Journal of Experimental Psychology:
Human Perception and Performance
2013, Vol. 39, No. 6, 1741–1762

0096-1523/13/$12.00 DOI: 10.1037/a0032103

A Signal Improvement to Signal Detection Analysis: Fuzzy SDT on the ROCs

James L. Szalma and Peter A. Hancock
University of Central Florida

Fuzzy Signal Detection Theory (FSDT) combines traditional Signal Detection Theory (SDT) with Fuzzy Set Theory to generalize signal detection analysis beyond the traditional categorical decision-making model. This advance upon SDT promises to improve measurement of performance in domains in which stimuli do not fall into discrete, mutually exclusive categories; a situation which characterizes many detection problems in real-world operational contexts. FSDT allows for events to simultaneously be in more than one state category (e.g., both signal and nonsignal). The present study derived FSDT Receiver Operating Characteristic (ROC) functions to test whether application of FSDT meets the Gaussian and equal variance assumptions of traditional SDT and, therefore, whether the standard representation of the SDT decision space can be extended to the broader case of FSDT. Results supported the contention that FSDT does meet these traditional SDT assumptions, and further, that it yields higher sensitivity scores than traditional SDT when the category membership of events is ambiguous. ROC analyses indicate that use of traditional SDT formulas with fuzzy hit and false alarm rates is thus justified. The implications of this advance to both theoretical and practical domains are adumbrated.

Keywords: Signal Detection Theory, Fuzzy Signal Detection Theory, ROC analysis

For more than half a century Signal Detection Theory (SDT; Green & Swets, 1966/1988; Swets, 1996; Tanner & Birdsall, 1958; Tanner & Swets, 1954) has provided perhaps the most useful analytical tool for evaluating human and machine performance in both simple and complex detection domains. As such, it has served simultaneously as one of the most important measurement tools and influential theories in all of behavioral research (Dember, 1998; Estes, 2002). The theory permits the independent evaluation of *perceptual sensitivity* and *response bias* (see Macmillan & Creelman, 2005). Perceptual sensitivity (often denoted by the term d') depends upon the perceptual ability of the observer to detect a signal or target or to discriminate signal from nonsignal events. Response bias (often denoted by the term β) represents the operator's decision criterion as to their propensity to say yes or no given the evidence to be evaluated.

Although SDT has been a critical achievement along our path to understanding decision-making performance, even preeminent theories have their limitations and SDT is no exception. In SDT the state of the world is forced into two distinct and mutually exclusive categories (i.e., signal vs. nonsignal; see Figure 1). However, this absolute division may not always represent an accurate depiction of the true state of the world. In many instances, events are sufficiently complex and/or perceptually ambiguous that they possess ongoing properties of both signal and nonsignal to varying degrees. It is important to note that this complexity does not result from low versus high signal strength (i.e., changes in the magnitude of the evidence variable) but rather a *change in the nature of the evidence variable itself*. That is, until absolute categorical identification has occurred (often after the fact), the signal itself may retain various nonsignal properties and vice versa. Indeed, it is such categorical (and often multidimensional) blending that induces at least some of the inherent stimulus-based uncertainty in decision-making in the first place. This circumstance is especially true of real-world operational settings.

Often, in such operational environments, characteristics of "targets" (e.g., dangerous items in luggage, markers for improvised explosive devices—IEDs, etc.) and "nontargets" (luggage with no prohibited or dangerous items, the absence of an IED) do not clearly meet the simple classification scheme of traditional SDT or the identification criteria prescribed by trainers or designers in their respective domains. For instance, the classical detection model assumes that items in the environment are either targets (e.g., weapons in a bag, IEDs) or nontargets. Items that are nontargets might have some degree of target value (e.g., sharp-edged, oblong metal object: such as a letter opener in baggage, changes in the environment such as new piles of trash not seen before or disturbed earth that can indicate the presence of an IED), but the observer has to force the observed object (or groups of objects) into just one of the two categories. Forcing ambiguous stimuli into two discrete categories can result in loss of threat information in real-world detection and decision-making. Such stimuli, which cannot be easily categorized into one group of only two predetermined object sets, often predominate in operational settings. Traditional signal detection theory is limited in that, in general, it is constrained to impose a discrete division on these natural continua of signal definition.

This article was published Online First April 15, 2013.

James L. Szalma and Peter A. Hancock, Department of Psychology, University of Central Florida.

Correspondence concerning this article should be addressed to James L. Szalma, Performance Research Laboratory, Psychology Department, PO Box 161390, University of Central Florida, Orlando, FL 32816. E-mail: James.Szalma@ucf.edu

State of the World	Response: Yes	Response: No
Signal Present (s=1)	Hit	Miss
Signal Absent (s=0)	False Alarm	Correct Rejection

Figure 1. 2 × 2 matrix of outcomes to stimulus events for traditional signal detection theory. *Note:* **s** = membership in the set (category) "signal."

Addressing such uncertainty, therefore, requires modification to the traditional "crisp" representations in SDT to allow for degrees of membership in the signal and nonsignal categories. Hancock, Masalonis, and Parasuraman (2000; see also Parasuraman, Masalonis, & Hancock, 2000) have articulated this more general case by combining SDT with elements of Fuzzy Set Theory (Zadeh, 1965). Here, category membership is not mutually exclusive and a stimulus event can, therefore, be simultaneously assigned to more than one category. Thus, in Fuzzy Signal Detection Theory (FSDT) a given stimulus, or more formally, a stimulus selected at random from a distribution of possible stimulus events, may be simultaneously categorized as *both* a target and a nontarget depending on the relative degrees of signal-like properties versus nonsignal properties that comprise the stimulus. For instance, a convenient range for a stimulus dimension is one in which the membership value of the stimulus varies from 0 (100% membership in the nonsignal category) to 1 (100% membership in the signal category). These end points correspond to the categories of traditional SDT. However, in FSDT values between 0 and 1 reflect different degrees of membership in the two categories. A signal value of .5 represents maximal uncertainty in the category membership status of the stimulus itself because a stimulus with a signal value of .5 has properties of both a nonsignal and a signal to an equal degree. Implicit in this model is the assumption that signal uncertainty exists not only within the observer (a major insight provided by traditional SDT) but also in the *state-of-the-world itself* (a major supposition of the present work). Both the SDT and FSDT models capture the uncertainty in the variation in magnitude of the evidence variable. However, through fuzzification, FSDT also captures the uncertainty inherent in the categorization of stimulus events and the appropriate responses to them.

In defining the state-of-the-world, it is important to note that the categorization (whether fuzzy or crisp) depends on the properties of an object or event that make it a signal or a nonsignal. For instance, in the context of IED detection, small piles of trash or debris have been used by insurgents to camouflage the explosives. The trash is, therefore, a feature of the category "signal." There may not be an IED present (from a FSDT perspective, this simply means that $\mathbf{s} \neq 1$), but one of the properties of such a threat is present (indicating that $\mathbf{s} \neq 0$). The state of the world is fuzzy in this instance because the stimulus to be evaluated has a nonzero degree of membership in the category "signal to be detected" ("signalness"). FSDT does not purport that signals to be detected may not emerge completely over successive observations (i.e., attain a state in which $\mathbf{s} = 1$), but only that there are instances in which some of the properties that define an event as a signal are present, and that the degree to which an event has these properties can be quantified rather than be (implicitly) treated as categorically equivalent to stimuli that possess no signal properties ($\mathbf{s} = 0$).

The Nature of Uncertainty in Signal Detection

One of the successful aspects of traditional SDT is its capacity for quantification of uncertainty in the evidence variable and the response of the observer to that uncertainty. There is also precedent in research on the traditional model for examining uncertainty in the signal itself. A case in which the signal is defined without ambiguity (e.g., how a positive test result is defined—a signal on an x-ray may be a spot with well-defined characteristic of size, shape, density, etc.) is referred to as a "signal known exactly" (SKE; Tanner & Birdsall, 1958). It is the SKE circumstance that is described in most textbook treatments of SDT. A case in which the magnitude of the evidence variable that comprises a signal is variable and, thus, is itself uncertain is referred to as a "signal known statistically" (SKS; e.g., a signal is a "spot" on an x-ray that may vary randomly from case to case in the magnitude of its size, shape, density, etc., but nevertheless belongs in the category "signal"). An SKS occurs when, in addition to shifting the mean of the noise distribution, addition of a signal also introduces random variability in the signal itself, which can increase the variability of the signal-noise distribution (Swets, 1996). However, in FSDT the uncertainty derives not from *random* variation in signal magnitude (as in the crisp SKS case), but from *the categorical and nonrandom definition of the signal itself.*

Mapping Functions

FSDT can be applied to all circumstances in which the state-of-the-world is fuzzy. For instance, in their formal description of the model, Parasuraman, Masalonis, and Hancock (2000) showed how conflict between commercial aircraft (in terms of relative distance in nautical miles) can map to fuzzy set membership to provide a more accurate definition of a "signal to be detected" than the traditional crisp definition (in which a signal is defined as a distance of 5 nautical miles or less; see Figure 2). As illustrated, the fuzzy mapping function permits degrees of membership in the set "signal" (the ordinate axis in the figure), permitting distances outside the traditional definition of a signal (e.g., 6 nautical miles) to have a high level of signal value. In contrast, in the traditional approach, also shown in the figure, any value above 5 would be considered a nonsignal (e.g., 6 miles would be considered equivalent to any other number of miles of separation greater than 5). The mapping function is the most crucial step in application of FSDT because it specifies the correspondence between a physical variable and its membership in the set "signal." The mapping function defines the state-of-the-world for FSDT analyses.

A mapping function is also generated for the set "response," but this can be operationalized using confidence ratings typically used in traditional SDT (Hancock, Masalonis, & Parasuraman 2000; Parasuraman et al., 2000). However, use of confidence ratings in SDT constitutes loss of information because the state of the world

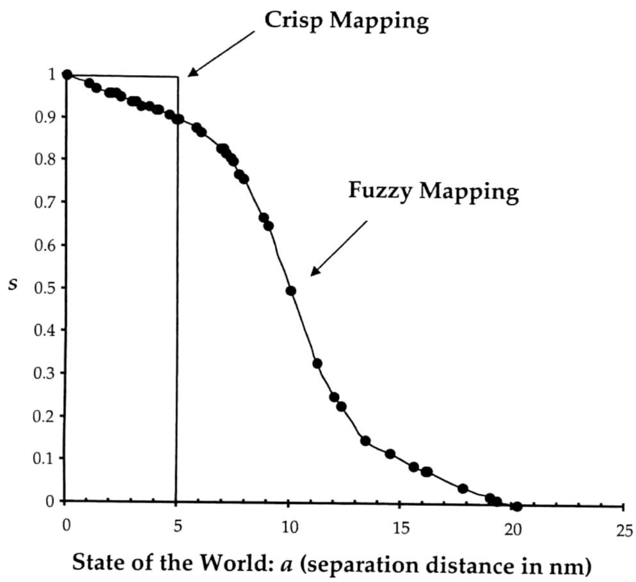

Figure 2. Mapping function showing the degree of membership in the set signal (**s**) as a function of a physical variable (a). From Parasuraman, Masalonis, and Hancock (2000).

is constrained to be organized into mutually exclusive categories, and because the confidence ratings are treated analytically as different criterion levels (i.e., crisp response categories), not different levels of judgment regarding the "signalness" of an event. In other words, the traditional analytic approach to the use of confidence intervals retains the crisp categorical response sets and may not necessarily reflect the confidence (fuzziness) of the response itself.

FSDT Procedures

The model and procedures for FSDT have been described in detail by Parasuraman et al. (2000) and Hancock et al. (2000) and, thus, are only briefly summarized here. Based on the fuzzy mapping function (e.g., Figure 2), the degree of membership in the set "signal" (**s**) is assigned for each stimulus presented, and the degree of membership in the set "yes response" (**r**) is assigned for each response by the observer (for examples see Table 1). Fuzzy membership in the sets hit (**H**), miss (**M**), false alarm (**FA**), and correct rejection (CR) are then assigned using the mixed implication functions shown in Equations 1–4. Parasuraman et al. (2000) reported that these functions were compared with others, but were the only ones found to not be "invalid for one reason or another" (p. 644). The minimum functions reflect the degree of overlap of the signal and response sets. The maximum functions reflect the degree of over- or underresponding for membership in the sets "false alarm" or "miss," respectively. Note that for computation of sensitivity and response bias only the implication functions for hits and false alarms are necessary.

$$H = \min(s, r) \quad (1)$$

$$M = \max(s\text{-}r,\ 0) \quad (2)$$

$$FA = \max(r\text{-}s,\ 0) \quad (3)$$

$$CR = \min(1\text{-}s,\ 1\text{-}r) \quad (4)$$

It is important to emphasize that Equations 1–4 do *not* produce proportions, but instead indicate fuzzy membership in the respective sets H, M, FA and CR (e.g., see Table 2). The proportions of correct detections, missed signals, false alarms, and correct rejections are computed by summing the fuzzy membership values over trials and dividing by the sum across trials of the fuzzy membership values for the set "signal." These computational procedures are shown in Equations 5–8.

$$p(H) = \sum(H_i)/\sum(s_i) \text{ for Trials } i = 1 \text{ to } N \quad (5)$$

$$p(M) = \sum(M_i)/\sum(s_i) \text{ for Trials } i = 1 \text{ to } N \quad (6)$$

$$p(FA) = \sum(FA_i)/\sum(1 - s_i) \text{ for Trials } i = 1 \text{ to } N \quad (7)$$

$$p(CR) = \sum(CR_i)/\sum(1 - s_i) \text{ for Trials } i = 1 \text{ to } N \quad (8)$$

Note that in their original presentation, Parasuraman et al. (2000) referred to these as "rates" (e.g., hit rate; HR, etc.). As Equations 5–8 yield proportions, we have clarified this meaning by replacing the original notations accordingly. This substitution should also serve to underscore that Equations 5–8 are *proportions*, but Equations 1–4 denote *degrees of categorical membership*.

