# Examining the standard model of signal detection theory in motion discrimination

**Ju Liang**
School of Life Sciences,
University of Science and Technology of China,
Hefei, China

**Yifeng Zhou**
School of Life Sciences,
University of Science and Technology of China,
Hefei, China

**Zili Liu**
Department of Psychology, University of California,
Los Angeles, Los Angeles, CA, USA

**We studied a fundamental assumption in signal detection theory that is applied to motion discrimination. Using random-dot motion stimuli of 100% coherence, it is natural to represent the two directions as two normal distributions. Then a same–different task and a forced-choice task should give rise to consistent $d'$ estimates, because both tasks share the same underlying signal detection theory model. To verify this prediction, we used 4°, 8°, and 12° as the angular difference in motion discrimination. In a between-subjects design, we found the predicted result only with the 4° angular difference. With 8° and 12°, the estimated same–different $d'$ was 32% greater than the two-alternative forced-choice $d'$. In a subsequent within-subject design with counterbalancing, the first half of the data confirmed this finding. Interestingly, there was now within-subject consistency. Namely, the second task's $d'$ was comparable to the first task's, as if the first task's discrimination was carried over to the second task. This carryover effect diminished when the time gap between the two tasks was lengthened.**

## Introduction

The standard model of signal detection theory (SDT) is commonly assumed to be applicable to psychophysics in general. In this model, two Gaussian distributions with equal variance represent two categories of experimental stimuli, which are usually referred to as *noise* and *signal*. A participant's discriminating performance of these two categories is characterized by the distance between the two distributions divided by their shared standard deviation. This quantity is defined as the discrimination sensitivity $d'$.

The standard model is commonly assumed because, in the literature, $d'$ is typically calculated without verifying the underlying assumptions. When the assumptions are verified, oftentimes only the equal-variance assumption is verified. In this study, we present a case in which the equal-variance assumption appears to hold but other aspects of the assumptions in the standard model appear to be violated.

The case we consider is motion-direction discrimination, when random dots with 100% coherence are used. That is, in a motion stimulus made up of random dots, all dots move in a single direction with a constant speed and are visible inside an aperture during the entire stimulus presentation. Using the language of the standard model, the two distributions correspond to the two motion directions. The associated variance along each direction is primarily due to intrinsic uncertainty and internal noise in the visual system (Pelli, 1985), because the external noise is small due to the 100% coherence. Therefore, when the two motion directions are symmetric about a cardinal direction or ±45° axis, it should be expected that the two variances associated with the two directions are equal.

If the equal-variance assumption is expected to hold, how might one verify the validity of the standard model of SDT? In this study, we rely on a mathematical property that is derived by Macmillan and Creelman (2005) from the standard model. This property ensures that the $d'$ estimates from different experimental methods are self-consistent. This property is not restricted to motion discrimination, but applies gener-

Citation: Liang, J., Zhou, Y., & Liu, Z. (2016). Examining the standard model of signal detection theory in motion discrimination. *Journal of Vision, 16*(7):9, 1–12, doi:10.1167/16.7.9.

doi: 10.1167/16.7.9 Received January 14, 2015; published May 13, 2016 ISSN 1534-7362

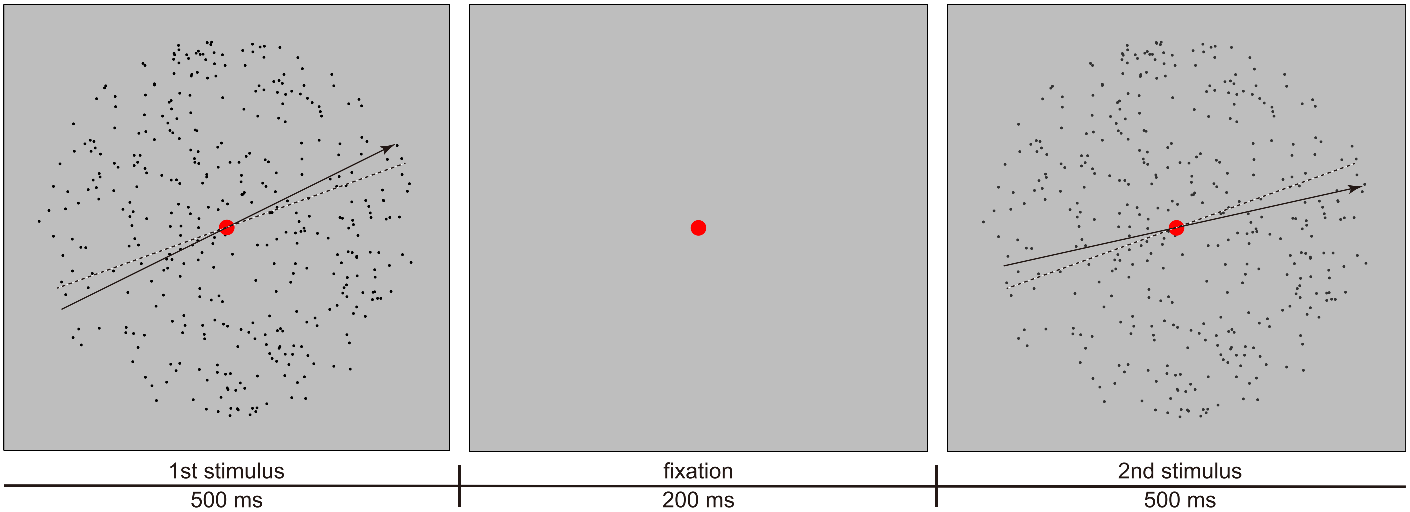


Figure 1. Illustration of one trial in the 2AFC task. In each of the two stimuli, all dots move along a single direction (represented by the solid arrow). The dashed line indicates the bisecting direction. Participants decide whether the first or second direction is more clockwise. In the corresponding same–different task, two samples are randomly drawn with replacement from the two directions. Participants decide whether the two stimulus directions are the same or different.

ally to all psychophysical stimuli. According to Macmillan and Creelman, the standard model has provided natural paradigms to study the detection of weak signals. Whether or not it applies to stronger signals remains an empirical question. In the current study, we tested the standard model across a range of signal strengths.

## The same–different task and estimation of d′

With the standard model as the starting point, one can derive a number of experimental tasks for participants to do in order to estimate the underlying $d'$. The two tasks we chose in this study were two-temporal-alternative forced-choice (2AFC) and temporal same–different. In the 2AFC task, two samples are randomly drawn *without* replacement from the two directions. In the same–different task, two samples are randomly drawn *with* replacement from the two directions. Figure 1 shows an example trial in a 2AFC task.

The 2AFC task is psychophysicist friendly, because it is straightforward to calculate the underlying $d'$. Let $S_1$ and $S_2$ denote the two motion stimuli from the two directions in a 2AFC task. Then the stimulus sequence in a trial will be either $<S_1S_2>$ or $<S_2S_1>$. According to the standard model, $d' = [Z(\text{proportion correct responses}|<S_1S_2>) + Z(\text{proportion correct responses}|<S_2S_1>)]/\sqrt{2}$. However, this 2AFC task is not participant friendly, which can be illustrated in the following example. Let us assume that the two motion directions are either 43° and 47° or 133° and 137° (0° is upward). In the first case, the response "more clockwise" corresponds to a rightward change of motion direction. However, when the two directions are 133° and 137°, the response "more clockwise" corresponds to a leftward change of motion direction. As a result, the response "more clockwise" is not intuitive.

