological studies, details of the stimuli (e.g., retinal position, motion direction, contrast etc.) are meticulously matched to the receptive field properties of the neuron currently under study, in order to maximize the chance that this neuron is actually involved in the psychophysical task. However, because this stimulus adaptation has to be done for each single unit recording (see Britten et al., 1992), it may have detrimental effects on the performance of the subject which is required to generalize the task across a large variety of stimuli, many of which it may never have seen before. For most sensory areas, it is commonly assumed that the neural response to a stimulus is thought to be largely unaffected by stimulus history, as long as some reasonable ISI is provided. Accordingly, neural representation of a given stimulus should remain unaltered by the number of stimuli in a stimulus set, while psychophysical performance is not. Therefore, experimenters should take care to meet the assumptions of SDT lest the subjects exhibit suboptimal performance.

Temporal and stimulus uncertainty

It is often neglected that ideal observer analysis of spike responses using SDT (construction of ROC curves) requires some assumptions that are frequently not met by experimental conditions. SDT analysis assumes that the observer knows everything about the signal, including starting time, duration, phase, frequency, amplitude, and location – a prerequisite sometimes referred to as “signal specified exactly.” If experimental subjects are uncertain as to any of these parameters, performance decreases (Shipley, 1960; Swets et al., 1961; Green and Weber, 1980; Green and Forrest, 1989).

Many neuroscience studies aiming at the NP comparison violate at least one of these assumptions; most often, multiple stimuli are used per experimental block (see stimulus range), or the timing of the stimulus is held uncertain (e.g., de Lafuente and Romo, 2005; Stüttgen et al., 2006). Hernandez et al. (1997) compared monkeys' performance for vibrotactile frequency discrimination in two different tasks: yes/no with reference stimulus and 2-IFC with variable stimulus pairs across trials. The monkey's difference limina in the first set of experiments were lower by ~30% (thus, sensitivity was higher). This effect is likely due to the added stimulus uncertainty, because performance in the second experiment would be expected to *increase* according to SDT. Assuming that neural responses were not systematically affected by task type, the NP comparisons would yield different results for the two sets of experiments.

WHICH TASK IS BEST SUITED FOR THE NEUROMETRIC–PSYCHOMETRIC COMPARISON?

Blackwell (1952) systematically compared psychophysical methods for measuring visual thresholds in human subjects. He concluded that the 2-IFC method is superior to YN on several indices of quality, including reliability of threshold measurement (variability of repeated assessments of threshold), vulnerability of threshold measurement to non-sensory biasing factors (i.e., procedural factors such as background illumination, number, spacing, and order of stimuli, whether feedback was provided, and whether financial incentives for optimal performance was offered), and the absolute magnitude of the psychophysical threshold. Jäkel and Wichmann (2006) reinvestigated this issue and confirmed Blackwell's earlier results for a visual detection task – however, only for experienced observers. For naïve observers, in contrast, spatial 4-AFC was superior in terms of reliability, bias, sensory determinacy, and efficiency of measurement. For animal subjects, Mentzer (1966) has conducted similar comparisons of YN, 2-AFC, and 4-AFC for light detection in pigeons, but could not find any performance differences. Frederick et al. (2011) conducted a comparison of GNG and YN for odor discrimination and also found no evidence for major differences in resulting performance.