Computational example. To illustrate this procedure, suppose that seven stimuli and seven responses are mapped to **s** and **r**, respectively, as shown in Table 1. In Table 2 are sample data for five stimulus presentations (from Participant 1 of Experiment 1 of the present work). Membership in the sets "hit," "miss," "false alarm," and "correct rejection" are determined using the mixed implication functions in Equations 1–4 and reproduced in Table 2.

Table 1
Membership in the Sets "Signal" (s) and "Response" (r) for Each of Seven Categories of Stimulus and Response

Stimulus category	s	Response category	r
1	0	1	0
2	0.1667	2	0.1667
3	0.3333	3	0.3333
4	0.5	4	0.5
5	0.6667	5	0.6667
6	0.8333	6	0.8333
7	1	7	1

Table 2
Example Computation of Fuzzy Outcome Measures

Trial	s	r	H = min(s,r)	M = max(s-r,0)	FA = max(r-s,0)	CR = min(1-s,1-r)
1	0.1667	0.3333	0.1667	0	0.1666	0.6667
2	0.8333	0.6667	0.6667	0.1666	0	0.1667
3	1	0.5	0.5	0.5	0	0
4	0.1667	0.3333	0.1667	0	0.1666	0.6667
5	0.6667	0.3333	0.3333	0.3334	0	0.3333

Computation of the corresponding proportions is accomplished by summing the membership values over trials and also determining the sum of s and 1-s over those trials using the formulae in Equations 5–8. Thus,

$$p(H) = \sum (H_i)/\sum (s_i) = (0.1667 + 0.6667 + 0.5 + 0.1667 + 0.3333)/(0.1667 + 0.8333 + 1 + 0.16667 + 0.6667)$$

$$p(H) = 0.65$$

$$p(M) = \sum (M_i)/\sum (s_i) = (0 + 0.1666 + 0.5 + 0 + 0.3334)/(0.1667 + 0.8333 + 1 + 0.16667 + 0.6667)$$

$$p(M) = 0.35$$

$$p(FA) = \sum (FA_i)/\sum (1\text{-}s_i) = (0.1666 + 0 + 0 + 0.16666 + 0)/[(1 - 0.1667) + (1 - 0.8333) + (1 - 1) + (1 - 0.1667) + (1 - 0.6667)]$$

$$p(FA) = 0.15$$

$$p(CR) = \sum (CR_i)/\sum (1\text{-}s_i) = (0.1667 + 0.1667 + 0 + 0.1667 + 0.3333)/[(1 - 0.1667) + (1 - 0.8333) + (1 - 1) + (1 - 0.1667) + (1 - 0.6667)]$$

$$p(CR) = 0.85$$

Application of SDT Procedures to Fuzzy Hit and False Alarm Proportions

In their original exposition, Parasuraman et al. (2000) argued that the established SDT procedures for computing sensitivity and response bias could be applied to the fuzzy hit and false alarm rates. This, they argued, is because "the fuzziness of the signal has already been captured in the definition of **s** and **r**" (p. 649). That is, the fuzzy hit and false alarm rates computed from FSDT procedures can be treated computationally as equivalent to the corresponding crisp rates.

However, we now propose that this equivalence may not be justified. Computation of sensitivity and bias rests on statistical assumptions regarding the decision space underlying the data. These assumptions may be briefly summarized as (a) noise is always present in any detection system, and it may be represented as a normal distribution; (b) signals are always embedded in noise, and the presence of a signal shifts the distribution (i.e., addition of a signal to the noise increases the magnitude of the evidence variable) but does not change its form (i.e., the variance of the noise distribution is equal to that of the signal + noise distribution), so that the distance in standard deviation units between the means of the noise and signal + noise distributions constitutes perceptual sensitivity (d'); (c) the observer establishes a criterion for responding affirmatively regarding the presence of a signal, and the placement of this criterion (response bias) is independent of the distance between the means of the two distributions (sensitivity). These assumptions manifest in the decision space shown in Figure 3.

There is an implicit assumption in FSDT that the decision space underlying FSDT data is identical to that of SDT, in spite of the fact that the state-of-the-world is explicitly assumed to be different. We can assert that it certainly *may* be the case that the assumptions underlying SDT, and represented in the decision space, also apply to FSDT, but crucially, this proposition needs to be empirically evaluated. Crisp SDT data are derived differently than FSDT data and these computational differences may potentially cause FSDT methods to fail to accurately represent the decision space shown in Figure 3. That is, it is possible that fuzzy hit and false alarm rates may be represented best by a decision space that differs from that shown in Figure 3. In crisp SDT each presentation of an event (a "trial") consists of either a signal or a nonsignal, and these categories are mutually exclusive. The computation of hit and false alarm rates results from repetitions of such presentation, which is necessary because SDT is a statistical model. FSDT is also a statistical model (as it is an extension of traditional SDT), but responses to mutually exclusive events are not summated, nor is the sum divided by the total number of signals or nonsignals presented. Rather, it is the sum (over trials) of the degree of membership in the category "hit" or "false alarm," divided by the degree of membership in the categories signal and nonsignal summed over the number of presentations, respectively (see Equations 5 and 7). It is not necessarily the case that the fuzzy signal and noise distributions are identical in form to those of crisp signal and noise distributions. However, if they are, then the recommendation of Parasuraman et al. (2000) regarding the use of standard SDT formulae would be warranted. If FSDT procedures produce results that do not reflect the underlying (assumed) decision space, then alternative measures to d' and β, which are based on these assumptions, should be developed. Fortunately, traditional SDT provides a methodology for evaluating the statistical assumptions of the model: Receiver Operating Characteristics (ROC; sometimes referred to as "Relative" Operating Characteristics; see Swets, 1973, 1996).

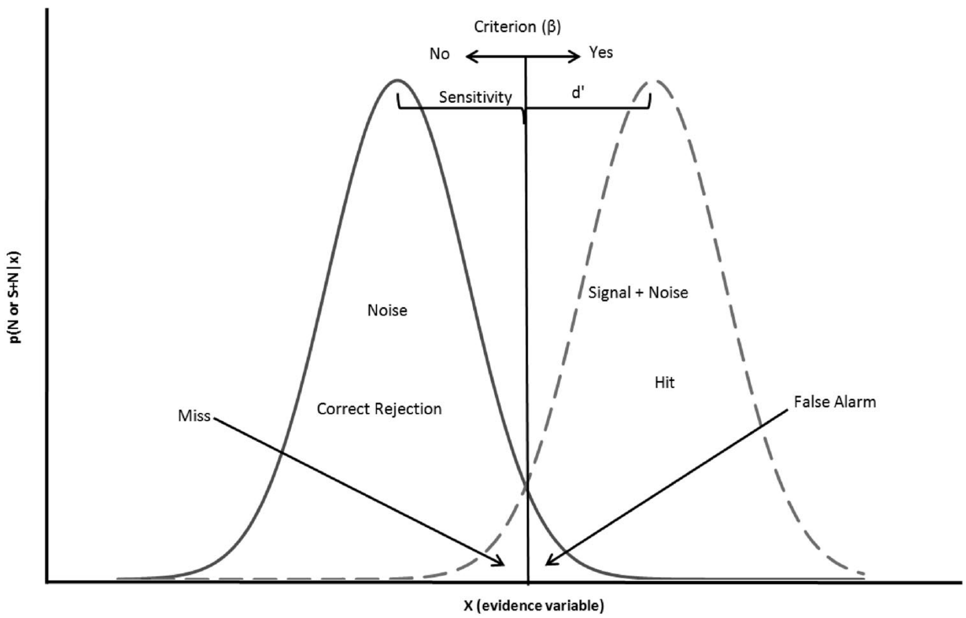

Figure 3. Illustration of the decision space assumed by traditional signal detection theory.

ROC Analysis

ROC analysis provides a methodology for determining the behavior of diagnostic systems (Swets, 1996; Swets & Pickett, 1982). Pairs of hit and false alarm rates are obtained at different levels of response criterion, either through a manipulation of instructions (e.g., a payoff matrix), of signal probability, or the use of confidence ratings (for more detailed descriptions of these methods, see Green & Swets, 1966/1988; Macmillan & Creelman, 2005). These data points comprise the basis for estimating the change in correct detections as a function of the false alarm rate. When the hit and false alarm rates are converted to z-score form, a good fit of the resulting function to a linear model indicates that the data are consistent with normally distributed noise and signal + noise distributions. If the slope of the function is one, the equal variance assumption is considered to be tenable. Finally, the statistical independence of sensitivity and response bias is indicated by manipulation of task difficulty resulting in parallel lines in the z-score form of the ROC. Cases in which the equal variance assumption is tenable also provide evidence for the independence of sensitivity and response bias. Independence can obtain when the equal variance assumption is violated, but in such cases the nonunit slopes of the two functions must be statistically equivalent to one another. Nevertheless, if the slopes are different from unity, then use of d' is inadvisable (Macmillan & Creelman, 2005; Swets, 1996).

Utility of fuzzy ROC analysis. Although a crucial test for FSDT is whether it is an improvement over traditional SDT methods because of the fuzzification of stimulus and response categories, such an evaluation depends on the validity of the FSDT procedures themselves. Thus, the key questions for FSDT are both whether and how the incorporation of fuzzy logic into the derivation of hit and false alarm rates influences whether such data support the assertion that the assumptions of traditional SDT extend to FSDT. That is, determination of whether the Gaussian and equal variance assumptions of traditional SDT extend to FSDT would provide evidence regarding the nature of the underlying decision space for the latter, which has implications for the computations of fuzzy sensitivity and fuzzy response bias. The reason that testing SDT assumptions is important is that if the application of FSDT results in ROC functions consistent with the model shown in Figure 3, it implies that (a) there is a common form of distributions in the decision spaces of FSDT and SDT; and (b) the use of standard formulae, which are based on assumptions regarding the decision space illustrated in the figure, are acceptable for fuzzy hit and false alarm rates. That is, if the assumptions of traditional SDT are also met by FSDT, then the standard formulae may be used for computing FSDT indices.

Evidence for FSDT

Using Monte Carlo methods, Szalma and O'Connell (2011) generated simulated SDT data and analyzed these data using both traditional and fuzzy methods. They applied the example mapping function for aircraft separation described in Parasuraman et al. (2000). They reported that when data are generated that conform to the statistical assumptions of SDT, applying FSDT methods resulted in ROC function that reflected this underlying structure. In other words, application of FSDT procedures yielded functions that accurately reflected the (known) underlying decision space. A purpose for the present work was to determine whether FSDT reflects the same underlying decision space using human observers engaged in a perceptual decision-making task.

Empirical Tests of FSDT

The true test for any theory lies in how well it describes empirical data and to what degree it differs from, improves on, or replaces existing theory. Initial investigations comparing FSDT and traditional SDT were reported by Masalonis and Parasuraman (2003), who used an air traffic control task and assigned fuzzy membership values to stimulus and response sets. They reported that the fuzzy false alarm rate was lower than the crisp false alarm rate computed from the same data. However, at that time Masalonis and Parasuraman (2003) were constrained to use data from previous experiments that were somewhat amenable to fuzzy SDT analysis but that were not designed to explicitly test this new theory. In addition, they could not evaluate the decision space for FSDT because they could not construct fuzzy ROC functions. What is therefore needed are studies which are explicitly designed to evaluate FSDT. The experiments reported here represent one of the few purpose-specific, empirical tests of FSDT using a perceptual discrimination task (but see Murphy, Szalma, & Hancock, 2004). The purpose for the ROC analyses was to elucidate the structure of the decision space underlying FSDT and to compare it with the decision space of the traditional SDT model applied to the same data but with crisp rather than fuzzy categorization.

We hypothesized that if application of standard formulas for sensitivity and bias to fuzzy hit and false alarm rates is valid, then applying common procedures to a FSDT task using the ROC paradigm should yield linear functions (in z-score form) of unit slope. In addition, variations in task difficulty should result in parallel functions (equivalent slopes) in which the line corresponding to lower sensitivity (a more difficult discrimination task) should be lower (i.e., closer to the origin) than the line for an easier version of a task.

The first experiment in the present series was designed to explore this form of fuzzy ROC space and evaluate the SDT assumptions at two levels of discrimination difficulty. These latter levels were operationalized as the magnitude of the differences between stimulus categories of a temporal discrimination task. This experiment represents a replication and extension of an earlier study by Murphy, Szalma, and Hancock (2004) who employed a temporal discrimination task in which a 200 ms stimulus was defined as a nonsignal (i.e., $\mathbf{s} = 0$), and six durations, in increments of either 20 ms or 80 ms, were defined as increasing degrees of signal membership ("signalness"). The two increment intervals (Δ) comprised a manipulation of task difficulty. Murphy et al. (2004) reported that for a stimulus magnitude interval of $\Delta = 20$ ms (designated the more difficult condition) both traditional and fuzzy SDT analyses indicated that the equal variance assumption was met, and for both difficulty conditions the data yielded ROC functions consistent with the assumption of Gaussian distributions. In addition, the FSDT analysis generally resulted in higher sensitivity estimates than traditional SDT analyses.

Murphy et al. (2004) interpreted their results in terms of the differences in categorization in the two analytic methods. In SDT, a missed signal or a false alarm is ubiquitously an all-or-none event. In contrast, in FSDT partially correct detections or rejections are possible. For instance, if on a given trial the observer's response belongs to the category "false alarm" to a degree of 0.7 and to the category "hit" 0.3, then in a crisp analysis (in which the data are forced into the 2×2 matrix shown in Figure 1) the response to this trial would be considered a false alarm. In FSDT, however, it is considered to be both a false alarm and a correct detection to differing degrees. Murphy et al. (2004) concluded that in many instances (i.e., those that do not naturally conform to the 2×2 matrix representation shown in Figure 1) traditional SDT underestimates the sensitivity of observers.

Murphy et al. (2004) reported that a discrimination difficulty manipulation resulted in an inconsistent pattern of results across participants and analyses (SDT vs. FSDT). As they noted, this result may have been due to the intrinsic covariation of stimulus range with the magnitude of the interval between stimulus categories. The range of temporal duration categories for the $\Delta = 20$ ms condition was narrower than that for the $\Delta = 80$ ms condition. Thus, the present work sought to evaluate the differential effects of interval magnitude and stimulus range by first replicating the experiment of Murphy et al. (2004), and then in a second experiment, evaluating controlled variations in intercategory intervals ($\Delta = 20$ ms vs. $\Delta = 80$ ms), range of stimuli (7 vs. 24 categories), and range of response set options (7 categories vs. a binary response). The latter manipulation allowed for comparisons under conditions favorable to each of the two analytic methods, that is, FSDT (a 7-category response set) versus SDT (a binary response set).

