Bayesian Cognitive Modeling applied to Signal Detection Theory: the Mirror Effect in a Perceptual Task as a special guest

Adriana F. Chávez-De la Peña, Michael D. Lee

September, 2019

Abstract:
Signal Detection Theory (SDT) has provided one of the most well-known and broadly applied statistical models within Cognitive Sciences to study a wide range of phenomena where decision-making systems depend on the detection of specific stimuli presented within their noisy environment to produce an optimal and adaptive response. Within Recognition Memory studies where SDT is applied to compare subjects' performance between two classes of stimuli that are known to be deferentially recognized, it has consistently been found that this difference is reflected in the identification of both target and lure stimuli, as measured by hit and false alarm rates in Signal Detection Theory. The implied order of the signal and noise distributions has led to the identification of this pattern of response as "the Mirror Effect". Since this phenomena has been predominantly studied within recognition memory tasks, most attempts to account for it theorize about high-level processes engaged in the study phase. To test the generalizability of the Mirror Effect to other domains where signal detection theory has been applied, we designed a perceptual task with two levels of discriminability defined using an optical illusion. By conducting a step by step replication of the mean-based analysis reported in the literature, we present evidence of the mirror effect outside recognition memory, and then, we continue to present a more detailed model based analysis, using Bayesian methods to assess the existence of the mirror effect at both the group and individual level.

Keywords: Signal Detection Theory; Bayesian Modeling; Perception; Recognition Memory; Mirror Effect; Ebbinghaus Illusion.

1 Introduction

We live in a world full of changes, noisy information and uncertainty. Every decision-making system is constantly exposed to a huge variety of stimuli that does not always provide relevant information for their adaptation. Once that organisms learn the most essential contingency rules operating in their environment, the detection of specific stimuli (i.e. signals) among all non-informative stimuli (i.e. noise), becomes a priority task in order for them to be able to adapt to the current state of the world. We can think of a huge variety of examples to illustrate the importance of this signal-detection kind of tasks, from a vulnerable animal that has to determine whether or not the sound it has just heard in the woods represents a threat or not to decide between running and keep gathering food, to the physician that has to determine if the tomography scan she's inspecting contains evidence of a cancerous tumor or not, in order to produce a diagnostic that could have a strong effect on the patient's lifestyle.

Signal Detection Theory (SDT) has provided a very intuitive framework to understand this kind of situations and that has been widely applied within and outside Psychological Science to account for a huge variety of phenomena involving this kind of dilemmas.

1.1 Signal Detection Theory Model

Signal Detection Theory appeared for the first time in 1954 and just like many other scientific and technological developments born around that time, its purpose was to contribute to the solution of a necessity derived from the Second World War: the development and improvement of radars (Peterson, Birdsall y Fox, 1954). Not long after that, this theoretical framework was brought into the Psychological domain to study organic perception (Tanner y Swets, 1954; Swets, Tanner y Birdsall, 1961).

The SDT model works as a descriptive model that captures the problem faced by any decision-making system who finds itself in the necessity to decide whether or not a signal is present within its immediate environment, in order to produce the most appropriate response to the contingencies its presence would be announcing, (Tanner y Swets, 1954; Swets, 1973).

The core contribution of the model is that it embodies the idea that there's always a certain level of uncertainty involved in this kind of tasks. This notion is represented by using probability distributions (typically, normal distributions) to describe the variability that both signal and noise stimuli present in terms of the values with which they can be associated along a "decision axis" that contains whatever kind of information the system is sensing and processing to produce a detection judgement ("Yes, the signal is there" / "No, the signal is absent"). This variability can be explained either by assuming stimuli vary in the way they appear in the world (e.g. there's a lot of variability in terms of animals that could be considered a threat) or because there's always variability in the way decision-making systems read their environment (as observed in Psychophysics). Uncertainty comes from the notion that a portion of the noise distribution overlaps the signal distribution, allowing for the existence of a range of values on the decision axis that can be associated with either kind of stimuli. This is what makes the task complicated: decision-making systems cannot rely blindly on the evidence they perceive to emit a judgement, (Killeen, 2014; Wickens 2002).