Most psychophysical studies employing unit recordings in primates have used the YN method, even though it is usually referred to by another name (e.g., Britten et al., 1992, 1996; Dodd et al., 2001; Uka and DeAngelis, 2003; Heuer and Britten, 2004; de Lafuente and Romo, 2005; Nienborg and Cumming, 2006, 2007, 2009); some have used GNG (Cook and Maunsell, 2002; Palmer et al., 2007); or some other method (YNR: Mountcastle et al., 1990; Purushothaman and Bradley, 2005; Qin et al., 2009; Bizley et al., 2010; same-different: Vogels and Orban, 1990). A series of studies by the Romo group has consistently employed 2-IFC (Romo et al., 1999; Hernandez et al., 2000; for review, see Romo and Salinas, 2003). To our knowledge, psychophysics with concomitant unit recordings in other species – most notably rats and mice – have so far almost exclusively relied on GNG (Stüttgen et al., 2006; Mehta et al., 2007; Stüttgen and Schwarz, 2008, 2010; Andermann et al., 2010; Gerdjikov et al., 2010; O'Connor et al., 2010a,b) or YN (Krupa et al., 2001; Prigg et al., 2002; Feierstein et al., 2006; von Heimendahl et al., 2007; Kepecs et al., 2008; Frederick et al., 2011). However, these species can be trained on FC tasks as well (pigeons: spatial 2-AFC: Blough, 1971; 4-AFC: Mentzer, 1966; mice: 2-AFC, Jacobs et al., 2009; Busse et al., 2011; Haiss et al., submitted; rats: 2-AFC: Knutsen et al., 2006; Adibi and Arabzadeh, 2011). We know of no study with these species which employed the *m*-IFC task; still, since rats, mice, and pigeons are known to learn delayed matching-to-sample problems (rats: Kesner et al., 1996; mice: Goto et al., 2010; pigeons: Lissek and Güntürkün, 2004), it should be possible to train them on *m*-IFC as well. To sum up, while most studies have so far employed YN, other methods seem feasible. It is common understanding in the community of researchers (based on anecdotal evidence) that GNG is trained faster than YN (but see Frederick et al., 2011), which again may be trained faster than IFC. More effort is required to make all psychophysical tasks routinely available for future psychophysical research. Spatial *m*-AFC has the disadvantage that, since several stimuli are presented simultaneously, it is difficult to control for

repetitive shifts of attention during the course of a single trial, and to attribute modulations in unit activity to any one stimulus, as opposed to the entire stimulus display. *m*-IFC avoids this problem because stimuli are presented successively. On the other hand, *m*-IFC requires working memory for the stimulus during the ISI (unless the interval is zero). In addition, all FC variants (as well as YNR) leave room for different decision strategies (see above), which need to be properly assessed before conducting the NP comparison. GNG and YN methods have the advantage that no sequential or simultaneous stimulus presentation is required. Accordingly, no working memory for a sensory stimulus is necessary, which potentially simplifies the task. We believe that the YN method is particularly well suited for NP comparisons. Unlike GNG, lapses of attention, or impulsive responding do not directly contaminate the response measure, compared to FC and YNR, there are less degrees of freedom in terms of strategy to employ, although we regret to say that there are no good data to back up this claim, and these data are badly needed.

CONCLUSION

The comparison of neurometric and psychometric sensitivity is fraught with problems. We have argued in this review that, in stark contrast to estimation of neurometric sensitivity, problems with the estimation of psychometric sensitivity have been largely ignored in the literature on the physiology of perception. Nevertheless, on both sides significant progress will be needed to make NP measurements more precise. Here we list some recommendations for future work originating from the points raised in this review.

On the neurometric side, we see the research program based on recording single neurons while activating them with sensory stimuli coming to an end. This approach has been invaluable to demonstrate that the neurometric sensitivity of single cells most often reaches close to (but hardly surpasses) that of the observer, thus fostering a central tenet of theories of sparse coding, as predicted by Barlow and Mountcastle. However, beyond showing sparse coding to be feasible in principle, this approach helps little in elucidating the role of the large neural populations activated even by near-threshold stimulation. The goal of today must be to characterize the neuronal code of the population of neurons carrying precisely the information leading to behavior. The need to define and access informational bottlenecks renders this a tough task. Retinal ganglion cells have been spotted to be one such bottleneck and should be exploited further. The creation of bottlenecks by juxtacellular stimulation and soon by optogenetic means will allow carrying this research program further both in rodents and in monkeys. In passing, we point out that bottlenecks can be found and/or created very easily in invertebrate model systems which sometimes employ just single or a few neurons to carry lifesaving, and thus, evolutionary relevant information. An instructive example has been provided by Roeder in his studies of noctuid moths. These insects use auditory information from just two neurons per ear to decide on different tactics to escape foraging bats (Roeder, 1966). Insects exhibit complex types of behavior, such as working memory and decision making (Menzel and Giurfa, 2001; Pompilio et al., 2006). Also, they offer exquisite experimental flexibility in terms of genetic manipulation

and optical imaging of neuronal function (Briggman et al., 2005; Haehnel et al., 2009). Accordingly, invertebrates may serve as valuable model systems to investigate the physiology of perception, and to offer useful insights for the studies of mechanisms of perceptual decision making in mammals.