Experiment 1

Method

In the first experiment we derived ROC functions for a temporal discrimination task using multiple categories of stimulus and response sets. Temporal discriminations were selected because the power law exponent for this class of discrimination is generally agreed to be close or equivalent to unity (i.e., the psychophysical function is linear; see Stevens, 1961; but see Hancock, 2013). This simplifies the application of FSDT to the perceptual dimension by allowing for equal-interval increments in duration across the categories to be discriminated. Response bias was manipulated using a payoff matrix which encouraged conservative, lenient, or unbiased ("neutral") responding, respectively. In addition, two levels of discrimination difficulty were employed, which were manipulated by changing the differences in duration between stimulus categories.

Participants. Six students from the University of Central Florida (three men and three women) participated in this study for monetary compensation. They were treated according to the APA guidelines on participation and all procedures were reviewed and approved by the internal Institutional Review Board (IRB).

Stimuli. The stimulus employed in this experiment consisted of a 6×6 cm light gray square with a black surround. The square was presented at seven different durations depending on the difficulty level. For the "less difficult" condition, seven durations were used ranging from 200 ms to 680 ms, with a difference of 80 ms between each category and the category below or above it ($\Delta = 80$ ms condition). In the "more difficult" condition the seven durations ranged from 200 ms to 320 ms, with a difference of 20 ms between adjacent categories ($\Delta = 20$ ms condition). Participants were not explicitly informed of either the number of stimulus categories or of the specific durations presented.

Procedure. On the first day, participants received instructions and completed two 40-min practice sessions on the task. One session was designated for each difficulty level of the task ($\Delta = 20$ ms and $\Delta = 80$ ms). The order in which the participants received the difficulty levels was counterbalanced. For the practice and experimental sessions, each condition was divided into four, 10-min blocks of trials, with a brief rest interval between each block.

On the subsequent 3 days of the experiment, participants engaged in a total of six, 40-min detection tasks (each preceded by a 5-min warm-up session) consisting of the two temporal difficulty levels ($\Delta = 20$ ms and $\Delta = 80$ ms) at three levels of instruction bias (lenient, neutral, and conservative). Participants received the treatments over 3 days during a 2-week period. Each day they received one of the bias manipulations and each difficulty level. This was done in order to avoid problems with asking observers to change their response criterion (i.e., as a result of the payoff matrix) in the middle of an experimental session. The order in which the participants received the different treatments was counterbalanced such that each participant received one of the six possible combinations of bias condition. Within each bias condition, three of the participants received the difficult task ($\Delta = 20$) first; the other three received the easy task ($\Delta = 80$) first.

The task required the participants to judge the relative duration that the gray square appeared on the screen. Participants were instructed to respond to stimuli by rating the degree to which each stimulus was shorter in duration (a lower level of "signalness") versus longer in duration (a higher level of "signalness") by pressing keys 1 through 7 on a computer keyboard, with the response "1" indicating that the stimulus was a "full nonsignal" and the response "7" indicating that the stimulus was a "full signal." For each condition the event rate was 21 events/min, and each of the seven different stimuli was presented 72 times during each session. The order in which the stimuli were presented within each condition was randomized. Note that an odd number of stimulus categories was selected in order to evaluate the influence of a middle category on the crisp SDT analysis, as this category can be viewed as the maximum level of uncertainty in the stimulus (**s**) or, subjectively, in the observer (**r**).

For the manipulation of response bias, three instructional sets were employed to induce lenient, unbiased, and conservative responding (see Appendix). During the task with the lenient instruction set, participants were informed that they would receive (+10) points for each correct identification, defined as a correct rating of the duration the square was on the screen. However, they were told that they would lose (−10) points for each missed signal, which was defined as an underestimation of the duration the square appeared on the screen. Finally, they were informed that they would be penalized (−1) point for each false alarm, which was defined as an overestimation of the duration of the square. In the unbiased ("neutral") conditions participants were informed that they would receive (+1) point for each correct identification. However, they were told that they would be penalized (−1) point for each missed signal and that they would be penalized a (−1) point for each false alarm. In the conservative conditions participants were informed that they would receive (+10) points for each correct identification, and were told that they would be penalized (−1) point for each missed signal and (−10) points for each false alarm.

Results

Current algorithms for estimation of ROC curves require provision of frequencies of response categories (e.g., ROCKIT; Metz, 1998; FitROC; Wickens, 2002). However, these programs do not permit application of fractional frequencies such as those associated with an FSDT analysis. Due to this constraint of available estimation procedures, fuzzy ROC functions were therefore estimated by rounding the frequencies to the nearest whole number and entering them into FitROC: Parameter Estimation for Gaussian Signal Detection Model (Wickens, 2002). Although some precision is lost in rounding, the frequencies obtained are fuzzy because they were derived using the mixed implication functions (equations 1–4). Version 2.1 of FitROC provides (a) the intercept and slope for the z-score form of the ROC ($z_H = a + b\, z_F$); (b) a χ^2 test of how well an ROC curve fits the Gaussian model; (c) estimates of perceptual sensitivity ($A_z/d'/d_a$); and (d) estimates of criterion (λ) and response bias ($\ln\beta$). The program also provides standard errors for each parameter estimated. Difficulty conditions were compared using z tests as described in Wickens (2002). The quality of ROC model fit was evaluated using the χ^2 goodness-of-fit test as also recommended by Wickens (2002). A statistically significant ($p < .05$) χ^2 test is interpreted as an indication of poor model fit. The FitROC software provides this test for both the equal variance and unequal variance models. The degrees of freedom for this test are the number of cells in a stimulus type (t) by response category (r) matrix minus the number of parameters (p) to be estimated, as shown in Equation 9:

$$\mathrm{df} = \big(\mathrm{t}(\mathrm{r}\text{-}1)\text{-}\mathrm{p}\big) \tag{9}$$

Note that the procedures for calculating parameters were the same for both FSDT and traditional SDT analyses. The two models differ in how hit and false alarm rates are calculated. The formulae for computing FSDT hits and false alarm proportions (see Equations 1–8) were obtained from Parasuraman et al. (2000), and the method for computing crisp hit and false alarm rates followed standard procedures (e.g., Macmillan & Creelman, 2005; Wickens, 2002). The data for the latter analyses were simplified by bifurcating the seven categories in two ways; either with the middle category considered a "nonsignal" or with it considered as a "signal." The mapping functions for the fuzzy and crisp analyses are shown in Figure 4. The goodness of fit (χ^2), intercept (i.e., "a"), slope (i.e., "b"), perceptual sensitivity, and response bias estimates for both the traditional SDT and FSDT methods for the $\Delta = 20$ ms and $\Delta = 80$ ms are reported in Tables 3 (Participants 1 and 2), 4 (Participants 3 and 4), and 5 (Participants 5 and 6).

Traditional SDT analysis. In the $\Delta = 20$ condition results depended on both the observer and the way in which the data were bifurcated. Thus, when the middle stimulus category was considered a "nonsignal" the equal variance model did fit for three observers (3, 4, and 5; see Tables 4 and 5 and Figures 5 and 6), yet the unequal variance model fitted for observer 2 (see Table 3). In the case of the latter participant, the slope of the function was greater than one (i.e., b = 1.181), indicating that $\sigma_s < \sigma_n$. This is because in its most general form the ROC function is,

$$z_H = \left(\frac{\sigma_N}{\sigma_S}\right) z_F + a \tag{10}$$

If one sets the noise distribution to be of standard form ($\sigma_N = 1$), then

$$z_H = \left(\frac{1}{\sigma_S}\right) z_F + a \tag{11}$$

Note that when the equal variance assumption is met, a = d' $\sigma_s = 1$, and the familiar equation $d' = z_H - z_F$ can be derived.

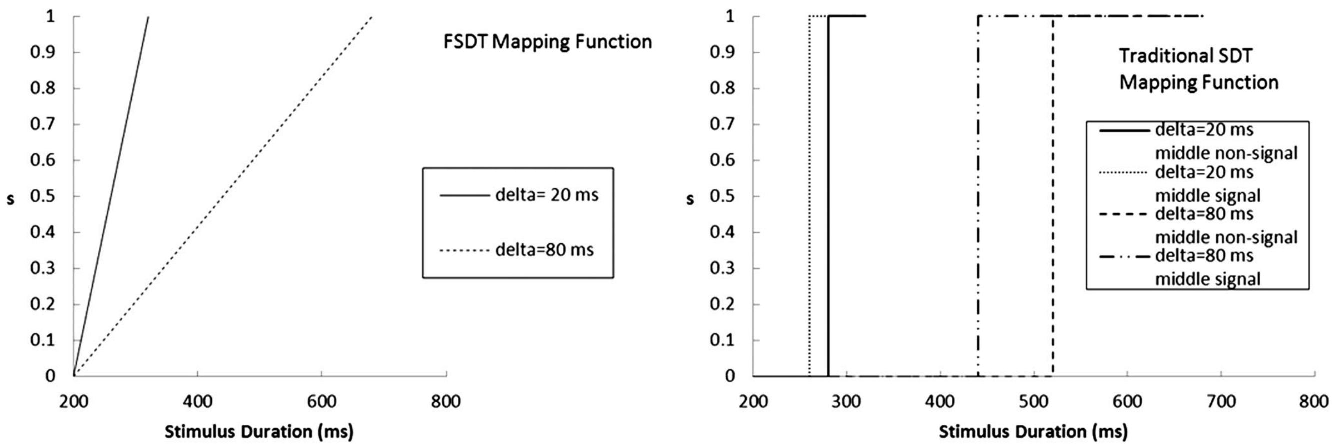

Figure 4. FSDT and traditional SDT mapping functions for Experiment 1.

For the remaining two observers (1 and 6) neither the equal nor the unequal variance model fitted the data when the middle category was classified as a nonsignal. Sensitivity scores (A_z) for the other four observers ranged from .593 to .709 (M = .652). The corresponding range for d' (d_a for Participant 2) was .334 to .780 (M = .558).

When the middle stimulus was coded as a "signal" the (same) data for Observer 1 fit the equal variance model (although the unequal variance model did not fit the data; see Table 3 and Figure 5). Observers whose data fit the equal variance assumption when the middle category was coded as a nonsignal also fit that model when the middle stimulus was coded as a signal, although for

Table 3
Experiment 1: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participants 1 and 2 (Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
Participant 1 (Female)							
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as nonsignal							
6.720[n] ($p < .05$)	.603 (.033)	.369 (.123)	.371 (.143)	1.008 (.122)	.479 (.100)	.545 (.100)	.135 (.144)
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as signal							
5.598[e] ($p = .061$)[*]	.601 (.015)	.361 (.054)	.361 (.054)	1.00	.160 (.031)	.267 (.046)	−.116 (.023)
Fuzzy SDT Analysis Δ = 20 ms condition							
.563[e] ($p > .75$)	.783 (.012)	1.107 (.056)	1.107 (.056)	1.00	.518 (.061)	.659 (.068)	−.043 (.051)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as nonsignal							
.521[e] ($p > .75$)	.713 (.015)	.795 (.063)	.795 (.063)	1.00	.934 (.092)	1.116 (.109)	.148 (.037)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as signal							
1.740[e] ($p > .40$)	.753 (.013)	.967 (.057)	.967 (.057)	1.00	.476 (.058)	.709 (.070)	−.361 (.050)
Fuzzy SDT Analysis Δ = 80 ms condition							
1.468[e] ($p > .47$)	.862 (.009)	1.539 (.059)	1.539 (.059)	1.00	.762 (.088)	1.016 (.096)	−.171 (.075)
Participant 2 (Male)							
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as nonsignal							
2.779[u] ($p = .096$)	.709 (.018)	.780 (.075)	.853 (.079)	1.181 (.111)	1.069 (.664)	.246 (.097)	−.851 (.112)
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as signal							
.467[e] ($p > .75$)	.691 (.019)	.707 (.076)	.707 (.076)	1.00	1.257 (.160)	−.278 (.043)	−1.294 (.160)
Fuzzy SDT Analysis Δ = 20 ms condition							
1.692[e] ($p > .42$)	.821 (.013)	1.298 (.073)	1.298 (.073)	1.00	2.252 (.186)	−.144 (.063)	−1.506 (.132)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as nonsignal							
27.612[n] ($p < .001$)	.746 (.018)	.937 (.077)	.781 (.083)	.624 (.076)	2.119 (.324)	−.144 (.135)	−.580 (.115)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as signal							
59.355 [n] ($p < .001$)	.826 (.011)	1.329 (.060)	1.059 (.062)	.519 (.063)	.781 (.163)	−.860 (.131)	−1.263 (.093)
Fuzzy SDT Analysis Δ = 80 ms condition							
18.366[n] ($p < .001$)	.849 (.013)	1.459 (.079)	1.130 (.082)	.447 (.067)	1.384 (.248)	−.677 (.167)	−1.237 (.127)

Note. C = Conservative; N = unbiased; L = Lenient; [u] data fits the unequal variance model; [e] data fits the equal variance model; [n] neither model fits; values presented correspond to the unequal variance model; CF = neither model fit, and a convergence failure was obtained for the unequal variance case; [*] although the equal variance model fit marginally, the unequal variance model did not fit ($\chi^2(1) = 5.028$, $p = .025$); [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