In comparison, the same–different task is participant friendly, because the responses "same" and "different" are intuitive regardless of the stimulus directions. Hence, the same–different task is often used in the literature, particularly when multiple bisecting directions are involved. For example, Ball and Sekuler (1982, 1987) used the same–different task on motion perceptual learning, which is the first study on perceptual learning of motion discrimination.

The compromise of the participant-friendly design of the same–different task is that a nontrivial calculation is needed to recover the underlying $d'$. Macmillan and Creelman (2005) worked out how to calculate this $d'$ by assuming no bias in the same–different task. They offered two insights. The first is that the four possible stimulus combinations $<S_1S_1>$, $<S_2S_2>$, $<S_1S_2>$, and $<S_2S_1>$ correspond to four distributions with centers at (0, 0), ($d'$, $d'$), (0, $d'$), and ($d'$, 0), respectively. These centers are located in the two-dimensional space spanned by two orthogonal axes that represent the two intervals (Figure 2). The optimal, unbiased decision criterion consists of two lines $x = d'/2$ and $y = d'/2$ that divide the two-dimensional space into four quadrants. Because of zero bias, the overall proportion correct $p(c)$ is equal to the proportion correct with any of the four stimulus combinations. For example, $p(c)$ is equal to the proportion correct in responding "different" when the stimuli are $<S_1S_2>$. This proportion correct is

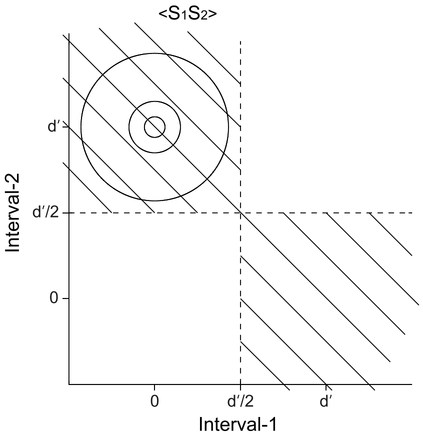


Figure 2. Schematic illustration of the decision space in the same–different task, with the two distributions $d'$ apart. The optimal (and unbiased) decision rule consists of two lines: $x = d'/2$ and $y = d'/2$. These two lines divide the stimulus space into four quadrants. The proportion correct is equal to the hit rate (responding "different") when the stimulus sequence is $<S_1S_2>$ or $<S_2S_1>$. The sequence $<S_1S_2>$ corresponds to the probability distribution that is centered at (0, $d'$), and the unbiased hit rate is the volume under this distribution in the upper left and lower right quadrants (the shaded area). Adapted from Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.)*. New York: Lawrence Erlbaum Associates.

equal, when the stimuli are $<S_1S_2>$, to the probability volume in the top left quadrant plus the probability volume in the bottom right quadrant. The probability volume in the upper-left quadrant is

$$\left(\int_{-\infty}^{\frac{d'}{2}} N(x)\,dx\right)\left(\int_{\frac{d'}{2}}^{+\infty} N(y-d')\,dy\right) = \left(\int_{-\infty}^{\frac{d'}{2}} N(x)\,dx\right)\left(\int_{-\infty}^{\frac{d'}{2}} N(y)\,dy\right) = \left[\Phi\left(\frac{d'}{2}\right)\right]^2, \quad (1)$$

where $N()$ is the normal distribution. The probability volume of the bottom right quadrant is $[\Phi(-d'/2)]^2$. The proportion correct is therefore

$$p(c) = \left[\Phi\left(\frac{d'}{2}\right)\right]^2 + \left[\Phi\left(-\frac{d'}{2}\right)\right]^2. \quad (2)$$

Given the symmetry $\Phi(-d'/2) = 1 - \Phi(d'/2)$, the recovered $d'$ is therefore

$$d' = 2z\left(\frac{1}{2}\left(1 + \sqrt{2p(c) - 1}\right)\right), \quad (3)$$

i.e., equation 9.3 of Macmillan and Creelman (2005), where $p(c) \geq 0.5$. Note that the $p(c)$ here is the unbiased proportion correct. In order to obtain this, the next insight from Macmillan and Creelman (2005) is needed.

Let us first define the hit rate $H = P(\text{"different"}|<S_1S_2> \text{ or } <S_2S_1>)$ and the false-alarm rate $F = P(\text{"different"}|<S_1S_1> \text{ or } <S_2S_2>)$. When there is no bias, $H = P(\text{"same"}|<S_1S_1> \text{ or } <S_2S_2>) = 1 - P(\text{"different"}|<S_1S_1> \text{ or } <S_2S_2>) = 1 - F$. Therefore, $Z(H) = Z(1 - F) = -Z(F)$. The unbiased proportion correct $p(c)$ is

$$p(c) = \Phi\Big(Z(H)\Big) = \Phi\left(\frac{Z(H) - Z(F)}{2}\right). \quad (4)$$

Having established that the proportion correct $p(c)$ is a function of $Z(H) - Z(F)$ in a same–different task for an unbiased observer, we can now introduce Macmillan and Creelman's second insight. They observed that the same–different receiver operating characteristic (ROC) is *approximately* a straight line with a 45° slope in the Gaussian space, $Z(H) = Z(F) + \text{constant}$. Hence, $Z(H) - Z(F)$ remains approximately constant regardless of the decision criterion. Therefore, even if an observer is biased, $Z(H) - Z(F)$ remains approximately the same as when there is no bias. The unbiased proportion correct $p(c)$ can be calculated from Equation 4 by directly plugging in $Z(H)$ and $Z(F)$ from the data. Consequently, $d'$ can then be recovered from Equation 3.

Although a participant does not have to be bias free in order for this calculation to be valid, the participant has to use the same criterion when responding "different" no matter if the stimulus sequence is $<S_1S_2>$ or $<S_2S_1>$. That is to say, the calculation assumes that $P(\text{"different"}|<S_1S_2>) = P(\text{"different"}|<S_2S_1>)$. This is an assumption that can be verified empirically.

What can also be empirically verified is the following inequality. Macmillan and Creelman (2005) showed that the same–different proportion correct $p(c)_{\text{SD}}$ and the bias-free yes–no proportion correct $p(c)_{\text{yes-no}}$ obey the equation

$$p(c)_{\text{SD}} = p(c)^2_{\text{yes-no}} + \left[1 - p(c)_{\text{yes-no}}\right]^2. \quad (5)$$

Equation 5 can be obtained as follows. In a yes–no task, the bias-free criterion is located at $d'/2$. Hence, the bias-free $p(c)_{\text{yes-no}} = H = \Phi(d'/2)$ (Equation 4); from Equation 2 we then obtain Equation 5. Starting from Equation 5, we have $p(c)_{\text{SD}} - 1/2 = 2(p(c)_{\text{yes-no}} - 1/2)^2$. Let us assume that $p(c)_{\text{yes-no}} \geq 1/2$, then $0 \leq 2(p(c)_{\text{yes-no}} - 1/2) \leq 1$. Therefore, $p(c)_{\text{SD}} - 1/2 \leq p(c)_{\text{yes-no}} - 1/2$. Hence, $p(c)_{\text{SD}} \leq p(c)_{\text{yes-no}}$. Given that $p(c)_{\text{yes-no}} = \Phi(d'/2)$ (Equation 4) and $p(c)_{\text{2AFC}} = \Phi(d'/\sqrt{2})$, we have $p(c)_{\text{SD}} \leq p(c)_{\text{yes-no}} < p(c)_{\text{2AFC}}$. That is, the same–different task is always more difficult than the corresponding 2AFC task.