On the psychometric side, the importance of task structure and other non-sensory factors relevant for psychophysical performance must be acknowledged. More effort is needed to validate the measurements of psychometric sensitivity by deliberate variation of task structure while maintaining a constant stimulus set. For instance, results from YNR or FC studies that allow ambiguous interpretations in terms of underlying cognitive processes can be validated by applying YN tasks. Formal models of the cognitive processes underlying different tasks need to be refined and pitted against each other both with purely behavioral tests (Gomez et al., 2007; Jang et al., 2009; Wolfe and Van Wert, 2010; Frederick et al., 2011; Stüttgen et al., 2011) and with neural recordings (Smith and Ratcliff, 2004; Gold and Shadlen, 2007; Churchland et al., 2008; Kepecs et al., 2008). It is unclear what kind of

comparison process underlies perceptual decisions, i.e., what is actually compared (Stüttgen et al., 2011). The effect of storing sample stimuli and/or decision criteria in long-term and working memory against which current sensory information can be compared demands clarification. As shown in **Figure 5**, psychometric performance for identical stimulus discriminations can be wildly different dependent on presentation strategy. Thus, psychometric performance must be compared with presenting pairs of stimuli vs. a whole stimulus array, and algorithms to calculate optimal neurometric sensitivity must be adjusted to reflect the animals' optimal strategy, given these circumstances. For further studies of neural coding in sensory system, we hold it vital to acknowledge that clean estimates of "true" psychophysical sensitivity cannot be obtained without appropriate models of perceptual decision-making. Such models need not only isolate sensitivity from response bias (Tanner and Swets, 1954; McCarthy and Davison, 1981; Busse et al., 2011) but from other factors affecting observed performance as well, be they inherent to the psychophysical task or not.

REFERENCES

- Adibi, M., and Arabzadeh, E. (2011). A comparison of neuronal and behavioral detection and discrimination performances in rat whisker system. *J. Neurophysiol.* 105, 356–365.
- Andermann, M. L., Kerlin, A. M., and Reid, R. C. (2010). Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front. Cell. Neurosci.* 4:3. doi:10.3389/fncel.2010.00003
- Arabzadeh, E., Panzeri, S., and Diamond, M. E. (2006). Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *J. Neurosci.* 26, 9216–9226.
- Baird, J. C., and Noma, E. J. (1978). *Fundamentals of Scaling and Psychophysics*. New York: Wiley.
- Barlow, H. B. (1961). "Possible principles underlying the transformations of sensory messages", in *Sensory Communication*, ed. W. Rosenblith (Cambridge: MIT Press), 217–234.
- Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394.
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., and Warland, D. (1991). Reading a neural code. *Science* 252, 1854–1857.
- Bizley, J. K., Walker, K. M. M., King, A. J., and Schnupp, J. W. H. (2010). Neural ensemble codes for stimulus periodicity in auditory cortex. *J. Neurosci.* 30, 5078–5091.
- Blackwell, H. R. (1952). Studies of psychophysical methods for measuring visual thresholds. *J. Opt. Soc. Am.* 42, 606–616.
- Blough, P. M. (1971). Visual acuity of pigeon for distant targets. *J. Exp. Anal. Behav.* 15, 57–67.
- Boneau, C. A., and Cole, J. L. (1967). Decision theory, the pigeon, and the psychophysical function. *Psychol. Rev.* 74, 123–135.