Table 4
Experiment 1: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participants 3 and 4 (Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
Participant 3 (M)							
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as nonsignal							
.526[e] ($p > .75$)	.593 (.018)	.334 (.065)	.334 (.065)	1.00	.703 (.143)	.273 (.056)	−.166 (.038)
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as signal							
.012[e] ($p > .99$)	.611 (.016)	.397 (.060)	.397 (.060)	1.00	.311 (.054)	.084 (.023)	−.579 (.093)
Fuzzy SDT Analysis Δ = 20 ms condition							
1.748[e] ($p > .41$)	.789 (.012)	1.136 (.057)	1.136 (.057)	1.00	.728 (.071)	.346 (.058)	−.509 (.063)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as nonsignal							
12.351[n] ($p < .01$)	.545 (.032)	.161 (.113)	.134 (.098)	.611 (.079)	1.296 (.245)	.155 (.134)	−.407 (.144)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as signal							
7.708[n] ($p < .01$)	.694 (.013)	.719 (.053)	.574 (.046)	.526 (.062)	−.049 (.114)	−.633 (.123)	−.859 (.075)
Fuzzy SDT Analysis Δ = 80 ms condition							
2.202[u] ($p > .10$)	.793 (.021)	1.156 (.102)	.914 (.123)	.501 (.128)	.086 (.249)	−.314 (.266)	−.815 (.240)
Participant 4 (F)							
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as nonsignal							
1.14[e] ($p > .55$)	.679 (.014)	.657 (.054)	.657 (.054)	1.00	.074 (.030)	.186 (.034)	.426 (.047)
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as signal							
5.323[e] ($p = .070$)*	.697 (.013)	.731 (.053)	.731 (.053)	1.00	−.302 (.040)	−.270 (.039)	−.023 (.033)
Fuzzy SDT Analysis Δ = 20 ms condition							
.324[u] ($p > .56$)	.800 (.012)	1.193 (.058)	1.110 (.218)	.856 (.351)	−.317 (.401)	−.188 (.412)	.071 (.401)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as nonsignal							
CF							
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as signal							
CF							
Fuzzy SDT Analysis Δ = 80 ms condition							
CF							

Note. C = Conservative; N = unbiased; L = Lenient; [u] data fits the unequal variance model; [e] data fits the equal variance model; [n] neither model fits; values presented correspond to the unequal variance model; * although the equal variance model fit marginally, the unequal variance model did not fit ($\chi^2(1) = 4.227$, $p = .040$). [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

Participants 4 and 5 the fit was associated with a failure of the more general, unequal variance model to fit. In addition, when the middle stimulus was coded as a signal, the data for Observer 2 the equal variance model fit. By contrast, the other categorization of the middle stimulus resulted in the unequal variance model fit for that participant. Note that for the Δ = 20 condition both instances in which neither model fit the data occurred when the middle category was classified as a nonsignal.

For the Δ = 80 condition, the two categorization schemes were more consistent within observers (see Tables 3, 4, and 5 and Figures 5 and 6). Regardless of whether the middle category was coded as a signal or nonsignal, the equal variance assumption was met by the data of Observer 1 and the unequal variance assumption was met by the data of Observers 5 and 6. Neither model fit for Observers 2 and 3; convergence failures were observed for Participant 4. Note, however, that for the latter individual, no convergence failures were observed in the fuzzy analysis of the same data, as we show in the subsequent section on FSDT analysis directly. This indicates that the failure to converge resulted from the procedures used to categorize the data for crisp analysis rather than the pattern of responding of this latter participant per se. In general, model fits were better in the Δ = 20 condition than in the Δ = 80 condition. In the case of the latter, the way in which the middle category was classified had little effect on model fit. Thus, the effect of the middle category depended on difficulty level and affected whether the Gaussian model fit the data, and, if so, whether the fit was for the equal variance or the unequal variance model.

Comparison of difficulty conditions using traditional SDT. Comparisons of the two difficulty conditions were accomplished by examining the equal variance models for Participant 1 and the unequal variance models for Participants 5 and 6. Note that comparisons were computed only in cases where either the equal variance or the unequal variance models fit for *both* difficulty conditions. Such comparisons were not possible for Participants 2, 3, or 4. For Participants 1 and 6 the comparison was for the middle category as signal, because neither model fit when that category was classified as a nonsignal. For Participant 5 comparisons could be computed for each classification of the middle stimulus.

When the middle stimulus was coded as a nonsignal, the sensitivity of Observer 5 was higher in the Δ = 80 than in the Δ = 20 condition ($z = 8.09$; see Table 5 and Figure 6). The same pattern was observed for this participant when the middle category was coded as a signal ($z = 6.38$; see Table 5 and Figure 6). In addition, when the middle category was coded as a signal, higher sensitivities were associated with the Δ = 80 condition compared to the Δ = 20 condition for Participants 1 ($z = 7.66$; see Table 3 and Figure 5) and 6 ($z = 2.62$; see Table 5 and Figure 6).

FSDT analysis. For the purpose of the present work, perhaps the most important analyses concerned the FSDT interpretations of the recorded data. Here, in the Δ = 20 condition, the fuzzy data for each observer fit the Gaussian equal variance model with the

Table 5

Experiment 1: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participants 5 and 6 (Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
Participant 5 (M)							
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as nonsignal							
5.313[e] (p = .070)*	.628 (.015)	.463 (.057)	.463 (.057)	1.00	.545 (.073)	.346 (.049)	−.016 (.021)
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as signal							
3.449[e] (p > .10)#	.622 (.015)	.440 (.056)	.440 (.056)	1.00	.296 (.046)	.151 (.030)	−.357 (.050)
Fuzzy SDT Analysis Δ = 20 ms condition							
.747[e] (p > .68)	.790 (.011)	1.14 (.055)	1.14 (.055)	1.00	.563 (.064)	.367 (.059)	−.294 (.056)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as nonsignal							
.225[u] (p > .60)	.794 (.014)	1.16 (.070)	1.716 (.281)	1.838 (.291)	1.227 (.091)	1.040 (.164)	.607 (.175)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as signal							
.003[u] (p > .95)	.753 (.014)	.967 (.063)	1.197 (.109)	1.437 (.207)	.628 (.125)	.339 (.149)	−.560 (.123)
Fuzzy SDT Analysis Δ = 80 ms condition							
1.443[e] (p > .48)	.850 (.009)	1.464 (.057)	1.464 (.057)	1.00	.508 (.076)	.103 (.070)	−.433 (.079)
Participant 6 (F)							
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as nonsignal							
17.435[n] (p < .001)	.699 (.021)	.737 (.085)	.787 (.140)	1.130 (.150)	.634 (.114)	.690 (.102)	.109 (.134)
Traditional SDT Analysis Δ = 20 ms condition. Middle stimulus coded as signal							
3.679[u] (p = .055)	.673 (.017)	.633 (.067)	.701 (.063)	1.207 (.173)	.148 (.146)	.235 (.139)	−.476 (.098)
Fuzzy SDT Analysis Δ = 20 ms condition							
3.628[e] (p > .15)+	.817 (.010)	1.280 (.055)	1.280 (.055)	1.00	.242 (.062)	.293 (.063)	−.311 (.063)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as nonsignal							
.081[u] (p > .75)	.610 (.026)	.393 (.094)	.321 (.084)	.576 (.064)	.840 (.147)	.614 (.134)	−.562 (.106)
Traditional SDT Analysis Δ = 80 ms condition. Middle stimulus coded as signal							
.461[u] (p > .49)	.729 (.013)	.864 (.058)	.686 (.053)	.511 (.054)	−.089 (.120)	−.081 (.120)	−.980 (.058)
Fuzzy SDT Analysis Δ = 80 ms condition							
.703[u] (p > .40)	.804 (.018)	1.209 (.092)	.929 (.096)	.428 (.087)	.059 (.217)	−.105 (.219)	−1.128 (.171)

Note. C = Conservative; N = unbiased; L = Lenient; [u] data fits the unequal variance model; [e] data fits the equal variance model [n] neither model fits; values presented correspond to the unequal variance model; CF = neither model fit, and a convergence failure was obtained for the unequal variance case; * although the equal variance model fit marginally, the unequal variance model did not fit ($\chi^2(1) = 3.831$, $p = .050$). # although the equal variance model fit, the unequal variance model was marginal ($\chi^2(1) = 3.408$, $p = .065$); + although the equal variance model fit, the unequal variance model was marginal, $\chi^2(1) = 3.043$, $p = .081$. [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

exception of Participant 4, for whom the unequal variance model fit. As can be seen in Tables 3, 4, and 5, sensitivities across observers ranged from .78 to .82 (M = .80). The corresponding values of d' (d_a for Participant 4) ranged from 1.11 to 1.30 (M = 1.19). In the Δ = 80 condition the data of two observers met the Gaussian equal variance assumption (Observers 1 and 5; see Tables 3 and 5 and Figures 5 and 6). The data of two other observers (3 and 6) met the unequal variance assumption, and in each case the slope of the ROC (z-score form) was less than one, indicating that $\sigma_s > \sigma_n$. This meant that for only two observers (2 and 4) did neither model fit the data. Figures 5 and 6 illustrate why this occurred. For both of these observers, each data point was associated with a different level of sensitivity in the Δ = 80 condition. Essentially, these points fall on different ROCs. We might suspect that the reason for this might be that these two observers had difficulty with the task itself. Although this may be true (in each session the Δ = 80 condition was completed first for these observers), it is unlikely because across sessions these participants were able to perform the Δ = 20 condition such that the data conformed to the Gaussian equal variance (4) or unequal variance (2) assumptions.

Comparison of difficulty conditions for FSDT analysis. Comparisons of the two difficulty conditions for FSDT were accomplished by comparing the equal variance models for Participants 1 and 5 and the unequal variance models for Participants 3 and 6. Statistically significant differences were obtained for observer 1 (z = 5.27) and observer 5 (z = 4.22), with the higher sensitivity associated with the Δ = 80 condition in each case. Comparable tests for observers 3 (z = .28) and 6 (z = −.68) revealed no statistically significant differences. Thus, the difficulty manipulation manifested only in the cases where both conditions met the Gaussian equal variance assumption. Note that in this case the slopes of the two functions were equivalent (b = 1), indicating that the independence of sensitivity and response bias asserted by traditional SDT may extend to FSDT.

Comparison of traditional SDT and FSDT. Where comparisons were possible (i.e., for all participants in the Δ = 20 condition and Participants 1, 5, and 6 in the Δ = 80 condition), FSDT analyses revealed higher sensitivity scores than the traditional SDT analysis. This interpretation is supported by z tests, summarized in Table 6. As can be seen, in every instance the FSDT analysis yielded statistically significantly higher sensitivity scores than the traditional analysis. These results indicate that FSDT consistently captured observer sensitivity to different categories of duration that was not reflected in the traditional SDT analysis, most likely because in the latter the stimulus and response sets were collapsed into two categories. Further, these findings confirm the arguments by previous researchers (Hancock et al., 2000; Parasuraman et al., 2000) that FSDT is a useful extension of the traditional model for instances in which multiple stimulus and response categories obtain.

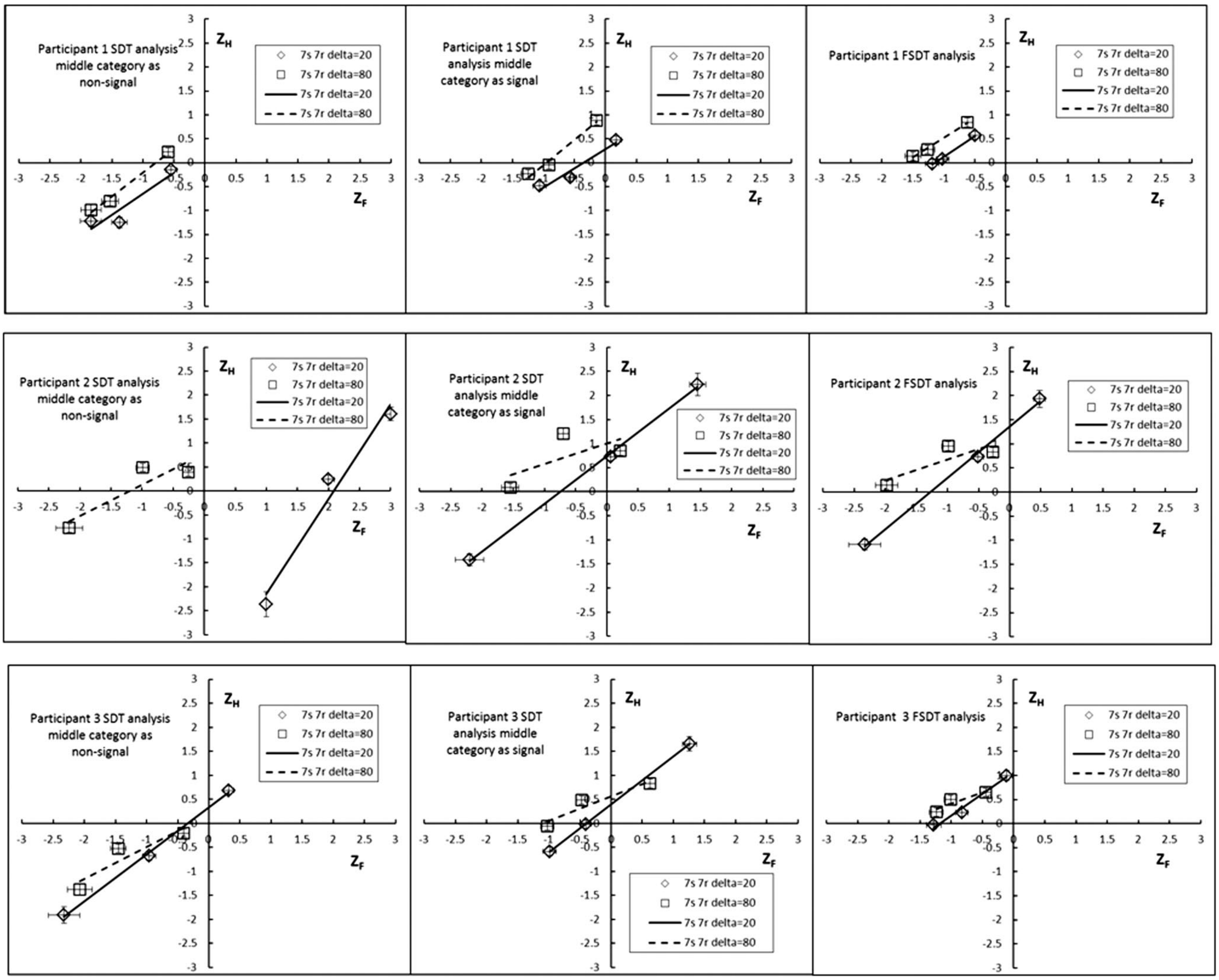

Figure 5. Crisp and Fuzzy ROC functions for Participants 1−3 of Experiment 1. *Note:* Error bars are 95% confidence intervals.