When organisms are presented with a stimulus located at the overlapping area of the signal and noise distributions, they have to decide which binary judgment to make, by weighting the information they're currently receiving from the environment with all the information available about its structure, which can be classified in two big categories: 1) Information about the probabilistic nature of the environment and 2) Information regarding the consequences at risk, (Killeen, 2014). Let's think again about the physician who's evaluating some tomography scans: if she's not confident enough about the information being presented to diagnose the presence of a certain disease, she would have to take into account the information she has about the probability at which this disease is usually presented and the risks involved with letting a sick patient go without a treatment versus treating someone who's not sick with a very invasive procedure.

The SDT model is a decision model that assumes that binary detection judgments are produced by comparing the evidence that's currently being observed with a decision criterion which cuts the "decision axis" and which location is determined by the decision-making system. Thus, the probability of making a "Hit" or a "False alarm" depend on which portion of the signal and noise distributions fall above the criterion and the probability of making a "Correct rejection" or a "Miss" is related to how much is left below it, (See Figure 1)

One of the most valued features of the SDT model is the way in which it allows to measure separately the influence of two great factors on the decision-making system's performance: The discriminability of the signal stimuli from noise, (as represented by the distance between the mean of the signal and the noise distribution, captured by the d' parameter), and the system's bias towards a specific binary detection judgment over the other. SDT provides two distinct measures of bias: the likelihood ratio between the points at which the criterion cuts the signal and the noise distribution (the β parameter) and the distance between the observed criterion and the ideal location of the criterion assuming no bias at the point where the two distributions overlap (measured by a parameter typically identified as c).

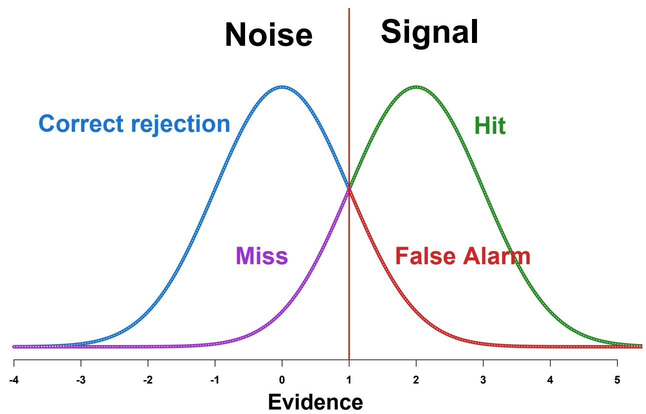


Figure 1: Graphical representation of the Signal Detection Theory model

To this day, SDT is one of the best-known models in Psychological Science. It has been applied to the study of a huge variety of phenomena, both as a tool for data analysis and as a cognitive model to describe the cognitive processes behind the emission of binary choices being made under uncertainty. One can find examples of its application to the study of perception (Rosenholtz, 2001; Pessoa, Japee y Ungerleider, 2005; Wallis y Horswill, 2007), where it was first adopted within Psychology; the emission of clinical diagnosis (Grossberg y Grant, 1978; Swets, Dawes y Monahan, 2000; Boutis, Pecaric, Seeto y Pusic, 2010); the syntomathology associated with all kinds of clinical conditions (Westermann y Lincoln, 2010; Bonnel y cols., 2002; Brown, Kosslyn, Breiter, Baer y Jenike, 1994; Naliboff y Cohen, 1981); eyewitness testimonies (Gronlund, Wixted y Mickes, 2014; Wixted y Mickes, 2014; Wixted, Miches, Dunn, Clark y Wells, 2016), and an incredibly ongroing etcetera (Gordon y Clark, 1974; Nuechterlein, 1983; Harvey Jr., Hammond, Lusk y Mross, 1992; Verghese, 2001).