- Boneau, C. A., Holland, M. K., and Baker, W. M. (1965). Color-discrimination performance of pigeons – effects of reward. *Science* 149, 1113–1114.
- Briggman, K. L., Abarbanel, H. D., and Kristan, W. B. Jr. (2005). Optical imaging of neuronal populations during decision-making. *Science* 307, 896–901.
- Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., and Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* 13, 87–100.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765.
- Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Scholvinck, M. L., Zaharia, A. D., and Carandini, M. (2011). The detection of visual contrast in the behaving mouse. *J. Neurosci.* 31, 11351–11361.
- Celebrini, S., and Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J. Neurosci.* 14, 4109–4124.
- Churchland, A. K., Kiani, R., and Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nat. Neurosci.* 11, 693–702.
- Cohen, M. R., and Newsome, W. T. (2008). Context-dependent changes in functional circuitry in visual area MT. *Neuron* 60, 162–173.
- Cohen, M. R., and Newsome, W. T. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* 29, 6635–6648.
- Cook, E. P., and Maunsell, J. H. (2002). Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nat. Neurosci.* 5, 985–994.
- de Lafuente, V., and Romo, R. (2005). Neuronal correlates of subjective sensory experience. *Nat. Neurosci.* 8, 1698–1703.
- de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science* 275, 1805–1808.
- Diester, I., Kaufman, M. T., Mogri, M., Pashaie, R., Goo, W., Yizhar, O., Ramakrishnan, C., Deisseroth, K., and Shenoy, K. V. (2011). An optogenetic toolbox designed for primates. *Nat. Neurosci.* 14, 387–397.
- Dodd, J. V., Krug, K., Cumming, B. G., and Parker, A. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *J. Neurosci.* 21, 4809–4821.
- Evarts, E. V. (1966). "Methods for recording activity of individual neurons in moving animals", in *Methods in Medical Research*, ed. R. F. Rushmer (Chicago: Year Book), 241–250.
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf.
- Feierstein, C. E., Quirk, M. C., Uchida, N., Susulski, D. L., and Mainen, Z. F. (2006). Representation of spatial goals in rat orbitofrontal cortex. *Neuron* 51, 495–507.
- Feng, S., Holmes, P., Rorie, A., and Newsome, W. T. (2009). Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Comput. Biol.* 5, e1000284. doi:10.1371/journal.pcbi.1000284
- Fine, I., and Jacobs, R. A. (2002). Comparing perceptual learning tasks: a review. *J. Vis.* 2, 190–203.
- Frederick, D. E., Rojas-Libano, D., Scott, M., and Kay, L. M. (2011). Rat behavior in go/no-go and two-alternative choice odor discrimination: differences and similarities. *Behav. Neurosci.* 125, 588–603.
- Fruend, I., Hanel, N. V., and Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *J. Vis.* 11.
- Furuta, T., Urbain, N., Kaneko, T., and Deschenes, M. (2010). Corticofugal control of vibrissa-sensitive neurons in the interpolaris nucleus of the trigeminal complex. *J. Neurosci.* 30, 1832–1838.
- García-Pérez, M. A., and Alcalá-Quintana, R. (2011). Interval bias in 2AFC detection tasks: sorting out the artifacts. *Atten. Percept. Psychophys.* doi: 10.3758/s13414-011-0167-x. [Epub ahead of print].
- Geisler, W. S., and Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Vis. Neurosci.* 14, 897–919.
- Gerdjikov, T. V., Bergner, C. G., Stüttgen, M. C., Waiblinger, C., and Schwarz, C. (2010). Discrimination of vibrotactile stimuli in the rat whisker system: behavior and neurometrics. *Neuron* 65, 530–540.