Discussion

The sensitivity associated with the FSDT analysis was higher than that associated with the traditional SDT analyses. This is consistent with previous research (Masalonis & Parasuraman, 2003; Murphy, Szalma, & Hancock, 2003, 2004), and may be due to the loss of information that necessarily occurs when multiple stimulus levels and responses are constrained to collapse into two categories. That is, the higher sensitivity values of FSDT may be due to the fact that the latter theory allows a degree of membership in the correct detection category rather than full membership in the miss category, and similarly, allowing only a degree of false alarm rather than a "full false alarm." Partial hit rates in FSDT are thus categorized as either full hits *or* full misses in the crisp SDT analysis, and partial false alarms are categorized in the traditional SDT analyses as full false alarms or correct rejections. Thus, the loss of information resulting from forcing events into mutually exclusive categories, as required by traditional SDT, serves to produce sensitivity scores that may not always reflect the "true" level of sensitivity.

A central question for FSDT is whether the assumptions of traditional SDT hold for this new model. The current results indicate that at both difficulty levels and for all participants (with the exception only of Participants 2 and 4 in the $\Delta = 80$ ms condition), the assumption of normally distributed noise and signal plus noise distribution is tenable (see χ^2 tests in Tables 3 and 4). In addition, the equal variance assumption holds for the more difficult discrimination for five of the six participants in the $\Delta = 20$ ms condition, although the data of only two participants met this assumption in the $\Delta = 80$ ms condition. Although the equal variance assumption is not crucial to application of SDT, it can bias standard deviation measures of sensitivity (d') and response bias (c), and whether FSDT conforms to this assumption may have implications for selection of parametric measures to evaluate performance using the model.

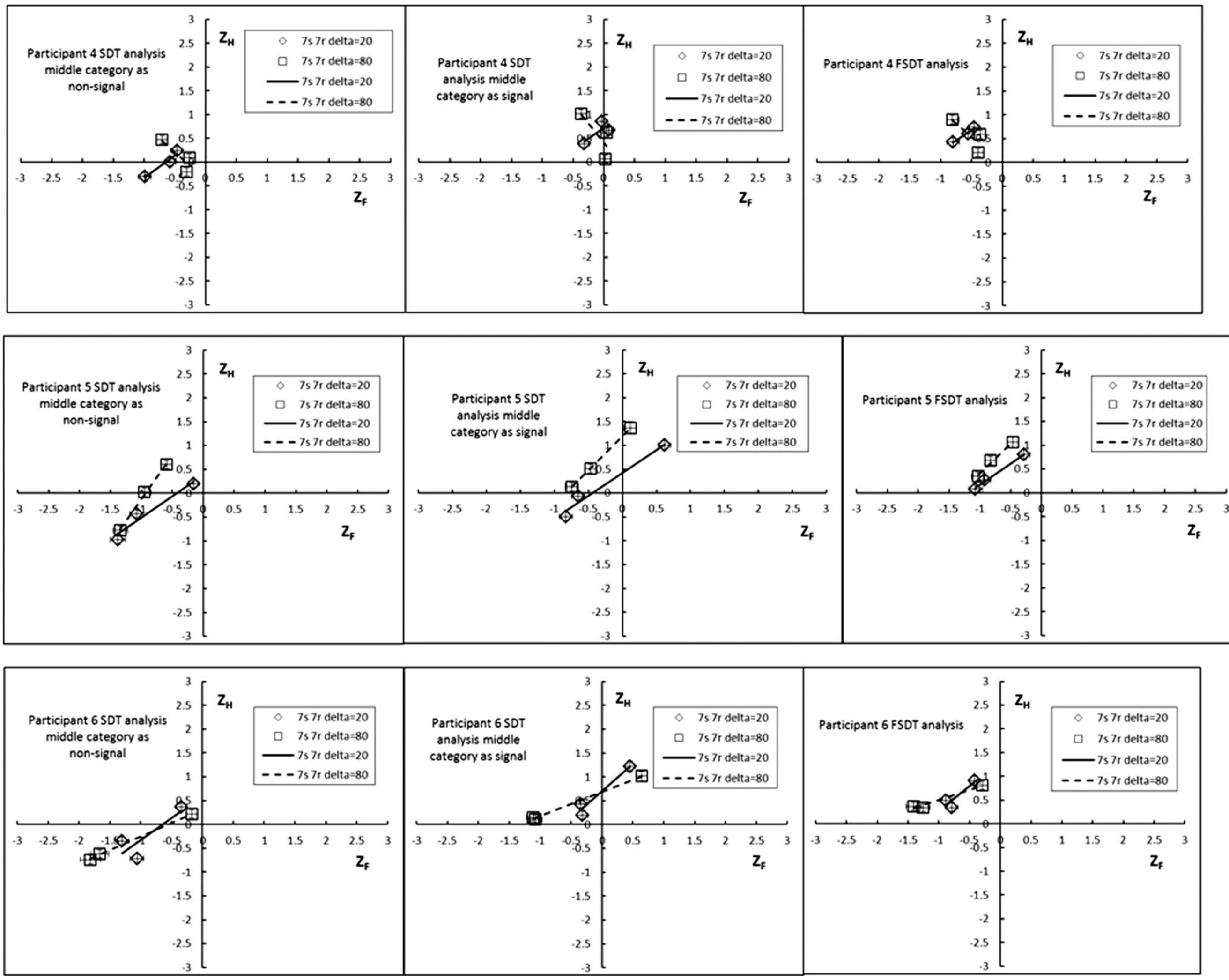

Figure 6. Crisp and Fuzzy ROC functions for Participants 4−6 of Experiment 1. *Note:* Error bars are 95% confidence intervals.

However, it is not clear whether the outcome differences for the two difficulty conditions are due to differences between the two absolute Δ values (i.e., 20 ms vs. 80 ms) or to the range of durations used (200–320 ms vs. 200–680 ms, respectively), because the two are necessarily confounded in the present protocol. Experiment 2, therefore, was designed to disambiguate this confound by examining the respective effects of the size of stimulus and response sets on signal detection using FSDT and by exploring the impact of stimulus range and perceptual "distance" between adjacent stimulus categories. In Experiment 2, the effect of discrimination difficulty was, therefore, replicated, and the range of stimulus categories was manipulated as well as the range of response categories (i.e., binary vs. multicategorical). A goal for this study was to investigate the effects of systematic variation in stimulus and response set size on fuzzy ROC functions.

Experiment 2

Data from Experiment 1 suggested that the fundamental assumptions of normally distributed noise and signal + noise distributions of equal variance are reasonably met by FSDT analysis, although there was greater consistency associated with the $\Delta = 20$ condition relative to the $\Delta = 80$ condition. These results suggest that the decision space shown in Figure 3 is an adequate representation of the data derived from fuzzy as well as from traditional SDT hit and false alarm rates. However, the less stable estimates of the traditional analyses suggest that bifurcation of the type employed here may introduce either artifact and/or error into SDT analyses. Given the instability in the $\Delta = 80$ ms condition, an additional replication of the SDT versus FSDT differences was considered advisable. In addition, in Experiment 1 the effectiveness of the discrimination difficulty manipulation was observed only in the two cases in which the equal variance model fit the data for both conditions. Inspection of the figures indicates that this is not an artifact. For the other four observers, the ROC points of the two conditions are quite close to one another.

Experiment 1 was, therefore, replicated and extended to manipulate response set size and to evaluate the effects of

Table 6
Z-Tests for Comparison of FSDT and Traditional SDT for Each Observer in Each Condition in Experiment 1

Observer	Model compared	z
	$\Delta = 20$	
FSDT vs. SDT, nonsignal		
1	—	—
2	Unequal variance	4.78
3	Equal variance	9.06
4	Unequal variance	4.40
5	Equal variance	8.71
6	—	—
FSDT vs. SDT, signal		
1	Equal variance	9.47
2	Unequal variance	5.16
3	Equal variance	8.63
4	Unequal variance	3.93
5	Equal variance	9.03
6	Unequal variance	7.35
	$\Delta = 80$	
FSDT vs. SDT, nonsignal		
1	Equal variance	8.52
2	—	—
3	—	—
4	—	—
5	Unequal variance	3.25
6	Unequal variance	6.13
FSDT vs. SDT, signal		
1	Equal variance	6.89
2	—	—
3	—	—
4	—	—
5	Unequal variance	5.64
6	Unequal variance	3.65

Note. The criterion for statistical significance was $z_{\alpha = .05} = 1.96$.

interval differences and range of stimulus (**s**) and response (**r**) values. The two conditions of Experiment 1, seven stimulus levels and seven response levels with two different interval sizes (7s, 7r, $\Delta = 20$ and 7s, 7r, $\Delta = 80$), were replicated and two additional conditions were added: one in which the interval was held constant at $\Delta = 20$ but the range of stimulus categories was extended to 24 levels (24s, 7r), and one in which the response set was binary (7s, 2r, $\Delta = 20$). This permitted comparison of response set sizes, as well as the effects of interval differences with the number of stimulus categories held constant (i.e., 7s, 7r, $\Delta = 20$ vs. 7s, 7r, $\Delta = 80$) and the influence of the range of stimulus values with interval held constant (i.e., 7s, 7r, $\Delta = 20$ and 24s, 7r, $\Delta = 20$).

Method

Participants. Four participants, two males and two females, volunteered to participate in the experiment. They ranged in age from 23 to 32, and were unfamiliar with the purpose of the experiment or the stimuli used. These were different individuals from those who participated in Experiment 1. Each of the procedural and approval strictures that were applied in Experiment 1 also pertained here.

Stimuli and procedure. The stimuli and procedures were the same as those used in Experiment 1, except that with the addition of two task conditions participants completed the experiment over 12 days rather than 6 days as in Experiment 1.

Results

Traditional SDT analysis.

Replication of experiment 1. As in the first experiment, the pattern of results for the traditional SDT analyses depended on both the observer and the way in which the data were bifurcated. In the $\Delta = 20$ condition, when the middle stimulus category was considered a "nonsignal" the equal variance model fit for Observer 3 (see Table 9 and Figure 7), and the unequal variance model fit for Observer 1 (see Table 7 and Figure 7). In the case of the latter, the slope of the function was greater than one, indicating that $\sigma_s < \sigma_n$. Neither model fit the data for Participant 4, and a convergence failure was obtained for Participant 2. A similar pattern was observed when the middle category was classified as a signal, but here the equal variance model did fit the data for both Participants 1 and 3.

For the $\Delta = 80$ condition, when the middle stimulus category was considered a "nonsignal" the equal variance model fit for two observers (1 and 2; see Tables 7 and 8, respectively, and Figure 7), and the unequal variance model fit for one observer (4; Table 10 and Figure 7). In the case of the latter, the slope of the function was greater than one, indicating that $\sigma_s < \sigma_n$. A convergence failure was obtained for Participant 3. However, also as in Experiment 1, when the middle category was classified as a signal the model fit was poor. The unequal variance model fit for Participant 1 where the slope of the function was less than one, indicating that $\sigma_s > \sigma_n$. However, neither model fit for the other three participants.

The two difficulty conditions could be compared only for Observer 1 because of failures of either model to fit in both conditions for the other participants. Sensitivity scores for the $\Delta = 80$ condition were statistically significantly higher than that for the $\Delta = 20$ condition for both the middle category classified as nonsignal analysis as well as for when that category was classified as a signal. Note that for both of these analyses the unequal variance models were compared. It is clear that introducing crisp categorization leads to general failures in model fit, indicating that such crisp categorization does not reflect the assumed decision space of SDT and that, therefore, measures such as d' and β would be biased if used in such cases.

Traditional SDT analysis: binary response condition. In this condition the number of stimuli was held at seven but the response set was restricted to a simple yes/no outcome. Unfortunately, data for one of the observers could not be used in analyses of this condition. Observer 3 failed to follow instructions and responded here inappropriately by using multiple categories. Of the remaining three observers, the equal variance model fit for two of them (1 and 2) but neither model fit for Participant 4. This pattern was observed regardless of whether the middle stimulus was classified as a nonsignal or as a signal (Tables 7, 8, and 10, and Figure 7). Hence, when responses are constrained to be binary, traditional SDT reasonably reflects the decision space shown in Figure 3. For Participant 1 sensitivity was greater for the binary response condition relative to the 7s, 7r, $\Delta = 20$ condition when the middle stimulus was classified either as a nonsignal ($z = 3.35$) or as a signal ($z = 2.53$). Thus, in a crisp analysis, traditional SDT yields higher sensitivity scores when a binary response set is used