Now that the $d'$ can be recovered from the 2AFC and same–different tasks separately, we can verify the consistency between the two $d'$ values thus computed. We used 2AFC and same–different tasks because these are common tasks in motion discrimination and motion perceptual learning in the literature. We compared these two tasks with both between-subjects and within-subject designs. We also used angular differences of 4°, 8°, and 12° in order to parametrically manipulate the signal strength.

Recall that, in recovering the same–different $d'$, Macmillan and Creelman (2005) assumed the optimal decision rule. They also introduced a suboptimal rule, the differencing rule. In the case of motion discrimination, differencing takes the directional difference between the first and second stimulus. If the magnitude of this difference is below a certain threshold, a "same" response will be chosen; otherwise, "different" will be chosen. Geometrically, this policy is to carve out the space in Figure 2 by two lines: $y - x = \pm$threshold. The region between these two lines corresponds to the stimuli with the "same" responses. The regions outside correspond to the "different" responses.

Other suboptimal rules were subsequently introduced and examined by Petrov (2009). Petrov's significant contribution was working out, for each of the eight suboptimal decision rules, the equalities or inequalities for the proportions of responding "different" between the stimulus sequences $<S_1S_1>$ and $<S_2S_2>$ and between $<S_1S_2>$ and $<S_2S_1>$.

It is important to note that in inferring the underlying $d'$ in the same–different task from the same behavioral data, the recovered $d'$ from the optimal decision rule is always smaller than a $d'$ recovered from a suboptimal model. This is because, in order for a suboptimal rule to produce the same level of behavioral performance as the optimal rule does, the corresponding $d'$ necessarily has to be greater (i.e., the task has to be easier) in order to compensate for the inefficiencies of the suboptimal model (Petrov, 2009).

## Summary of the results in this study

To anticipate, we found in our first experiment with a between-subjects design that the two tasks gave rise to consistent $d'$ estimates only for the weak signal case, at the 4° angular difference. When the angular difference increased to 8° and 12°, the same–different $d'$ as recovered using the optimal model became greater than the 2AFC $d'$.

In a subsequent within-subject experiment (Experiment 2), half of the participants performed the 2AFC task first and the same–different task second. The other half of the participants ran in the opposite order. The first half of the data from all participants was consistent with the data in Experiment 1. Interestingly, however, after the participant switched immediately from the first to the second task, the $d'$ estimate in the second task was comparable with that in the first task, indicating that the participant adopted a consistent computation from the first to the second task. More specifically, when the angular differences were 8° and 12°, the participants whose first task was same–different gave rise to a higher average $d'$ from both tasks than those participants whose first task was 2AFC. When the angular difference was 4°, the $d'$ estimates remained consistent within and between subjects.

In Experiment 3, the within-subject Experiment 2 was repeated except that the temporal gap from the first to the second task was increased from 5 min to 1 week. The carryover effect from the first to the second task was much reduced.

# Experiment 1: Between-subjects design

## Methods

### Stimuli and tasks

The stimuli and tasks were similar to those used by Ball and Sekuler (1982, 1987), Huang, Lu, Tjan, Zhou, and Liu (2007), Liu (1995, 1999), Liu and Weinshall (2000), and Wang, Zhou, and Liu (2013). The dots were darker than the background. Specifically, 400 dots were randomly distributed within a circular aperture of 8° diameter (262 pixels). Each dot was 0.09° in size (a $3 \times 3$ pixel square), and the dots' luminance was 0.0 cd/m$^2$. A central red fixation disk had a diameter of 0.5° (16 pixels) and a luminance of 5.6 cd/m$^2$. The background luminance was 22.0 cd/m$^2$. In each stimulus, all dots moved along a single direction with a speed of 10°/s. The duration of each stimulus was 500 ms, and the interstimulus interval was 200 ms (Figure 1).

In the 2AFC task, the motion directions of the two stimuli differed by either 4°, 8°, or 12° in a blocked design. The participants fixated at the central red disk and decided which of the two intervals had the more clockwise direction. In the same–different task, the two motion directions, when different, differed by 4°, 8°, or 12° in a blocked design. The participant decided whether the two directions were the same or different. Trial-wise feedback was provided by a computer beep in both tasks.

### Procedure

Six groups of randomly assigned participants took part in this experiment. Groups 1, 2, and 3 participated in the same–different task. Groups 4, 5, and 6 participated in the 2AFC task. The three groups in each

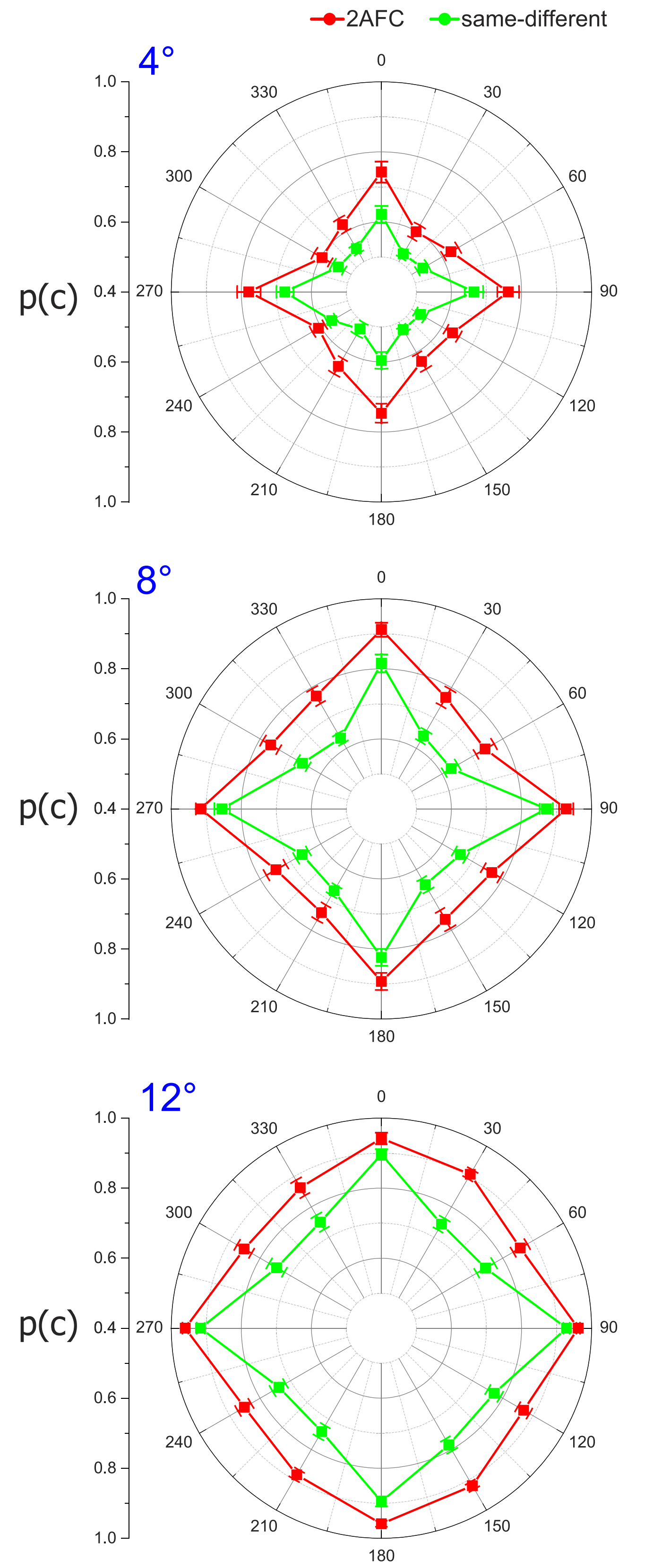


Figure 3. Average proportion correct for the same–different and 2AFC tasks, and for the 4°, 8°, and 12° angular differences, respectively. Error bars are standard errors of the means.