The huge number of publications on the SDT is not limited to its application to different phenomena, its utility has motivated the production of several tutorials and manuals, which are aimed to provide either an introduction to the benefits of the model (Killeen, 2014); a detailed description of its assumptions and implications (McNicol, 2005; Wickens 2002); or a guide to conduct data analysis under the SDT framework, (Stainslaw y Todorov, 1999).

1.2 The mirror effect

Among the previously described variety of phenomena where SDT has been applied, its relevance to the study of Recognition Memory is definitely one of the most notorious. In a typical recognition memory task, participants are presented with a set of stimuli that they are either explicitly asked to study or with which they are expected to interact in any given way; this first phase is typically referred as the Study phase and it can involve an intentional or an incidental study of the stimuli presented. After some time has passed, participants are faced again with the same stimuli they experienced on the first phase plus another set of stimuli that hasn't been used before. The recognition memory task, as its name clearly suggest, requires subjects to identify the stimuli that they had encountered in the first phase (the signal) from those that are totally new (the noise).

When SDT has been used to analyze data obtain from Recognition Memory tasks where two different classes of stimuli (where one is known to be more easily recognized (the A class) than the other (the B class)) are presented without subjects being aware of it, and their performance is analyze separately between the A and the B class, a consistent pattern of response has been reported, showing that not only more Hits are being made on the A class, but also less False alarms. This particular phenomenon has been identified in the literature as the Mirror Effect, given the implications it has in terms of the underlying signal and noise distributions, (Glanzer y Bowles, 1976; Glanzer, Adams, Iverson y Kim, 1993). Evidence for the Mirror Effect has been reported across different procedures typically used to collect detection judgements (the "Yes/No" task; two alternative forced choice and confidence ratings) and a wide variety of variables used to define the A and B classes. (Glanzer y Adams, 1990).

The Mirror Effect has only been studied within Recognition Memory literature and thus, the majority of the models and theories proposed to account for it tend to do it in terms of high-level processes involved in the study phase, (Glanzer, Adams, Iverson y Kim, 1993; Glanzer, Kim y Adams, 1998; Glanzer, Hilford y Maloney, 2009), as if it was a reflection of the inner cognitive processes engaged in recognition memory tasks.

The data reported in the present study provides evidence for the existence of the Mirror Effect outside Recognition Memory, within a merely perceptual task where two classes of stimuli were proposed based on the literature on Optical Illusions (in particular, the Ebbinghaus illusion). This study aims to explore the generalizability of the Mirror Effect while providing a more detailed analysis of this phenomenon at the individual level and its implications in terms of the SDT model, by means of the application of Bayesian methods.

2 Experiments

To evaluate the generalizability of the Mirror Effect, we proposed a binary perceptual detection task where participants had to indicate whether two circles presented for comparison on screen had the same size (signal trials) or not (noise trials). The same task was used in two different settings, referred as Experiment 1 and Experiment 2. On Experiment 1, one of these two circles was used as the center circle of an Ebbinghaus illusion, while in Experiment 2 both circles were the central circles of their own Ebbinghaus illusion. Details about the stimuli design are discussed in a later section.

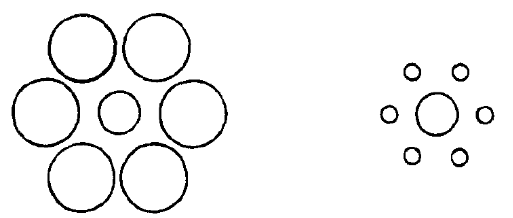


Figure 2: Illustration of the Ebbinghaus Illusion. The center circles have the same size, (illustration taken from Massaro y Anderson, 1971)

The Ebbinghaus Illusion (see Fig 2), is a well-known and widely studied optical illusion that has been typically explained as the result of a perceptual mechanism that processes and estimates sizes based on its contrast witb regards of the surrounding stimuli. This illusion tends to be illustrated with the Ebbinghaus figures (also known as the Titchener's circles), where the size of an inner circle is either overestimated or underestimated due to its relation to a set of surrounding circles which are uniformally bigger (Underestimation effect), or smaller (Overestimation effect), (Coren, 1971; Coren y Miller, 1974; De Fockert, Davidoff, Fagot, Parron y Goldstein, 2007).