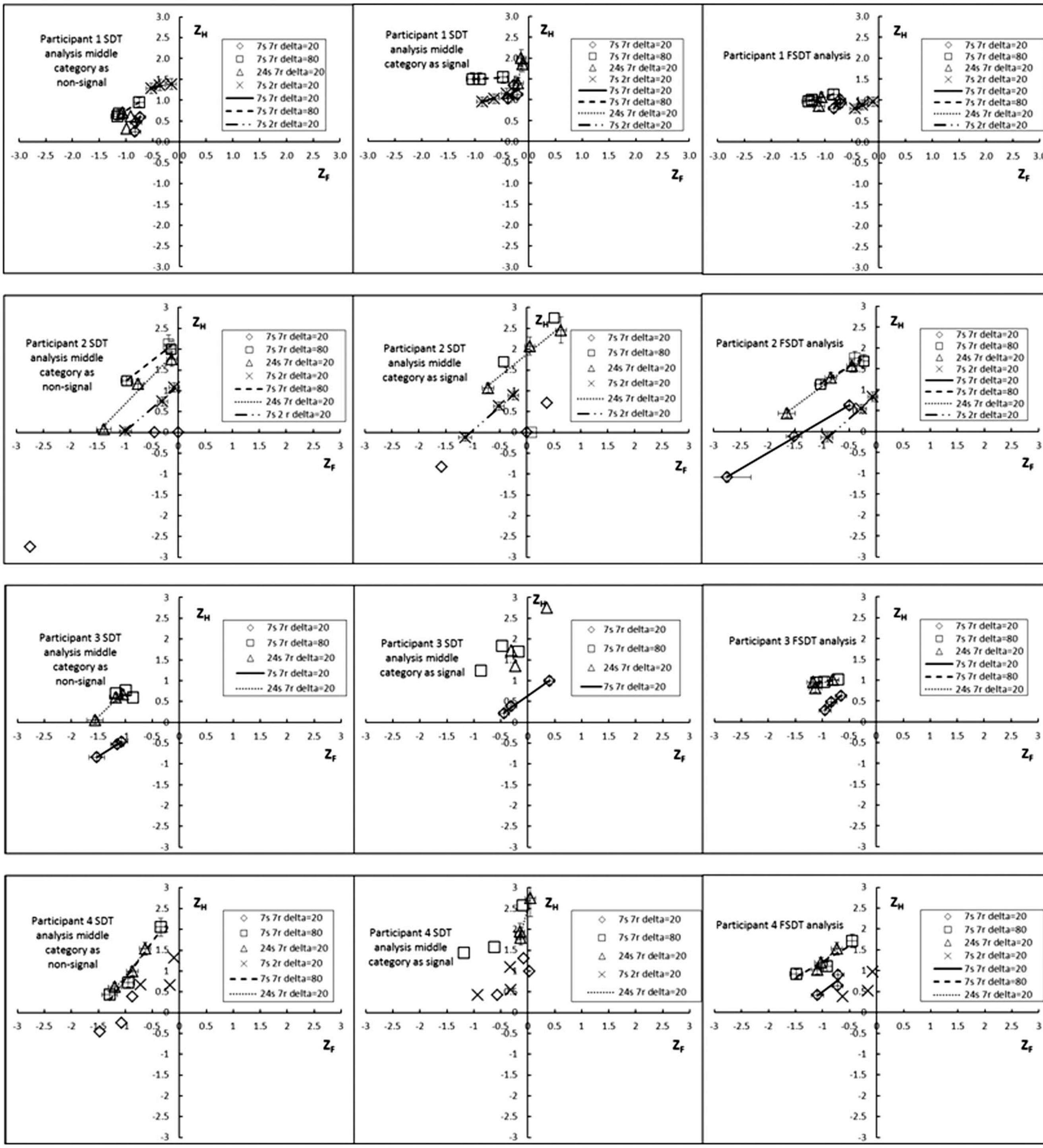

Figure 7. Crisp and Fuzzy ROC functions for Participants 1−4 of Experiment 2. *Note:* Error bars are 95% confidence intervals.

relative to when participants can respond using multiple categories. This comparison could not be done for Participant 2 because of convergence failures in each analysis.

Traditional SDT analysis: 24s, 7r, Δ = 20 condition. To evaluate this condition using traditional SDT, the lower 12 time intervals were classified as "nonsignals" and the upper 12 intervals as "signals." As in other analyses, the middle response category was considered either as a "no" or a "yes" response in separate analysis. When seven response categories were permitted but 24 stimuli presented, model fit varied across participants and de-

Table 7
Experiment 2: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participant 1 (Male; Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
Traditional SDT Analysis: Middle stimulus coded as nonsignal							
7s, 7r, Δ = 20 ms condition							
0.205[u] (p = .651)	.809 (.020)	1.239 (.104)	3.599 (3.624)	3.985 (4.554)	1.593 (1.148)	1.488 (1.170)	1.709 (1.103)
7s, 7r, Δ = 80 ms condition							
0.484[e] (p = .785)	.893 (.008)	1.757 (.064)	1.757 (.064)	1.000	.482 (.100)	−.174 (.097)	.397 (.099)
24s, 7r, Δ = 20 ms condition							
CF							
7s, 2r, Δ = 20 ms condition							
2.795[e] (p = .247)	.884 (.010)	1.693 (.070)	1.693 (.070)	1.000	−.831 (.109)	−1.111 (.120)	−.605 (.102)
Traditional SDT Analysis: Middle stimulus coded as signal							
7s, 7r, Δ = 20 ms condition							
3.196[e] (p = .202)	.847 (.011)	1.448 (.063)	1.448 (.063)	1.000	−.671 (.086)	−.755 (.089)	−.464 (.080)
7s, 7r, Δ = 80 ms condition							
.006[u] (p = .940)	.941 (.020)	2.211 (.242)	1.568 (.197)	.074 (.230)	−3.164 (3.195)	−3.666 (2.999)	−3.312 (3.151)
24s, 7r, Δ = 20 ms condition							
.286[u] (p = .593)	.645 (.095)	.526 (.359)	2.898 (1.370)	7.735 (8.949)	.086 (1.150)	.322 (1.168)	1.091 (1.80)
7s, 2r, Δ = 20 ms condition							
2.752[e] (p = .253)	.883 (.009)	1.683 (.064)	1.683 (.064)	1.000	−.322 (.093)	−.628 (.098)	−.087 (.091)
FSDT Analysis							
7s, 7r, Δ = 20 ms condition							
.508[e] (p = .776)	.880 (.009)	1.658 (.062)	1.658 (.062)	1.000	−.175 (.090)	−.236 (.090)	.025 (.089)
7s, 7r, Δ = 80 ms condition							
.028[u] (p = .866)	.911 (.028)	1.905 (.243)	1.423 (.287)	.341 (.248)	−.708 (.756)	−1.367 (.675)	−.799 (.752)
24s, 7r, Δ = 20 ms condition							
1.318[e] (p = .252)	.929 (.006)	2.074 (.067)	2.074 (.067)	1.000	−.001 (.120)	−.013 (.067)	.277 (.121)
7s, 2r, Δ = 20 ms condition							
1.581[e] (p = .454)	.794 (.012)	1.162 (.060)	1.162 (.060)	1.000	−.334 (.063)	−.509 (.069)	−.199 (.060)

Note. C = Conservative; N = unbiased; L = Lenient; [e] data fits the equal variance model; [u] data fits the unequal variance model; [n] neither model fits; values presented correspond to the unequal variance model; CF neither model fit, and a convergence failure was obtained for the unequal variance case. [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

pended on how the middle response category was bifurcated. When the middle category was classified as a "no" or "nonsignal" response, the equal variance model fit for Observer 3 and the unequal variance model fit for Participants 2 and 4 (see Tables 8, 9, and 10, and Figure 7). In each of the latter two cases the slope of the function was greater than one, indicating that $\sigma_s < \sigma_n$. A convergence failure was observed for Participant 1.

When the middle category was classified as a "yes" or "signal" response, the equal variance model fit for Observer 2 and the unequal variance model fit for Participants 1 and 4 (see Tables 7, 8, and 10, and Figure 7). In each of the latter two cases the slope of the function was greater than one. Neither model fit for Participant 3. For those participants for whom a comparison with the 7s, 7r, Δ = 20 condition could be made (i.e., Participant 3 when the middle category was classified "no" and Participant 1 when the middle category was classified as a "yes"), the relative sensitivity scores differed in opposite directions. Thus, for Participant 3 a higher sensitivity was obtained in the 24s, 7r, Δ = 20 relative to the 7s, 7r, Δ = 20 condition (z = 11.18), and for Participant 1 a higher sensitivity were obtained in the 7s, 7r, Δ = 20 relative to the 24s, 7r, Δ = 20 condition (z = 2.11). Hence, in traditional SDT analysis bifurcation of a larger stimulus range exhibits inconsistencies in model fit similar to those associated with a narrower stimulus range. That is, as in the 7s, 7r, Δ = 20 condition, in the 24s, 7r, Δ = 20 condition the fit of the data to the model corresponding to the assumed decision space is highly variable when the data are bifurcated. Although the Gaussian assumption generally held, confirming the general form of the decision space, failure to meet the equal variance assumption indicates that standard deviation based measures of sensitivity and bias (d' and c, respectively) would not be advisable (Swets, 1996).

FSDT analysis.

Replication of experiment 1. The results for the two conditions replicating Experiment 1 confirmed that for the most part the Gaussian model did fit the data for both the 7s,7r, Δ = 20 and 7s, 7r, Δ = 80 conditions. In Experiment 1, data of three of the six participants met the unequal variance assumption and for one participant the data fit the equal variance model for the 7s, 7r, Δ = 80 condition. The same condition in Experiment 2 resulted in two of four participants whose data fit the equal variance model and two whose data fit the unequal variance model. As in Experiment 1, in the 7s, 7r, Δ = 20 condition the data in most cases of Experiment 2 met the equal variance assumption. In all cases the data were consistent with the Gaussian assumption of SDT. Thus, the replication indicates that FSDT analysis of multicategorical data adequately reflects the decision-space representation of Figure 3 and does so more consistently than the traditional SDT analyses of the same data. Note that in both experiments, two

Table 8
Experiment 2: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participant 2 (Female) (Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
			Traditional SDT Analysis: Middle stimulus coded as nonsignal				
			7s, 7r, Δ = 20 ms condition				
CF							
			7s, 7r, Δ = 80 ms condition				
0.544[e] (p = .762)	.941 (.007)	2.205 (.088)	2.205 (.088)	1.000	−.322 (.136)	−2.056 (.201)	−2.109 (.204)
			24s, 7r, Δ = 20 ms condition				
1.883[u] (p = .170)	.886 (.010)	1.702 (.074)	2.044 (.158)	1.373 (.171)	1.316 (.104)	−1.370 (.181)	−.039 (.160)
			7s, 2r, Δ = 20 ms condition				
.992[e] (p = .609)	.774 (.013)	1.065 (.061)	1.065 (.061)	1.000	.507 (.063)	−.242 (.057)	−.519 (.068)
			Traditional SDT Analysis: Middle stimulus coded as signal				
			7s, 7r, Δ = 20 ms condition				
CF							
			7s, 7r, Δ = 80 ms condition				
4.882[n] (p = .027)	.897 (.039)	1.786 (.303)	2.466 (.193)	1.677 (.483)	−.856 (.387)	−3.047 (.452)	−4.710 (1.028)
			24s, 7r, Δ = 20 ms condition				
1.149[e] (p = .563)	.905 (.010)	1.852 (.088)	1.852 (.088)	1.000	−.341 (.109)	−2.889 (.240)	−1.856 (.180)
			7s, 2r, Δ = 20 ms condition				
.751[e] (p = .687)	.782 (.013)	1.104 (.061)	1.104 (.061)	1.000	.713 (.078)	−.065 (.055)	−.355 (.060)
			FSDT Analysis				
			7s, 7r, Δ = 20 ms condition				
.032[u] (p = .859)	.788 (.020)	1.131 (.097)	.997 (.122)	.744 (.099)	3.039 (.573)	.822 (.146)	−.371 (.136)
			7s, 7r, Δ = 80 ms condition				
2.228[e] (p = .328)	.929 (.007)	2.078 (.075)	2.078 (.075)	1.000	−.114 (.121)	−1.378 (.152)	−1.611 (.161)
			24s, 7r, Δ = 20 ms condition				
.375[e] (p = .829)	.929 (.007)	2.073 (.073)	2.073 (.073)	1.000	1.243 (.145)	−1.196 (.143)	−.463 (.124)
			7s, 2r, Δ = 20 ms condition				
.802[e] (p = .670)	.719 (.014)	.820 (.058)	.820 (.058)	1.000	.433 (.054)	−.112 (.041)	−.314 (.048)

Note. C = Conservative; N = unbiased; L = Lenient; [e] data fits the equal variance model; [u] data fits the unequal variance model; [n] neither model fits; values presented correspond to the unequal variance model; CF neither model fit, and a convergence failure was obtained for the unequal variance case. [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

individuals showed significant differences in performance, with the Δ = 80 condition yielding superior performance relative to the Δ = 20 condition. Thus, increasing the distance between categories but holding the number of categories constant has either a facilitative or null effect on sensitivity across observers.

FSDT Analysis: Binary response condition. In this condition, the number of stimuli was held at seven but the response set was restricted to only yes/no. Here, the equal variance model fit for two observers (1 and 2), but neither model fit for Observer 4. Examining the pattern within each of the three participants between the binary and s7, r7, Δ = 20 conditions, in one case (Participant 1) both conditions fit the equal variance model. For Participant 2 the data for the binary condition fit the equal variance model but the s7, r7, Δ = 20 condition conformed to the unequal variance model (see Table 8 and Figure 7). For Observer 4 the s7, r7, Δ = 20 condition fit an equal variance model, but, as noted above, neither model fit the data for the binary condition. Hence, when this participant was forced to make binary decisions to fuzzy stimuli, the SDT model illustrated in Figure 3 failed to provide an adequate fit to the data.

Note that the binary response set resulted in lower sensitivity scores relative to the s7, r7, Δ = 20 condition. The differences in performance between the two conditions for the two participants for whom comparisons could be made were evaluated via z tests, and in each case the 7s, 7r, Δ = 20 condition yielded higher A_z scores relative to the binary condition (Participant 1, z = 5.73; Participant 2, z = 2.83). This provides support for the notion that constraining response sets impairs task performance when fuzzy stimulus sets are used.

FSDT analysis: 24s, 7r, Δ = 20 condition. The equal variance model fit for all observers when seven response categories are permitted but there are 24 stimulus categories. Relative to the 7s, 7r, Δ = 20 condition, higher sensitivities were obtained in the 24s, 7r, Δ = 20 condition for each participant. Thus, extending the stimulus range consistently facilitated fuzzy sensitivity, and relative to the crisp analysis, the FSDT analysis for this condition was more consistent with the assumptions of SDT regarding the structure of the decision space.