task used 4°, 8°, and 12°, respectively. All groups were tested along the following 12 average directions: 0°, 30°, 60°, . . . 330°. These 12 directions were run in a blocked design, in 24 blocks. The first 12 blocks corresponded to the 12 directions, randomly ordered. The second 12 blocks repeated the same sequence backward, for counterbalancing. Each block had 32 trials. Hence, each participant ran 768 trials in total that took less than 1 hr to complete. Participants in all experiments in the current study practiced 15–30 trials before their data collection.

#### *Participants*

A total of 108 students from the University of Science and Technology of China, Hefei, participated in accordance with the Helsinki Declaration. They were unaware of the purposes of the study and had normal or corrected-to-normal visual acuity.

#### *Apparatus*

The stimuli were presented on a 17-in. Sony Multiscan G220 monitor. The monitor resolution was 1024 × 768 pixels, and the refresh rate was 100 Hz. The participants viewed the stimuli binocularly from a chin rest. The viewing distance was 60 cm. MATLAB software (The Mathworks, Inc., Natick, MA) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) were used.

### Results and discussion

Figure 3 shows the averaged proportion correct, for each of the 12 directions, each task, and each angular difference. Although subsequent analyses will focus on the discrimination sensitivity $d'$, it is nevertheless informative to visualize the proportion-correct data, whose calculation requires no assumptions. The 2AFC proportion correct was greater than the same–different one, consistent with Macmillan and Creelman's (2005) prediction, even though this proportion correct was the actual and not the unbiased one.

#### *Calculating d′ using the independent and differencing rules*

We now calculate the discrimination sensitivity $d'$ for each of the two tasks (Figure 4), and using the optimal independent model and the suboptimal differencing model for the same–different task. If a participant's hit rate $H$ or false-alarm rate $F$ was 1 or 0, correction was made by subtracting $1/2n$ from 1 or adding $1/2n$ to 0, where $n$ is the number of trials used to calculate this particular rate. It turned out that 6% of the data needed

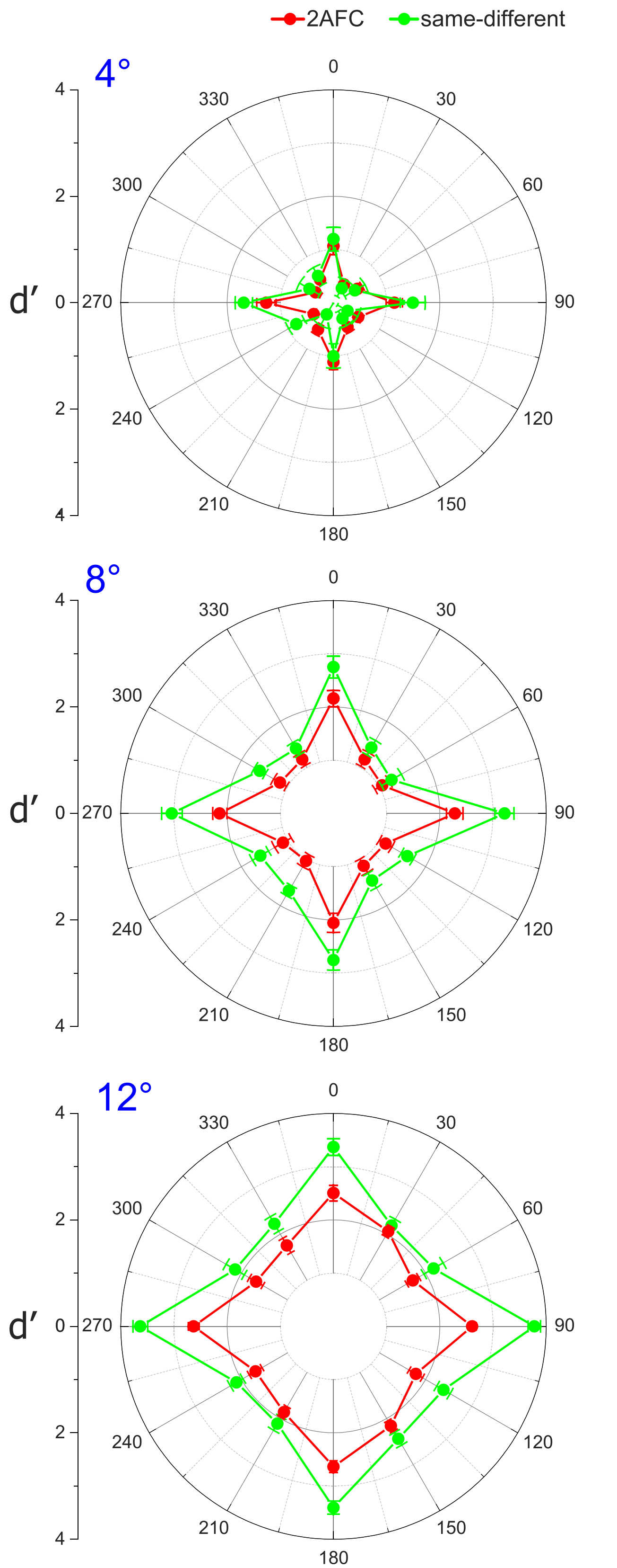


Figure 4. Discrimination sensitivity $d'$ for each of the two tasks, and for the 4°, 8°, and 12° angular differences.

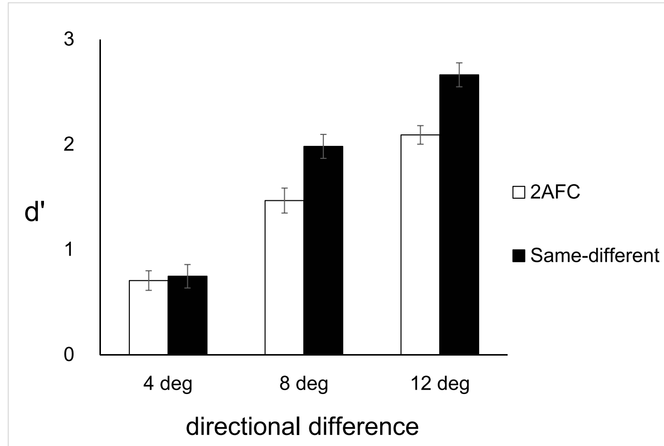


Figure 5. Recovered $d'$ values for the two tasks for each of the three directional or angular differences.

correction, which will be addressed later. From $Z(H)$ and $Z(F)$, the same–different $d'$ could be calculated using the independent model, which is the optimal model that assumes that the two stimuli in a trial are perceptually independent and separable (Macmillan & Creelman, 2005). For the 2AFC task, the $d'$ calculation was straightforward: $d' = (Z(H) - Z(F))/\sqrt{2}$.