- Gold, J. I., and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci. (Regul. Ed.)* 5, 10–16.
- Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574.
- Gomez, P., Ratcliff, R., and Perea, M. (2007). A model of the go/no-go task. *J. Exp. Psychol. Gen.* 136, 389–413.
- Goto, K., Kurashima, R., and Watanabe, S. (2010). Delayed matching-to-position performance in C57BL/6N mice. *Behav. Processes* 84, 591–597.
- Green, D. M., and Forrest, T. G. (1989). Temporal gaps in noise and sinusoids. *J. Acoust. Soc. Am.* 86, 961–970.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Green, D. M., and Weber, D. L. (1980). Detection of temporally uncertain signals. *J. Acoust. Soc. Am.* 67, 1304–1311.
- Haehnel, M., Froese, A., and Menzel, R. (2009). In vivo Ca²⁺-imaging of mushroom body neurons during olfactory learning in the honey bee. *J. Vis. Exp.* 30, e1353. doi:10.3791/1353.
- Harris, J. A., Harris, I. M., and Diamond, M. E. (2001). The topography of tactile working memory. *J. Neurosci.* 21, 8262–8269.
- Hautus, M. J., van Hout, D., and Lee, H. S. (2009). Variants of A Not-A and 2AFC tests: signal detection theory models. *Food Qual. Prefer.* 20, 222–229.
- Hecht, S., Shlaer, S., and Pirenne, M. H. (1942). Energy, quanta, and vision. *J. Gen. Physiol.* 25, 819–840.
- Hernandez, A., Salinas, E., Garcia, R., and Romo, R. (1997). Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *J. Neurosci.* 17, 6391–6400.
- Hernandez, A., Zainos, A., and Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6191–6196.
- Heuer, H. W., and Britten, K. H. (2004). Optic flow signals in extrastriate area MST: comparison of perceptual and neuronal sensitivity. *J. Neurophysiol.* 91, 1314–1326.
- Houweling, A. R., and Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. *Nature* 451, 65–68.
- Houweling, A. R., Doron, G., Voigt, B. C., Herfst, L. J., and Brecht, M. (2010). Nanostimulation: manipulation of single neuron activity by juxtacellular current injection. *J. Neurophysiol.* 103, 1696–1704.
- Jacobs, A. L., Fridman, G., Douglas, R. M., Alam, N. M., Latham, P. E., Prusky, G. T., and Nirenberg, S. (2009). Ruling out and ruling in neural codes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5936–5941.
- Jäkel, F., and Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naive observers. *J. Vis.* 6, 1307–1322.
- Jang, Y., Wixted, J. T., and Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *J. Exp. Psychol. Gen.* 138, 291–306.
- Johansson, R. S., and Vallbo, A. B. (1979). Detection of tactile stimuli. Thresholds of afferent units related to psychophysical thresholds in the human hand. *J. Physiol.* 297, 405–422.
- Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231.
- Kesner, R. P., Hunt, M. E., Williams, J. M., and Long, J. M. (1996). Prefrontal cortex and working memory for spatial response, spatial location, and visual object information in the rat. *Cereb. Cortex* 6, 311–318.
- Kingdom, F. A. A., and Prins, N. (2010). *Psychophysics: A Practical Introduction*. London: Academic Press.
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: a commentary. *Percept. Psychophys.* 63, 1421–1455.
- Knutsen, P. M., Pietr, M., and Ahissar, E. (2006). Haptic object localization in the vibrissal system: behavior and performance. *J. Neurosci.* 26, 8451–8464.
- Krupa, D. J., Matell, M. S., Brisben, A. J., Oliveira, L. M., and Nicolelis, M. A. (2001). Behavioral properties of the trigeminal somatosensory system in rats performing whisker-dependent tactile discriminations. *J. Neurosci.* 21, 5752–5763.
- Lages, M., and Treisman, M. (1998). Spatial frequency discrimination: visual long-term memory or criterion setting? *Vision Res.* 38, 557–572.
- Lee, H. S., van Hout, D., and Hautus, M. J. (2007). Comparison of performance in the A-Not A, 2-AFC, and same-different tests for the flavor discrimination of margarines: the effect of cognitive decision strategies. *Food Qual. Prefer.* 18, 920–928.