Comparison of traditional SDT and FSDT. Formal comparisons were computed to determine whether the differences in sensitivity observed for the two methods were statistically reliable. *Z* tests for comparisons of SDT and FSDT analyses are shown in Table 11. Note that comparisons were made only in cases in which an adequate model fit was obtained for either the equal or unequal variance model. Where significant differences were observed FSDT analysis again yielded higher sensitivity scores than traditional SDT analysis when a large number of categories were presented (i.e., 24) and multiple responses were permitted. Even in cases in which there were nonsignificant differences (e.g., Participants 1 and 2 for the 7s, 7r, Δ = 80 condition; see Table 11) the

Table 9
Experiment 2: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participant 3 (Male) (Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
Traditional SDT Analysis: Middle stimulus coded as nonsignal							
7s, 7r, Δ = 20 ms condition							
0.323[e] (p = .851)	.672 (.017)	.632 (.066)	.632 (.066)	1.000	.485 (.062)	.747 (.091)	.531 (.067)
7s, 7r, Δ = 80 ms condition							
CF							
24s, 7r, Δ = 20 ms condition							
.941[e] (p = .625)	.887 (.009)	1.710 (.067)	1.710 (.067)	1.000	.485 (.097)	1.319 (.125)	.317 (.095)
7s, 2r, Δ = 20 ms condition							
—							
Traditional SDT Analysis: Middle stimulus coded as signal							
7s, 7r, Δ = 20 ms condition							
.494[e] (p = .781)	.676 (.015)	.645 (.058)	645 (.058)	1.000	.074 (.033)	−.028 (.031)	−.450 (.045)
7s, 7r, Δ = 80 ms condition							
4.996[n] (p = .025)	.939 (.012)	2.185 (.135)	1.991 (.150)	.812 (.240)	−.616 (.347)	−1.494 (.292)	−1.841 (.232)
24s, 7r, Δ = 20 ms condition							
4.391[n] (p = .036)	.809 (.047)	1.234 (.246)	2.083 (.187)	2.168 (.682)	−.471 (.362)	−.210 (.370)	−3.205 (.801)
7s, 2r, Δ = 20 ms condition							
—							
FSDT Analysis							
7s, 7r, Δ = 20 ms condition							
.321[e] (p = .852)	.814 (.011)	1.261 (.060)	1.261 (.060)	1.000	.433 (.071)	.227 (.066)	.024 (.064)
7s, 7r, Δ = 80 ms condition							
.008[u] (p = .928)	.867 (.046)	1.576 (.302)	1.130 (.268)	.170 (.283)	−1.608 (1.710)	−1.768 (1.683)	−2.036 (1.616)
24s, 7r, Δ = 20 ms condition							
3.530[e] (p = .171)	.916 (.007)	1.95 (.066)	1.95 (.066)	1.000	.192 (.111)	.319 (.113)	−.202 (.111)
7s, 2r, Δ = 20 ms condition							
—							

Note. C = Conservative; N = unbiased; L = Lenient; [e] data fits the equal variance model; [u] data fits the unequal variance model; [n] neither model fits; values presented correspond to the unequal variance model; CF neither model fit, and a convergence failure was obtained for the unequal variance case. [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

A_z score for FSDT was higher than that for traditional SDT. In contrast, for the two cases that could be compared (Participants 1 and 2) traditional SDT yielded higher A_z scores when a binary response was required relative to the FSDT analysis. Thus, the sensitivity scores one observes with respect to the two analyses varies according to the size of the stimulus and response set. The size of the temporal interval seems to exert less of an effect, except that the more difficult discrimination yielded more consistent model fits across participants. Given that the same data were used for the SDT and FSDT analyses, these results indicate that how the data are structured, based on common assumptions regarding the underlying decision space, can affect the level of sensitivity estimated by the respective procedures.

Discussion

For the binary response condition the equal variance model fitted appropriately for Participants 1 and 2 whether the analysis was fuzzy or crisp, and neither model fit the data of Participant 4 across all analyses. For the other conditions in which multicategorical responses were permitted, the FSDT analysis fitted better than the crisp analyses. Further, the FSDT analyses yielded higher sensitivities for these conditions relative to the crisp analyses (this general pattern was also observed in the study reported by Murphy et al., 2004).

For the two conditions that were replicated, in each experiment the slope was always less than one when the unequal variance model fit. This pattern is consistent with that reported by Murphy et al. (2004). Recall that when b < 1 the implication is that the variance of the S + N distribution is larger than that of the N distribution. This is based on the following representation of an ROC function,

$$z_H = (\sigma_n/\sigma_s)z_F + (\mu_s/\sigma_s).$$

For the unequal variance model, then, the slope is $b = \sigma_n/\sigma_s$ so that when b < 1, then $\sigma_s > \sigma_n$, and when b > 1, then $\sigma_s < \sigma_n$. The relatively consistent b < 1 suggests that in cases where the unequal variance model fit, the fuzzy signals shifted the position of the distribution and also increased its variability. Although it is possible that some participants had difficulty in setting response criteria, in every case in which the unequal variance model fit in one condition, the equal variance model fit for that same individual in the other condition (this only with the exception of Participant 4 in Experiment 2, whose data did not converge in the 7s, 7r, Δ = 80 condition). Thus, whether the equal variance or unequal variance model fitted depended on both observer and condition. In no case did an observer produce data that fit the unequal variance model in both conditions in the FSDT analysis. This is important because it suggests that the form of the distributions in the decision

Table 10

Experiment 2: Goodness of Fit, Sensitivity, and Criterion (Response Bias) Statistics Calculated From the Hits, False Alarms, Misses, and Correct Rejections for Participant 4 (Female) (Standard Errors in Parentheses)

χ^2	A(z)	d'[a]	a	b	C ln(β)	N ln(β)	L ln(β)
Traditional SDT Analysis: Middle stimulus coded as nonsignal							
7s, 7r, $\Delta = 20$ ms condition							
4.412[n] ($p = .036$)	.816 (.022)	1.273 (.115)	1.689 (.422)	1.588 (.376)	1.287 (.170)	1.142 (.227)	.746 (.269)
7s, 7r, $\Delta = 80$ ms condition							
2.402[u] ($p = .121$)	.896 (.010)	1.783 (.077)	2.612 (.279)	1.814 (.284)	1.263 (.144)	.822 (.178)	−1.366 (.247)
24s, 7r, $\Delta = 20$ ms condition							
.727[u] ($p = .394$)	.906 (.011)	1.858 (.092)	2.533 (.371)	1.649 (.399)	.385 (.267)	1.007 (.221)	−.455 (.238)
7s, 2r, $\Delta = 20$ ms condition							
19.494[n] ($p < .001$)	.804 (.016)	1.213 (.084)	1.157 (.100)	.905 (.233)	−.051 (.269)	−.444 (.260)	−.742 (.210)
Traditional SDT Analysis: Middle stimulus coded as signal							
7s, 7r, $\Delta = 20$ ms condition							
6.211[n] ($p = .013$)	.752 (.023)	.962 (.105)	1.186 (.098)	1.430 (.303)	.421 (.199)	−.222 (.211)	−.373 (.199)
7s, 7r, $\Delta = 80$ ms condition							
6.285[n] ($p = .012$)	.961 (.007)	2.496 (.124)	2.368 (.186)	.895 (.214)	−.411 (.297)	−1.375 (.282)	−2.777 (.234)
24s, 7r, $\Delta = 20$ ms condition							
.188[u] ($p = .664$)	.701 (.083)	.747 (.339)	2.506 (.332)	4.637 (2.669)	−.302 (.660)	−.109 (.659)	−2.202 (.681)
7s, 2r, $\Delta = 20$ ms condition							
18.222[n] ($p < .001$)	.799 (.013)	1.187 (.065)	1.232 (.149)	1.074 (.263)	.335 (.225)	−.036 (.252)	−.393 (.228)
FSDT Analysis							
7s, 7r, $\Delta = 20$ ms condition							
2.898[e] ($p = .235$)	.854 (.010)	1.491 (.061)	1.491 (.061)	1.000	.505 (.085)	.071 (.078)	−.141 (.079)
7s, 7r, $\Delta = 80$ ms condition							
4.754[e] ($p = .093$)	.940 (.006)	2.198 (.073)	2.198 (.073)	1.000	.562 (.136)	−.213 (.131)	−1.407 (.158)
24s, 7r, $\Delta = 20$ ms condition							
.689[e] ($p = .708$)	.940 (.006)	2.200 (.070)	2.200 (.070)	1.000	−.182 (.131)	.076 (.131)	−.853 (.141)
7s, 2r, $\Delta = 20$ ms condition							
8.210[n] ($p = .004$)	.737 (.016)	.895 (.071)	.907 (.096)	1.025 (.259)	.134 (.246)	−.135 (.256)	−.373 (.224)

Note. C = Conservative; N = unbiased; L = Lenient; [e] data fits the equal variance model; [u] data fits the unequal variance model; [n] neither model fits; values presented correspond to the unequal variance model; CF neither model fit, and a convergence failure was obtained for the unequal variance case. [a] For cases where the equal variance model fit d' is reported. For all other cases, d_a is reported.

space representation is of the same general form for FSDT as it is for traditional SDT. However, the results also suggest that the structure of the implied decision space may vary somewhat (in terms of the variability of each distribution) depending on which analysis is computed, and, therefore, the appropriateness of d' and c may also depend on analysis method.

The results of Experiment 2 indicate that the effects of crisp versus fuzzy treatment of multicategorical data have similar implications for model fit regardless of the magnitude of stimulus range or the magnitude of differences between categories. However, the range of response set does affect the relative model fits. Although fuzzy SDT data were consistent with the SDT model fit for both binary and multicategorical response, in the latter FSDT fit the decision model better than crisp SDT.

General Discussion

For the FSDT analyses, the results generally confirmed those of Murphy et al. (2004), in that the equal variance model fit for the $\Delta = 20$ ms condition. With respect to the difficulty discrimination, results were mixed, but again this is consistent with the findings of Murphy et al. (2004). In some cases the difficult condition was associated with lower sensitivity, in other cases scores were similar, and in a few cases the difficult condition was associated with higher sensitivity. However, the results of the current experiments were more consistent across observers than those of Murphy et al. (2004). In both experiments, FSDT analysis (for the multicategorical response conditions) resulted in higher sensitivities than SDT analysis.

With respect to the comparison of binary versus categorical response, results of this study were generally consistent with those reported by Szalma and O'Connell (2011). That is, for the FSDT analysis higher sensitivity was observed for the multicategorical response set than for the binary response set. Comparison of binary SDT with binary FSDT also generally conformed to the results reported by Szalma and O'Connell (2011), in that for binary responses SDT yielded higher sensitivities than FSDT analysis. This may be due to the loss of information for an FSDT analysis that occurs when a binary decision is required.

The ROC experiments reported here formally confirm that the Gaussian assumptions of traditional SDT are reasonably met in FSDT. In a number of cases the equal variance assumption was also met, but the results for this assumption are, to a degree, somewhat less consistent. It is not clear whether the case-specific failures to meet the equal variance assumption arose from the FSDT model itself or the inability of individual participants to adjust their decision criteria appropriately. Recall that response bias was manipulated using traditional payoff matrices demonstrated to be effective in traditional ROC analysis (e.g., see Mac-

Table 11
Experiment 2: Z-Tests for Comparison of FSDT and Traditional SDT for Each Observer in Each Condition

Observer	Model compared	Z
7s, 7r, Δ = 20 ms condition		
FSDT vs. SDT, nonsignal		
1	Unequal variance	1.49
2	—	—
3	Equal variance	7.01*
4	—	—
FSDT vs. SDT, signal		
1	Equal variance	2.32*
2	—	
3	Equal variance	7.42*
4	—	
7s, 7r, Δ = 80 ms condition		
FSDT vs. SDT, nonsignal		
1	Unequal variance	0.74
2	Equal variance	1.21
3	—	—
4	Unequal variance	3.86*
FSDT vs. SDT, signal		
1	Unequal variance	0.87
2	—	—
3	—	—
4	—	—
24s, 7r, Δ = 20 ms condition		
FSDT vs. SDT, nonsignal		
1	—	—
2	Unequal variance	3.52*
3	Equal variance	2.54*
4	Unequal variance	1.78
FSDT vs. SDT, signal		
1	Unequal variance	2.23*
2	Equal variance	1.97*
3	—	
4	Unequal variance	2.79*
7s, 2r, Δ = 20 ms condition		
FSDT vs. SDT, nonsignal		
1	Equal variance	5.76*
2	Equal variance	2.88*
3	—	—
4	—	—
FSDT vs. SDT, signal		
1	Equal variance	5.93*
2	Equal variance	3.30*
3	—	—
4	—	—

Note. The criterion for statistical significance was $z_{\alpha = .05} = 1.96$.
* $p < .05$. two-tailed test.

millan & Creelman, 2005). However, in traditional ROC studies using such bias manipulations, the signals and nonsignals are unambiguously defined. That is, the observers know with certainty how a false alarm and a miss are defined. By contrast, with fuzzy stimuli, observers were informed that false alarms were "overestimations" of relative duration, and misses were "underestimations (see Appendix)." It may be that it is difficult for observers to set stable fuzzy criteria, possibly because they are being asked to establish a crisp decision threshold based on a fuzzy evidence variable. Indeed, an important issue for future theoretical work and empirical research is to clarify the structure of the underlying decision space in general and the construct of fuzzy criterion setting in particular.

An Alternative Representation of the Decision Space

The decision space underlying traditional SDT shown in Figure 3 lends itself to an intuitive interpretation of d' and β. Indeed, it is this intuitive interpretation that is arguably responsible for the widespread use of these measures despite evidence that other indices are more appropriate in many circumstances (e.g., See, Warm, Howe, & Dember, 1997; Swets, 1996). However, it is a logical possibility that fuzzy d' and β, computed by combining degrees of membership across multiple stimuli, derive from a decision space unlike that shown in Figure 3. The different categories may be represented as separate distributions, as shown in Figure 8 (see Macmillan & Creelman, 2005; Wickens, 2002). The multiple curves with multiple criteria are intuitively attractive as representations of FSDT decision space because each curve can represent different degrees of membership in the category "signal." In this representation the k categories are adjacent to one another and there are k-1 decision criteria for an ordered set of categories in which each stimulus type can be represented as a normally distributed random variable.

However, we believe the structure shown in Figure 8 does not represent a fuzzy decision space. First, use of multiple "fuzzy" criteria would require a redefinition of the nature of the criterion itself. This may prove necessary, but the computational procedures described by Parasuraman et al. (2000) generate a single criterion value, implying that there is a single criterion for a FSDT task. There is evidence that a single fuzzy response bias score is sensitive to manipulation of criterion setting (Stafford, Szalma, Hancock, & Mouloua, 2003). Second, the FSDT procedures require summation across the stimulus membership categories, but in traditional SDT tasks with multiple stimulus levels the categories are preserved by the analytic method (for computational details see Macmillan & Creelman, 2005; Wickens, 2002).