On these $d'$ scores, ANOVA was performed with the following factors: 12 motion directions $\times$ 2 tasks $\times$ 3 angular differences. All effects except the three-way interaction—$F(22, 1122) = 1.06$, $p = 0.39$—were significant. Specifically, the main effect of angular difference was significant, as expected, $F(2, 102) = 121.19$, $p = 1.51 \times 10^{-27}$, with the mean $d'$ for 4°, 8°, and 12° at $d' = 0.73$, 1.72, and 2.38, respectively. The main effect of motion direction was also significant, $F(11, 1122) = 112.28$, $p = 3.54 \times 10^{-172}$. This was also expected, since discrimination along cardinal directions ($d' = 2.29$) was better than along oblique directions ($d' = 1.27$). Importantly, the main effect of task was significant, $F(1, 102) = 18.41$, $p = 0.00004$. The same–different $d'$ was greater than the 2AFC $d'$ by 26% (1.80 vs. 1.42). The interaction between task and angular difference was also significant, $F(2, 102) = 3.68$, $p = 0.03$. This interaction revealed that, while the $d'$ values of the two tasks at the 4° directional difference were comparable ($d' = 0.75$ vs. 0.71, 6% difference), their difference became greater at larger angular differences (8°: $d' = 1.98$ vs. 1.47, 35% difference; 12°: $d' = 2.66$ vs. 2.09, 27% difference; Figure 5).

A main assumption in calculating $d'$ of the same–different task was that $P(\text{“different”}|<S_1S_2>) = P(\text{“different”}|<S_2S_1>)$. In order to check whether this assumption held, we compared these two measures from all 54 participants. With all three angular differences collapsed, $t(53) = 0.78$, $p = 0.44$. Given that the 4° data already showed consistency, we also looked at the 8° and 12° data separately. There, $t(35) = 0.59$, $p$

$= 0.56$, meaning that there was little asymmetry between the two "different" responses (but see later the more detailed analysis using the methods from Petrov, 2009).

We also wondered whether the task difference occurred only for cardinal or oblique motion directions. To find out, we conducted another ANOVA with 2 motion directions (cardinal, oblique) × 2 tasks × 3 angular differences. All effects except the three-way interaction—$F(2, 102) = 1.62$, $p = 0.20$—were statistically significant, with the largest $p = 0.02$, $F(2, 102) = 4.17$, for the two-way Task × Angular Difference interaction. This means that the task difference increased with the angular difference for cardinal and oblique directions alike.

In order to verify the robustness of our results so far, we performed two additional analyses. We first replaced the independent model with the differencing model (Macmillan & Creelman, 2005) when calculating the same–different $d'$. The differencing model is suboptimal with two fixed stimuli but optimal when the average of the two stimuli changes from trial to trial in a roving experiment (Dai, Versfeld, & Green, 1996). We performed an ANOVA as we did for the independent-model analysis, and found that the only difference from the previous ANOVA was the three-way interaction. Using the differencing model, this effect was significant: $F(22, 1122) = 1.63$, $p = 0.033$, as compared to $F(22, 1122) = 1.06$, $p = 0.39$, in the independent model. The main message that the task difference became substantial with larger angular differences, however, remained robust. We should also mention that, using the differencing model, the mean $d'$ was 2.21, which was significantly greater than the 1.80 in the independent model. This confirms that the optimal decision rule gives rise to a smaller estimate of $d'$ than a suboptimal rule.

The second additional analysis we performed was as follows. Recall that we made corrections when the hit and false-alarm rates were 1 or 0, in order to be able to calculate the $Z$ scores. Since the 2AFC accuracy was greater than the same–different accuracy, this correction was expected to be more frequent for the 2AFC task. In order to avoid the corrections that were arbitrary and asymmetric between the two tasks, rather than calculating for each of the 12 directions we recalculated the hit and false-alarm rates for two directions only: oblique and cardinal. This new calculation turned out to be sufficient to avoid any corrections. We then repeated the ANOVA with motion direction, task, and angular difference as the main factors. The results were qualitatively very similar as before. Namely, all effects except one were significant, with the largest significance being $p = 0.001$. The exception was the Angular Difference × Task interaction, $F(2, 102) = 2.13$, $p = 0.12$. At 4°, the $d'$ estimates were 2AFC $d' = 0.63$, same–different $d' = 0.97$, $t(34) = 2.61$, $p = 0.013$. At 8° and 12°, the $d'$ estimates were more different: 2AFC $d' = 1.63$, same–different $d' = 2.22$, $t(70) = 4.80$, $p = 0.000009$. Therefore, although the interaction effect was only marginal, we maintain that the analysis was reasonably consistent with the rest of the analyses.

Recall that $Z(H)$ and $Z(F)$ were calculated because the bias-free proportion correct was needed in order to calculate $d'$. Here, as an approximation, we used the actual proportion correct to calculate $d'$ in order to avoid the corrections when $H = 1$ or $F = 0$. An ANOVA was performed similarly as when we used the independent model with 12 directions. Completely consistent results were obtained. Namely, all effects except the three-way interaction were significant. In particular, the Task × Angular Difference interaction was significant, $F(2, 102) = 3.10$, $p < 0.05$.

#### *Calculating d′ using suboptimal rules from Petrov (2009)*

We applied the methods in Petrov (2009) and categorized the participants' potential suboptimal strategies in the same–different task by checking whether or not $P(\text{"different"}|<S_1S_1>) = P(\text{"different"}|<S_2S_2>)$ and whether or not $P(\text{"different"}|<S_1S_2>) = P(\text{"different"}|<S_2S_1>)$. Whenever possible, we also calculated the corresponding $d'$ values.

We should point out from the outset that a $d'$ estimated from a suboptimal strategy is necessarily greater than that from the optimal model, when the same behavioral data are used. This is because a task has to be easier (with a greater $d'$) in order to compensate for the inefficient suboptimal strategy. Given that our main result so far was that for larger angular differences, the same–different $d'$ estimated from the optimal model was greater already than the 2AFC $d'$, then the even greater same–different $d'$ estimated from a suboptimal strategy would not change the result. In the interest of space, we skip the details but note that, for the 4° angular difference, the same–different $d'$ values re-estimated from the suboptimal strategies were still statistically comparable to the 2AFC $d'$.