- Lissek, S., and Güntürkün, O. (2004). Maintenance in working memory or response selection? Functions of NMDA receptors in the pigeon “prefrontal cortex.” *Behav. Brain Res.* 153, 497–506.
- Luna, R., Hernandez, A., Brody, C. D., and Romo, R. (2005). Neural codes for perceptual discrimination in primary somatosensory cortex. *Nat. Neurosci.* 8, 1210–1219.
- Machens, C. K., Romo, R., and Brody, C. D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 307, 1121–1124.
- MacMillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User’s Guide* (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.).
- McCarthy, D., and Davison, M. (1981). Towards a behavioral-theory of bias in signal-detection. *Percept. Psychophys.* 29, 371–382.
- Mehta, S. B., Whitmer, D., Figueroa, R., Williams, B. A., and Kleinfeld, D. (2007). Active spatial perception in the vibrissa scanning sensorimotor system. *PLoS Biol.* 5, e15. doi:10.1371/journal.pbio.0050015
- Mentzer, T. L. (1966). Comparison of three methods for obtaining psychophysical thresholds from the pigeon. *J. Comp. Physiol. Psychol.* 61, 96–101.
- Menzel, R., and Giurfa, M. (2001). Cognitive architecture of a mini-brain: the honeybee. *Trends Cogn. Sci. (Regul. Ed.)* 5, 62–71.
- Mountcastle, V. B., Steinmetz, M. A., and Romo, R. (1990). Frequency discrimination in the sense of flutter: psychophysical measurements correlated with postcentral events in behaving monkeys. *J. Neurosci.* 10, 3032–3044.
- Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Res.* 46, 2456–2464.
- Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature* 341, 52–54.
- Nienborg, H., and Cumming, B. (2010). Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron’s causality? *Curr. Opin. Neurobiol.* 20, 376–381.
- Nienborg, H., and Cumming, B. G. (2006). Macaque V2 neurons, but not V1 neurons, show choice-related activity. *J. Neurosci.* 26, 9567–9578.
- Nienborg, H., and Cumming, B. G. (2007). Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nat. Neurosci.* 10, 1608–1614.
- Nienborg, H., and Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron’s causal effect. *Nature* 459, 89–93.
- Ochoa, J., and Torebjork, E. (1983). Sensations evoked by intraneural microstimulation of single mechanoreceptor units innervating the human hand. *J. Physiol.* 342, 633–654.
- O’Connor, D. H., Clack, N. G., Huber, D., Komiyama, T., Myers, E. W., and Svoboda, K. (2010a). Vibrissa-based object localization in head-fixed mice. *J. Neurosci.* 30, 1947–1967.
- O’Connor, D. H., Peron, S. P., Huber, D., and Svoboda, K. (2010b). Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* 67, 1048–1061.
- Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621.
- Palmer, C., Cheng, S. Y., and Seidemann, E. (2007). Linking neuronal and behavioral performance in a reaction-time visual detection task. *J. Neurosci.* 27, 8122–8137.
- Parker, A. J., and Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* 21, 227–277.
- Polley, D. B., Heiser, M. A., Blake, D. T., Schreiner, C. E., and Merzenich, M. M. (2004). Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16351–16356.
- Pompilio, L., Kacelnik, A., and Behmer, S. T. (2006). State-dependent learned valuation drives choice in an invertebrate. *Science* 311, 1613–1615.
- Poulton, E. C. (1975). Range effects in experiments on people. *Am. J. Psychol.* 88, 3–32.
- Prigg, T., Goldreich, D., Carvell, G. E., and Simons, D. J. (2002). Texture discrimination and unit recordings in the rat whisker/barrel system. *Physiol. Behav.* 77, 671–675.
- Purushothaman, G., and Bradley, D. C. (2005). Neural population code for fine perceptual decisions in area MT. *Nat. Neurosci.* 8, 99–106.
- Qin, L., Liu, Y. C., Wang, J. Y., Li, S. N., and Sato, Y. (2009). Neural and behavioral discrimination of sound duration by cats. *J. Neurosci.* 29, 15650–15659.
- Resulaj, A., Kiani, R., Wolpert, D. M., and Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature* 461, 263–266.
- Riehle, A., Grun, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953.