There are also theoretical reasons to reject the decision space shown in Figure 8. In both traditional and fuzzy signal detection theory the *x*-axis (evidence variable; see Figure 3) does not represent category membership (signal/nonsignal) but rather the *strength* or *intensity* of the evidence variable itself. Category membership is represented in the Gaussian distributions, and the fuzziness is not defined by the magnitude of the evidence variable but by the definition of the two categories represented by the two distributions. Thus, the "true" category membership of a given stimulus event is determined by the distribution from which it was

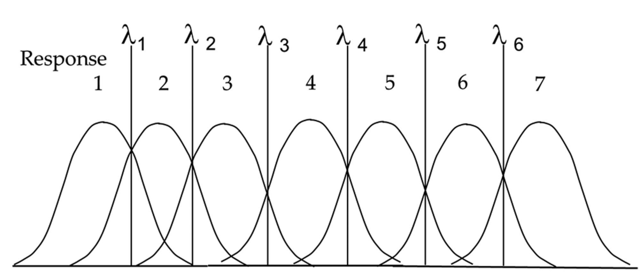

Figure 8. An alternative decision space for fuzzy signal detection theory.

sampled. As Wickens (2002) noted, ". . . each actual event [stimulus presentation] is drawn from one or another of these distributions. On any trial, only *one* of them applies." (p. 12; emphasis added). This observation confirms why the multiple category decision space shown in Figure 8 may not be an adequate representation of fuzzy decision space. If category membership is represented in distributions, then Figure 8 implies that there are discrete changes between fuzzy categories, so that the number of distributions approaches infinity for continuous fuzzy membership sets. A more parsimonious argument is to retain the representation in Figure 3, with the *x*-axis representing the strength/intensity of the evidence variable, and the distributions representing noise (N) and signal embedded in noise (S + N), as in traditional SDT. However, for FSDT the sampling of the distributions would be different than in traditional SDT.

As we observed previously, Wickens (2002) noted that only *one* distribution may be sampled in traditional SDT. In FSDT, however, this constraint is obviated. On any given trial, each possible distribution is sampled, depending on the degree of category membership. For instance, a fuzzy stimulus value of .5 would result from sampling the noise distribution to the degree .5 and sampling the signal + noise distribution to that same degree. A major challenge for the future is to identify ways to empirically sample the two different distributions to different degrees based on fuzzy membership values. However, such a representation would be consistent with the computation of a single response bias value for a given pair of hit and false alarm rates. This represents the next logical step in the elaboration of FSDT as a more encompassing theory. In sum, the multicategorical representation shown in Figure 8 seems to describe an underlying structure of a decision space very different from that of the FSDT model. Indeed, it should be noted that in the multicategorical (traditional) SDT task the categories are crisp, that is, mutually exclusive. Hence, the similarity of multiple categories to multiple degrees of category membership breaks down when the formal characteristics of the models are considered.

Comparison of Traditional SDT and FSDT

In general, the FSDT analyses of the multicategorical stimulus and response dimensions resulted in more instances of appropriate model fit than the crisp analyses of the same data (but collapsing **s** and **r** into two respective categories in the latter case). In addition, the relative changes in the functions were more sensitive to manipulations of the **s** and **r** dimensions in FSDT relative to the traditional SDT analyses. By permitting intermediate values of "signalness," FSDT facilitates a more accurate representation of the underlying state-of-the-world. This interpretation is underscored by the finding in some cases that the traditional SDT model fit of data, and, therefore, the underlying properties of the decision space implied by the model depended on how the middle category was classified. If one assumes a stable decision space underlying detection data, differences between the SDT analyses as a function of how the middle category is classified must represent an artifact of the bifurcating procedure itself. FSDT is not as vulnerable to this problem (although it is dependent on the validity of the function mapping the physical variable to fuzzy set membership). FSDT may, therefore, be a particularly useful extension of traditional SDT when multicategorical or continuous stimulus dimensions are examined, or when the intrinsic characteristics of stimuli do not permit precise assignment to one of two restricted categories.

To be sure, FSDT is not a panacea, nor indeed does it necessarily replace traditional (crisp) SDT in all circumstances. For instance, the sensitivity estimated for the binary response condition was substantially lower in the FSDT as compared to the SDT analysis. However, the FSDT analysis for this condition did illustrate the cost of collapsing the state-of-the world into two mutually exclusive categories: The loss of information resulted in a lower FSDT sensitivity scores when compared to FSDT sensitivity scores in conditions in which fuzzy responses were permitted. This indicates that perhaps the traditional procedure of forcing stimuli and responses into the 2 × 2 matrix shown in Figure 1 can distort the sensitivity of the observer to stimulus variation. Of course, in those cases in which the state-of-the-world is comprised of only the two possible categories of signal and nonsignal, traditional SDT remains an important tool for analysis. Indeed, in such particular cases the more general FSDT actually reduces to the traditional model (Hancock et al., 2000).

FSDT constitutes an advance of SDT because it reconsiders the perspective on the state-of-the-world as it is defined in traditional SDT. Traditional SDT forces membership into mutually exclusive categories, a classification that may not always be and indeed is probably not representative of many relevant task dimensions in actual operational environments. FSDT provides a quantitative model and procedures for incorporating the "fuzzy" nature of these real-world stimulus dimensions. We do not argue that FSDT is necessarily "superior" to SDT, only that it provides the more comprehensive case and, thus, a tool for a more fine-grained analysis of the state-of-the-world and, therefore, greater flexibility in determining the optimal observer response for a given magnitude of a stimulus event. The data in the present work demonstrates the viability of FSDT for modeling performance of multicategorical stimulus and response dimensions, and to explore the effects of variation in the structure of these dimensions. Note, however, that FSDT conceptualizes the multiple categories of stimulus and response differently than SDT. In the latter, the categories represent levels of stimulus magnitude (strength of the evidence variable) and levels of observer confidence in the judgment of the magnitude. However, FSDT conceptualizes multiple categories as different levels of membership in the category "signal" (**s**) and levels of perceived membership in that set (**r**).

The theoretical advance offered by FSDT is in the more accurate quantification of state-of-the-world and in the response (by a human or a machine) to such events. It addresses the inherent fuzziness in the state-of-the-world, the evidence variable, and the subsequent response to it. With respect to the latter, FSDT provides formalized procedures of fuzzy analysis of response formats heretofore analyzed as confidence ratings using crisp SDT procedures. In these cases, rather than representing confidence levels as separate criteria, FSDT represents them as differences in category membership in the set "response" (**r**). This similarity was previously noted in the original FSDT publications (Hancock et al., 2000; Parasuraman et al., 2000).

Note that as an extension or generalization of traditional SDT, FSDT is a model of decision making, specifically the relation of the stimulus structure and the response of the decision maker regarding that stimulus. As such, it is a statistical model of performance but it is *not* a model of cognition or perception. FSDT is thus not a new theory of human information processing. Rather it is a theory of the *relation* between the structure of the stimulus and response categories and how these are defined. It is no more a theory of information processing than the traditional SDT from which it was derived. Indeed, it is the relation of FSDT to SDT that was a motivation for the current empirical effort: to determine whether the assumptions that underlie SDT, and, thus, determine the structure of the assumed decision space, extend to FSDT.

It may be the case that the cognitive and perceptual mechanisms that drive fuzzy sensitivity or response bias may themselves be fuzzy, but such a test is beyond the scope of the current work, and it is not an element of the FSDT model itself. That is, like SDT, FSDT is independent of any particular theory of sensation, perception, or cognition. In this sense, FSDT differs from other applications of fuzzy logic to psychology in which fuzziness is incorporated into perceptual and cognitive mechanisms (e.g., Massaro, 1987).

The problem of aggregation. SDT and FSDT each provide summary analyses across multiple trials. Thus, they capture the macrostructure of overall responding and the reflections of the stable performance state of each observer. What they do not do is reveal the microstructure of each individual response or trial and, thus, the sequential dependencies of each of these discrete responses or trials. FSDT may represent one step toward the latter exposition in that it provides a greater focus on the uncertainty of the everyday world and the ambiguity within it. However, as we strive to understand both the task-based and person-based facets of an individual's performance, the manner in which each discrete response is contingent upon the events which precede it and prospectively that which the observer anticipates encountering, represents an important ongoing challenge to which FSDT looks to contribute.

Practical Applications

At the beginning of this work, we suggested that SDT was perhaps the most effective quantitative analytic method in the armory of the psychologist who ventures into the real-world. However, it is important to reiterate that traditional SDT represents only the selected case because it divides the world into necessarily discrete states of "signal" or "noise." We are all aware that the world itself resolves over space and time such that uncertainty—in the definition of the signal and not in the variability of the evidence variable—covaries with the magnitude of each of these linked dimensions of experience. For example, as an individual approaches an object, their recognition capacity often improves with increasing proximity.

Similarly, as any individual has greater observation time, their identification accuracy also often improves. Traditional SDT largely fails to capture these informational dimensions because the individual is constrained to respond after a certain time interval or at a certain distance from the stimulus event. This obligatory collapsing of the detection choice means that SDT is necessarily limited. Given that SDT actually plays such an important role in many critical real-world situations (e.g., Carlson, Gronlund, & Clark, 2008; Gronlund, Carlson, Dailey, & Goodsell, 2009; Hancock & Warm, 1989), any methodological or quantitative improvements are of vital practical and theoretical importance.

Here, we have explored and tested FSDT, which looks to circumvent some of these constraints of SDT. This we have done by using experimental conditions that in reality most favor the SDT paradigm. However, we can postulate many practical circumstances in which forcing people to "make" their decision, when ambiguity still dominates, clearly represents a disservice. We suggest rather, that in many actual performance circumstances an individual voluntarily seeks further sampling until the "to-be-detected" object reaches a threshold of signal membership. If their response is also less constrained than the traditional "yes-no" choice, additional value is again found in the FSDT paradigm. To conclude, the great contributions of SDT can be enhanced by the use of the broader FSDT model. This expansion, elaboration, and improvement we have sought to establish and advocate for here.

References

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 14,* 118–128. doi:10.1037/1076-898X.14.2.118

Dember, W. N. (1998). If you come to a fork in the road. In J. S. Warm, R. R. Hoffman, & M. F. Sherrick (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 629–649). Washington, DC: American Psychological Association. doi:10.1037/10290-030

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9,* 3–25. doi:10.3758/BF03196254

Green, D. M., & Swets, J. A. (1966/1988). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15,* 140–152. doi:10.1037/a0015082

Hancock, P. A. (2013). *A new law for time perception?* Manuscript submitted.

Hancock, P. A., Masalonis, A. J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomics Science, 1,* 207–230. doi:10.1080/14639220110038640

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. (2nd ed.). Mahwah, NJ: Erlbaum.

Masalonis, A. J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46,* 1045–1074. doi:10.1080/0014013031000121986

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). Cambridge: Cambridge University Press.

Metz, C. E. (1998). *ROCKIT 0.9B beta version*. Retrieved from http://metz-roc.uchicago.edu/MetzROC

Murphy, L., Szalma, J. L., & Hancock, P. A. (2003). Comparison of fuzzy signal detection and traditional signal detection theory: Approaches to performance measurement. *Proceedings of the Human Factors and Ergonomics Society, 47,* 1967–1971. doi:10.1177/154193120304701804

Murphy, L., Szalma, J. L., & Hancock, P. A. (2004). Comparison of fuzzy signal detection and traditional signal detection theory: Analysis of duration discrimination of brief light flashes. *Proceedings of the Human Factors and Ergonomics Society, 48,* 2494–2498. doi:10.1177/154193120404802111

Parasuraman, R., Masalonis, A. J., & Hancock, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42,* 636–659. doi:10.1518/001872000779697980

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39,* 14–29. doi:10.1518/001872097778940704

Stafford, S. C., Szalma, J. L., Hancock, P. A., & Mouloua, M. (2003). Application of fuzzy signal detection theory to vigilance: The effect of criterion shifts. *Proceedings of the Human Factors and Ergonomics Society, 47,* 1678–1682.

Stevens, S. S. (1961). The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 1–33). Cambridge, MA: MIT Press.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science, 182,* 990–1000. doi:10.1126/science.182.4116.990

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics*. Mahwah, NJ: Erlbaum.

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems*. New York, NY: Academic Press.

Szalma, J. L., & O'Connell, M. (2011). Fuzzy signal detection theory: A Monte-Carlo investigation. *Proceedings of Human Factors and Ergonomics Society, 55,* 1366–1369. doi:10.1177/1071181311551284

Tanner, W. P., & Birdsall, T. G. (1958). Definitions of d' and η as psychophysical measures. *Journal of the Acoustical Society of America, 30,* 922–928. doi:10.1121/1.1909408

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401–409. doi:10.1037/h0058700

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford, UK: Oxford University Press.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8,* 338–353. doi:10.1016/S0019-9958(65)90241-X

Appendix

Payoff Instructions for Fuzzy Criterion Manipulation

Instructions for the Payoff Manipulation in Experiments 1 and 2

Lenient Criterion

During this part of the experiment, you will receive (+10) points for each correct identification, which means that you correctly estimated the relative duration of the square.

However, you will be penalized (−10) points for each missed signal, which means that you underestimated the duration of the square.

Also, you will be penalized a (−1) point for each false alarm, which means you overestimated the duration of the square.

Conservative Criterion

During this part of the experiment, you will receive (+10) points for each correct identification, which means that you correctly estimated the relative duration of the square.

However, you will be penalized a (−1) point for each missed signal, which means that you underestimated the duration of the square.

Also, you will be penalized (−10) points for each false alarm, which means you overestimated the duration of the square.

Unbiased Criterion

During this part of the experiment, you will receive (+1) point for each correct identification, which means that you correctly estimated the relative duration of the square.

However, you will be penalized a (−1) point for each missed signal, which means that you underestimated the duration of the square.

Also, you will be penalized a (−1) point for each false alarm, which means you overestimated the duration of the square.

Received April 6, 2012
Revision received October 12, 2012
Accepted November 6, 2012 ■