## Experiment 2: Within-subjects design with a 5-min gap between tasks

### Methods

In Experiment 1, 12 average directions around the clock were used. The purpose was to obtain a result

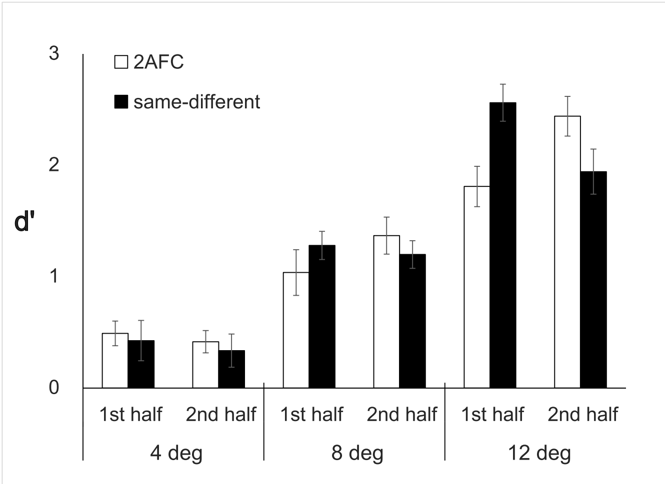


Figure 6. Recovered $d'$ values for the first and second halves of Experiment 2. The results from the first half were consistent with those from Experiment 1: While the $d'$ values were comparable between the two tasks at the 4° angular difference, at 8° and 12° the same–different $d'$ values were greater than the 2AFC $d'$ values.

that was not direction specific. Using 12 directions, however, makes the number of trials per direction small, namely 64. In addition, although data collection across the 12 directions was blocked, there was uncertainty at transitions between blocks, because the new block's average direction was randomly different and therefore unpredictable. Although this blocked design was used for both tasks, the transition uncertainty may not have been the same for the two tasks because the same–different task might be less vulnerable due to its participant-friendly nature. Another potential issue is that, among the 12 directions total, the four cardinal directions yielded nearly ceiling performance for the 12° condition. This at times created difficulties for calculating $Z(H)$ and $Z(F)$ such that arbitrary correction was needed.

In this experiment, we used only two average directions instead: 45° and 135°. A given participant used one average direction for the same–different task and the other direction for the 2AFC task. Hence, whether clockwise would be left- or rightward would be fixed and less confusing for this participant. In addition, by using ±45° as the average or bisecting orientations, the two stimulus directions (e.g., 41° and 49°) would be symmetric about 45° or 135°. Therefore, the assumption that these two directions were represented by two distributions of equal variance would be reasonable.

Three groups of fresh participants, 20 in each group, were similarly recruited as in Experiment 1. Participants did the same–different task along one average direction and the 2AFC task along the other. The direction–task pairings and the task sequence were counterbalanced. These three groups used 4°, 8°, and 12° angular differences, respectively. The number of trials per task was 360. A participant took a 5-min break between tasks.

## Results and discussion

Similarly as in Experiment 1, $d'$ was calculated where the optimal independent model was used for the same–different task. We first verified that there was no significant difference between the proportion of $P(\text{“different”}|\langle S_1S_2\rangle)$ and $P(\text{“different”}|\langle S_2S_1\rangle)$ responses for each of the three angular differences. The smallest $p$ value was $p = 0.19$, $t(19) = 1.37$, for the 8° angular difference. This confirmed that the independent model was applicable.

We then looked at the data from the first half of the experiment, where half of the participants ran the 2AFC task and the other half the same–different task. This was in effect a between-subjects comparison, similar to Experiment 1. A two-way ANOVA was performed with angular difference and task as the main factors. The main effect of angular difference was significant, $F(2, 54) = 55.40$, $p = 8.27 \times 10^{-14}$, as expected (4°: $d' = 0.46$; 8°: $d' = 1.16$; 12°: $d' = 2.19$). The main effect of task was also significant, $F(1, 54) = 5.29$, $p = 0.025$. In fact, the $d'$ values (same–different: $d' = 1.42$; 2AFC: $d' = 1.11$) were comparable to those in Experiment 1 averaged along the eight oblique directions ($d' = 1.69$ and 1.15). The interaction effect was marginally significant, $F(2, 54) = 3.12$, $p = 0.052$. That is to say, when the angular difference was 4°, the $d'$ values of the two tasks were comparable (same–different: $d' = 0.43$; 2AFC: $d' = 0.49$). When the angular differences were greater, the $d'$ differences between the two tasks became larger (8°: 1.28 and 1.04; 12°: 2.56 and 1.81). These numbers were again consistent with the $d'$ values along oblique directions in Experiment 1 (Figure 6).

We similarly analyzed the data from the second half of the experiment. The main effect of angular difference remained significant, as expected: $F(2, 54) = 66.79$, $p = 2.50 \times 10^{-15}$. The main effect of task was marginally significant, $F(1, 54) = 3.78$, $p = 0.057$. Interestingly, however, the sign of the task difference was reversed from the first half. That is, while in the first half the same–different $d'$ (1.42) was greater than the 2AFC $d'$ (1.11), in the second half the two $d'$ values became 1.16 (same–different) and 1.41 (2AFC). Another way to describe the results is that the participants who ran the same–different task first and 2AFC second had $d'$ values of 1.42 and 1.41, whereas those participants in the opposite task order had $d'$ values of 1.11 (2AFC) and 1.16 (same–different). This indicates that, when running the second task, the participants may have

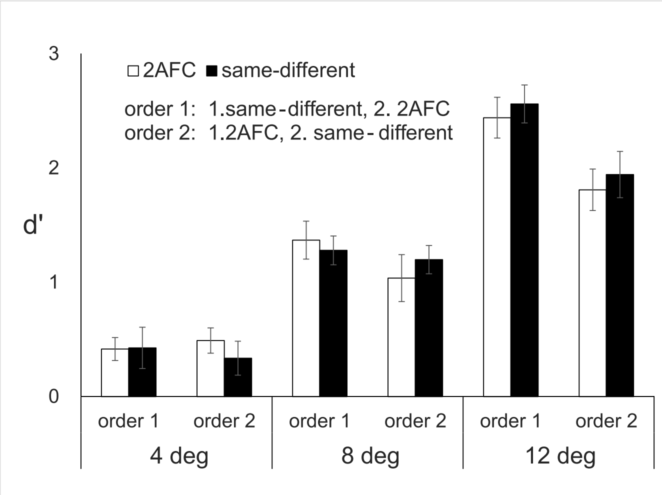


Figure 7. Recovered $d'$ values in Experiment 2 with the two task orders separately represented. For any particular participant, the second task's $d'$ was comparable to the first task's $d'$. In addition, while the $d'$ values were comparable at the 4° difference, for the 8° and 12° angular differences the participants who did the same–different task first enjoyed an overall greater $d'$ than those who did the 2AFC task first.

carried over their strategy from the first task, such that the recovered $d'$ values of the two tasks were comparable within subjects. It is also interesting to note that this carryover effect from one task to the next was also between directions that were 90° away. Due to this carryover effect, the first participant group (task order: same–different, 2AFC) had a greater overall $d'$ than the second group when the two tasks were combined, $t(38) = 3.49$, $p = 0.001$ (Figure 7).

Given the analyses of the two separate halves, the results from the following overall analysis using all the data were then not surprising. The main factors of the ANOVA were angular difference, task, and task order. The main effect of angular difference was significant, $F(2, 108) = 121.20$, $p = 2.50 \times 10^{-28}$, as expected. The Task × Task Order interaction was significant, $F(1, 108) = 9.03$, $p = 0.003$. The three-way interaction was also significant, $F(2, 108) = 3.83$, $p = 0.025$. All the remaining effects were nonsignificant, with $F$ values smaller than 1.