- Roeder, K. D. (1966). Auditory system of noctuid moths. *Science* 154, 1515–1521.
- Roitman, J. D., and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22, 9475–9489.
- Romo, R., Brody, C. D., Hernandez, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473.
- Romo, R., and Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nat. Rev. Neurosci.* 4, 203–218.
- Rorie, A. E., Gao, J., McClelland, J. L., and Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS ONE* 5, e9308. doi:10.1371/journal.pone.0009308
- Sakitt, B. (1972). Counting every quantum. *J. Physiol.* 223, 131–150.
- Salinas, E., Hernandez, A., Zainos, A., and Romo, R. (2000). Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *J. Neurosci.* 20, 5503–5515.
- Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012.
- Schwarz, C., Hentschke, H., Butovas, S., Hais, F., Stüttgen, M. C., Gerdjikov, T. V., Bergner, C. G., and Waiblinger, C. (2010). The head-fixed behaving rat: procedures and pitfalls. *Somatosens. Mot. Res.* 27, 131–148.
- Shadlen, M. N., Britten, K. H., Newsome, W. T., and Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* 16, 1486–1510.
- Sherrington, C. (1940). *Man on His Nature*. Cambridge: Cambridge University Press.
- Shipley, E. F. (1960). A model for detection and recognition with signal uncertainty. *Psychometrika* 25, 273–289.
- Shlens, J., Field, G. D., Gauthier, J. L., Greschner, M., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2009). The structure of large-scale synchronized firing in primate retina. *J. Neurosci.* 29, 5022–5031.
- Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266.
- Smith, P. L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci.* 27, 161–168.
- Stewart, N., Brown, G. D. A., and Chater, N. (2005). Absolute identification by relative judgment. *Psychol. Rev.* 112, 881–911.
- Stüttgen, M. C. (2010). Toward behavioral benchmarks for whisker-related sensory processing. *J. Neurosci.* 30, 4827–4829.
- Stüttgen, M. C., Rüter, J., and Schwarz, C. (2006). Two psychophysical channels of whisker deflection in rats align with two neuronal classes of primary afferents. *J. Neurosci.* 26, 7933–7941.
- Stüttgen, M. C., and Schwarz, C. (2008). Psychophysical and neuro-metric detection performance under stimulus uncertainty. *Nat. Neurosci.* 11, 1091–1099.
- Stüttgen, M. C., and Schwarz, C. (2010). Integration of vibrotactile signals for whisker-related perception in rats is governed by short time constants: comparison of neurometric and psychometric detection performance. *J. Neurosci.* 30, 2060–2069.
- Stüttgen, M. C., Yildiz, A., and Güntürkün, O. (2011). Adaptive criterion setting in perceptual decision-making. *J. Exp. Anal. Behav.* 96, 155–176.
- Swets, J. A. (1959). Indices of signal detectability obtained with various psychophysical procedures. *J. Acoust. Soc. Am.* 31, 511–513.
- Swets, J. A. (1961a). Detection theory and psychophysics: a review. *Psychometrika* 26, 49–63.
- Swets, J. A. (1961b). Is there a sensory threshold? *Science* 134, 168–177.
- Swets, J. A. (1973). The relative operating characteristic in psychology: a technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science* 182, 990–1000.
- Swets, J. A., Tanner, W. P., and Birdsall, T. G. (1961). Decision-processes in perception. *Psychol. Rev.* 68, 301–340.
- Tanner, W. P. (1961). Physiological implications of psychophysical data. *Ann. N. Y. Acad. Sci.* 89, 752–765.
- Tanner, W. P., and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.* 61, 401–409.
- Teichert, T., and Ferrara, V. P. (2010). Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. *Front. Neurosci.* 4:186. doi:10.3389/fnins.2010.00186
- Tolhurst, D. J., Movshon, J. A., and Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* 23, 775–785.
- Treisman, M., and Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychol. Rev.* 91, 68–111.
- Treue, S., and Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579.
- Uka, T., and DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J. Neurosci.* 23, 3515–3530.
- Ulrich, R. (2010). DLs in reminder and 2AFC tasks: data and models. *Atten. Percept. Psychophys.* 72, 1179–1198.
- Ulrich, R., and Vorberg, D. (2009). Estimating the difference limen in 2AFC tasks: pitfalls and improved estimators. *Atten. Percept. Psychophys.* 71, 1219–1227.
- Vallbo, A. B., Olsson, K. A., Westberg, K. G., and Clark, F. J. (1984). Microstimulation of single tactile afferents from the human hand. Sensory attributes related to unit type and properties of receptive fields. *Brain* 107(Pt 3), 727–749.
- Verplanck, W. S., Collier, G. H., and Cotton, J. W. (1952). Nonindependence of successive responses in measurements of the visual threshold. *J. Exp. Psychol.* 44, 273–282.
- Vogels, R., and Orban, G. A. (1990). How well do response changes of striate neurons signal differences in orientation: a study in the discriminating monkey. *J. Neurosci.* 10, 3543–3558.
- Voigt, B. C., Brecht, M., and Houweling, A. R. (2008). Behavioral detectability of single-cell stimulation in the ventral posterior medial nucleus of the thalamus. *J. Neurosci.* 28, 12362–12367.
- von Heimendahl, M., Itskov, P. M., Arabzadeh, E., and Diamond, M. E. (2007). Neuronal activity in rat barrel cortex underlying texture discrimination. *PLoS Biol.* 5, e305. doi:10.1371/journal.pbio.0050305
- Wichmann, F. A., and Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313.
- Wichmann, F. A., and Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept. Psychophys.* 63, 1314–1329.
- Wolfe, J. M., and Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Curr. Biol.* 20, 121–124.
- Yeshurun, Y., Carrasco, M., and Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: tests of the difference model. *Vision Res.* 48, 1837–1851.
- Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron* 71, 9–34.
- Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 September 2011; paper pending published: 04 October 2011; accepted: 14 October 2011; published online: 10 November 2011.

Citation: Stüttgen MC, Schwarz C and Jäkel F (2011) Mapping spikes to sensations. *Front. Neurosci.* 5:125. doi: 10.3389/fnins.2011.00125

This article was submitted to *Frontiers in Decision Neuroscience*, a specialty of *Frontiers in Neuroscience*.

Copyright © 2011 Stüttgen, Schwarz and Jäkel. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.