The conditions with the 12° angular difference can well illustrate these effects (Figure 7). When the 2AFC task was second, its $d'$ was greater than when it was the first task run by the other half of participants: $d' = 2.44$ versus 1.81, $t(18) = 2.49$, $p < 0.025$. This enhanced performance did not seem simply due to a sequential learning effect, because a similar comparison for the same–different task gave rise to the opposite effect. That is, the same–different task as second gave rise to a $d'$ smaller than as first: $d' = 1.94$ versus 2.56, $t(18) = 2.36$, $p < 0.03$. In order to take into consideration that the same–different $d'$ depends on the underlying decision rules (Petrov, 2009), we looked at the accuracies of the same–different task that are unchanged for all the assumed decision rules. We found that the same–different task as first was 0.81 in proportion correct, as compared to 0.72 as the second task, $t(18) = 2.23$, $p < 0.05$. This confirms that the effects were not simply due to the sequential learning effect. A similar comparison for the 2AFC task yielded 0.88 versus 0.94, $t(18) = 1.89$, $p = 0.07$.

Given that the same–different performance, calculated either in $d'$ using the optimal model or in proportion correct, was worse as the second task than as the first task, we wondered whether there was any difference in the decision rules used by the participants who ran the same–different task as the second task compared to the other half, who ran it as the first task. We again used the methods of Petrov (2009) to categorize the decision rules the participants used, and found that the majority of the participants were in two categories of suboptimal decision rules, for both task orders (28/30 and 27/30). However, in terms of the number of participants in each of these two categories, there was hardly any difference between task order—Category 1: 15 (same–different as Task 1) and 12 (same–different as Task 2); Category 2: 13 and 15. Unfortunately, therefore, using the methods of Petrov (2009) to narrow down the possible same–different strategies used by the participants did not help answer the question of what strategies participants might have used in the same–different task after performing the 2AFC task compared to before the 2AFC task. It is also an open question exactly how the first task influenced the second task such that the second task's $d'$ became comparable to the first's.

## Experiment 3: Repeating Experiment 2 with a 1-week gap between tasks

Results in Experiment 2 suggested that when the first task was followed immediately by the second task in a within-subject design, participants transferred their discrimination to the second task. It would therefore predict that, if the time gap between the two tasks were lengthened, this transfer would be weakened such that the $d'$ values for both tasks would depend less on the task order. Experiment 3 tested this prediction. Given that the $d'$ difference was biggest between the two task orders for the 12° angular difference, we focused on this angular difference in this experiment. This experiment was otherwise identical to Experiment 2, with 24 fresh participants similarly recruited.

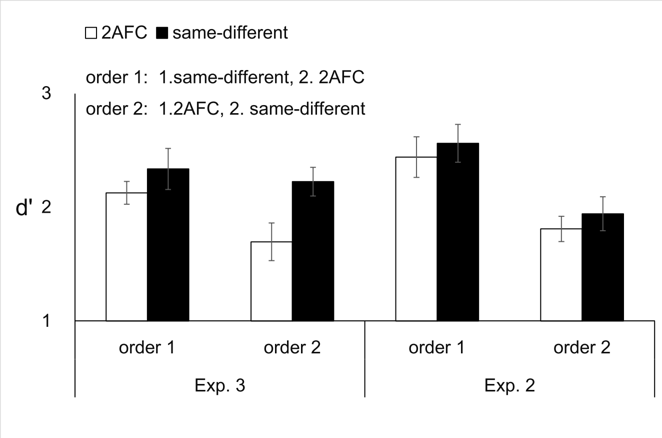


Figure 8. Left: $d'$ data from Experiment 3 (1-week gap), where 12 participants ran the same–different task first (Order 1) and the remaining 12 participants ran the 2AFC task first (Order 2). The same–different task gave rise to a higher $d'$ irrespective of the task order. Right: For comparison, Experiment 2's data with a temporal gap of 5 min between tasks.

Figure 8 shows the $d'$ values, which were calculated identically as in Experiment 2, using the optimal model for the same–different task. A two-way ANOVA with task and task order as the main factors showed that there was a significant main effect of task: 2AFC $d' = 1.91$, same–different $d' = 2.28$, $F(1, 22) = 7.75$, $p = 0.011$. The main effect of task order was not significant, $F(1, 22) = 1.06$, $p = 0.32$. The interaction was not significant either, $F(1, 22) = 1.44$, $p = 0.24$.

The data nevertheless indicate a weak asymmetry of task transfer. Namely, when the 2AFC task was second, its $d'$ was numerically greater than when it was first: $d' = 2.13$ vs. 1.70, $t(22) = 1.72$, $p = 0.05$ (one-tailed). A similar comparison for the same–different task did not reach significance. However, since this difference was only marginally significant using a one-tailed $t$ test, and since the interaction was nonsignificant, the implication of the asymmetric transfer requires further study.

In summary, when the temporal gap between the two tasks was lengthened from 5 min to 1 week, the carryover from the first to the second task was much reduced.

## General discussion and conclusions

In the literature, a number of studies on various perceptual modalities have compared $d'$ values that were computed from same–different and 2AFC or yes–no tasks. In most of these studies, the same–different $d'$ differed from the yes–no or 2AFC $d'$ by only $-14\%$ to 3%, suggesting reasonable applicability of the standard model of SDT (Chen & Macmillan, 1990; Creelman & Macmillan, 1979; Hautus & Irwin, 1995; Macmillan, Goldberg, & Braida, 1988). However, in an auditory study, Creelman and Macmillan (1979) found that their same–different $d'$ was 50% higher, where the underlying model failed (see also Taylor, Forbes, & Creelman, 1983). In visual motion perception, the applicability of the standard model of SDT has been little studied, as far as we know.

At the same time, however, the same–different $d'$ might have been calculated incorrectly in some studies of visual motion perception. For example, Ball and Sekuler (1987) calculated their same–different $d'$ in their pioneering study of motion-discrimination learning as follows: "Hit rates and false alarm rates for identifying 'different' trials in each block were converted by standard methods into $d'$, to provide a measure of discrimination performance (Green and Swets, 1966)" (p. 954). To our knowledge, nevertheless, Green and Swets (1966) did not discuss the same–different design or how to calculate $d'$ in that design. According to R. Sekuler (personal communication, February 2016), the $d'$ values for Ball and Sekuler (1987) may have been calculated by defining $d' = Z(\text{hit}) - Z(\text{false-alarm})$, where hit $= P(\text{"different"}|<S_1S_2>$ or $<S_2S_1>)$ and false-alarm $= P(\text{"different"}|<S_1S_1>$ or $<S_2S_2>)$. To be fair, in some of the studies by Liu (Liu, 1999; Liu & Weinshall, 2000), $d'$ was incorrectly calculated as $d' = Z(\text{hit}) - Z(\text{false-alarm})$.

In the current study, we tested the standard model of SDT by asking whether the same–different $d'$ was the same as the 2AFC $d'$ when the underlying stimuli were the same. We also tested our hypothesis with weak signals (angular difference = 4°), when the standard model was expected to work (Macmillan & Creelman, 2005), and with stronger signals (8° and 12°), when it was unknown whether the standard model would work or not.

Our data suggest that the standard model for strong signals did not work in motion discrimination. This is because when the angular difference was 4°, the same–different and 2AFC tasks gave rise to comparable $d'$ estimates. However, when the angular differences were 8° and 12°, the recovered $d'$ values from the same–different task were greater than those from the 2AFC task. Here the same–different $d'$ values were recovered using the optimal decision rule. If suboptimal decision rules were used instead, the recovered same–different $d'$ values were numerically even greater. Hence, the discrepancy between the same–different $d'$ and the 2AFC $d'$ was even greater for 8° and 12° (for 4°, the consistency between the two tasks remained). That is to say, for stronger signals, the standard model of SDT could not apply no matter which decision rule, optimal or otherwise, was used.

One possibility for why the two tasks gave rise to inconsistent $d'$ values with stronger signals is that the same–different ROC in the Gaussian space could no longer be approximated by a 45° straight line. This possibility can be verified experimentally, using a rating experiment in a same–different task. We will investigate this possibility in the future.

However, even if the same–different ROC turns out to be nonlinear, and the same–different $d'$ can no longer be estimated as shown earlier in this article, the participants still likely performed the two tasks with different $d'$ values. This is because, as shown in Experiment 2, the same $d'$ recovery method (the optimal model) for the same–different task gave rise to very different results for 8° and 12° depending on the order of the two tasks. This means that the underlying computations used in these two tasks were different when they were not run one immediately after another. When the two tasks were run one immediately after another, carryover occurred such that the second task's $d'$ was comparable to the first task's. This result is intriguing, because it suggests that these comparable within-subject $d'$ values may not be coincidental. It further indicates that the same–different $d'$ as recovered by the optimal model per Macmillan and Creelman (2005) may be reasonably accurate, and that the same–different ROC might not be severely nonlinear after all. This is another motivation for us to carry out an empirical study on the linearity of the same–different ROC in the future.

As a result of this carryover effect, the task sequence <same–different, 2AFC> gave rise to a higher averaged $d'$ than the sequence <2AFC, same–different>. Recall that in the 2AFC task, the two motion directions were symmetric about either 45° or 135°. This means that the two probability distributions corresponding to the two directions in the standard SDT model were likely to be identical in shape and differed only in their respective means. If one makes an additional and reasonable assumption that both distributions are Gaussians, then $d'$ is definable. This means that the 2AFC task may have a definable $d'$ (and not just $d'_{FC}$) that is also straightforward to calculate. Therefore, the 2AFC $d'$ values calculated in the current study are likely to be reasonably accurate.

Let us assume that the 2AFC $d'$ values were indeed accurate. Then—as shown in Experiment 2 for the 8° and 12° angular differences—when the 2AFC task was run first, the average $d'$ was 1.42. However, when the 2AFC task was run after the same–different task, $d' = 1.91$, probably because the preceding same–different task gave rise to a $d' = 1.92$. This was a remarkable result in that the 2AFC $d'$ increased by 35% when it immediately followed the same–different task, as compared to following no task at all. This suggests that in the original 2AFC task, there was internal noise that could be substantially reduced after only 360 trials of the same–different task. This type of learning appears different from the motion perceptual learning that exhibits gradual improvement through days of training (Epstein, 1967; Fahle & Poggio, 2002; Gibson, 1969; Sagi, 2011; Sasaki, Nanez, & Watanabe, 2009). This difference from traditional perceptual learning is further illustrated when we look at the opposite task order. When the same–different task followed the 2AFC task, the same–different $d' = 1.57$, which was close to the preceding 2AFC task's $d' = 1.42$. As a second task, this same–different $d' = 1.57$ was lower than when it was the first task, with $d' = 1.92$.

The pattern of results in Experiments 2 and 3 suggests that the two tasks used different strategies when run separately or with larger angular differences when run sequentially with a large time gap. Yet when one task immediately followed the other, the strategy appeared to be transferred such that the resultant $d'$ values were comparable.

Unfortunately, however, it remains unclear exactly how the strategies in the same–different task differed when it was the second task as compared to when it was the first task. Likewise, it is unclear how exactly the 2AFC task was performed such that its $d'$ was substantially greater as the second task than the first task.

To summarize, the standard SDT model is expected to work well with weak signals in basic psychophysical tasks, including yes–no, forced-choice, and same–different. Whether or not the model applies well to stronger signals remains an empirical question. Our study suggests that, in motion discrimination with stronger signals, the standard SDT model yields inconsistent results across tasks.

*Keywords: motion, direction discrimination, forced-choice, same–different, signal detection theory (SDT)*

## Acknowledgments

We thank Dr. Alex Petrov for helpful discussions regarding his work (2009) and for his codes calculating $d'$ under a number of suboptimal decision rules. We also thank Dr. Bob Sekuler for his kind correspondence regarding the same–different $d'$ calculations used by Ball and Sekuler (1987). This research was supported in part by the Natural NSF of China to ZL (NSFC 31228009) and to YZ (NSFC 31230032). This research was also supported by a fellowship to ZL at the Hanse Institute for Advanced Study, Delmenhorst, Germany.

Commercial relationships: none.
Corresponding author: Yifeng Zhou.
Email: zhouy@ustc.edu.cn.

Address: School of Life Sciences, University of Science and Technology of China, Hefei, China.

## References

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science, 218*(4573), 697–698.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*(6), 953–965.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Chen, H., & Macmillan, N. A. (1990). *Sensitivity and bias in same-different and 2AFC discrimination*. Paper presented at the Eastern Psychological Association, Philadelphia, PA.

Creelman, C. D., & Macmillan, N. A. (1979). Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance, 5*(1), 146–156.

Dai, H., Versfeld, N. J., & Green, D. M. (1996). The optimum decision rules in the same-different paradigm. *Perception & Psychophysics, 58*, 1–9.

Epstein, W. (1967). *Varieties of perceptual learning*. New York: McGraw-Hill Book Company.

Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Hautus, M. J., & Irwin, R. (1995). Two models for estimating the discriminability of foods and beverages. *Journal of Sensory Studies, 10*(2), 203–215.

Huang, X., Lu, H., Tjan, B. S., Zhou, Y., & Liu, Z. (2007). Motion perceptual learning: When only task relevant information is learned. *Journal of Vision, 7*(10):14, 1–10, doi:10.1167/7.10.14. [PubMed] [Article]

Liu, Z. (1995). *Learning a visual skill that generalizes.* Princeton, NJ: NEC Research Institute.

Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. *Proceedings of the National Academy of Sciences, USA, 96*, 14085–14087.

Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research, 40*(1), 97–109.

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

Macmillan, N. A., Goldberg, R. F., & Braida, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *The Journal of the Acoustical Society of America, 84*(4), 1262–1280.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2*, 1508–1532.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Petrov, A. A. (2009). Symmetry-based methodology for decision-rule identification in *same-different* experiments. *Psychonomic Bulletin & Review, 16*, 1011–1025.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research, 51*, 1552–1566.

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2009). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience, 11*(1), 53–60.

Taylor, M. M., Forbes, S. M., & Creelman, C. D. (1983). PEST reduces bias in forced choice psychophysics. *The Journal of the Acoustical Society of America, 74*(5), 1367–1374.

Wang, X., Zhou, Y., & Liu, Z. (2013). Transfer in motion perceptual learning depends on the difficulty of the training task. *Journal of Vision, 13*(7):5, 1–9, doi:10.1167/13.7.5. [PubMed] [Article]