We used what has been reported in the literature about the variables that have an impact on the intensity of the Ebbinghaus illusion (Massaro y Anderson, 1971; Girgus, Coren y Agdern, 1972; Roberts, HArris y Yates, 2005) to construct two levels of discriminability that would allow us to make our task analogous to those recognition tasks where the Mirror Effect has been reported. Figure 3 summarizes the results obtained in a study where the number (x-axis) and size (y-axis) of the external circles of several Ebbinghaus figures were manipulated to assess their effect on the intensity of the illusion elicited. This graph presents the mean estimations registered for the diameter of the central circle of the figures presented, and so, it can be observed that the discrepancy between the real diameter and participants' estimation increased as a function of the number of external circles included in the figures and, according to the vertical distance between every line, that the intensity of the illusion is proportional to the difference between the size of the external circles and the center circle, for either the Overestimation or the Underestimation effect, (Massaro y Anderson, 1971).

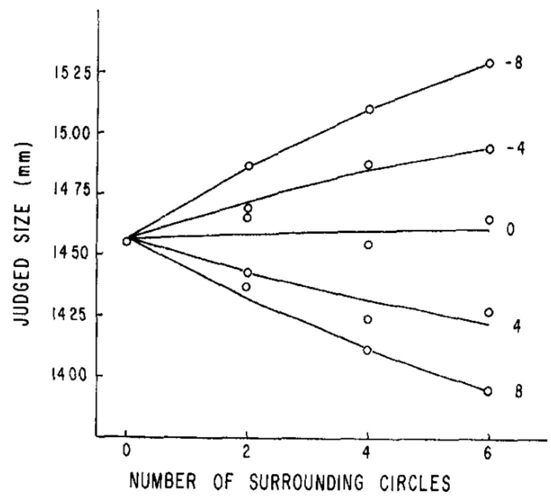


Figure 3: Results obtained in an experiment where the number and size of the external circles presented on different Ebbinghaus figures were manipulated to assess their effect on the illusion evoked, (Massaro y Anderson, 1971).

According to Massaro and Anderson's findings (1971), for our perceptual task we decided to construct the A class and the B class stimuli as follows:

- **Class A ("Higher discriminability"):** Ebbinghaus figures with few external circles (with two levels: 2 or 3 external circles)
- **Class B ("Lower discriminability"):** Ebbinghaus figures with more external circles (with two levels: 7 or 8 external circles)

For all figures constructed, external circles meant to induce an Overestimation effect were presented with a diameter of $0.5cm$ while external circles meant to induce an Underestimation effect had a constant diameter of $6cm$, these diameters were proposed so that they were at least $2x$ bigger and $0.5x$ smaller than the size of the central circles (which could vary from $1cm$ to $3cm$, with increments of $0.5cm$). The distance between the external and the central circles was not fixed nor controlled across the different figures created.

2.1 Experiment 1

On Experiment 1, participants had to compare the size of a constant stimulus (a circle with a diameter of $2cm$, which location on the left side of the screen remained fixed across trials) with the size of the central circle of an Ebbinghaus figure displayed on the right side of the screen.

Stimuli presented on Experiment 1 were constructed following a 5x2x2 factorial design, illustrated in Figure 4, for each class of stimuli A and B. The reference circle with a constant diameter of $2cm$ was compared with the center circle of an Ebbinghaus figure which could present five diameter lengths (from $1cm$ to $3cm$ in increments of $0.5cm$), with the external circles inducing either an Overestimation or Underestimation effect and two different levels of number of external circles per class (2 or 3 external circles on the A class and 7 or 8, for the B class). We ended up with 16 different "Noise stimuli" and 4 different "Signal stimuli" for each class of stimuli. To account for the variability of choices made by participants, and to make both signal and noise trials equal in terms of the number of times they were presented, we repeated each of the noise stimuli 10 times and each signal stimulus 40 times. This lead to a total of 160 noise trials and 160 signal trials per class of stimuli, with a total of 640 trials for the entire experiment.

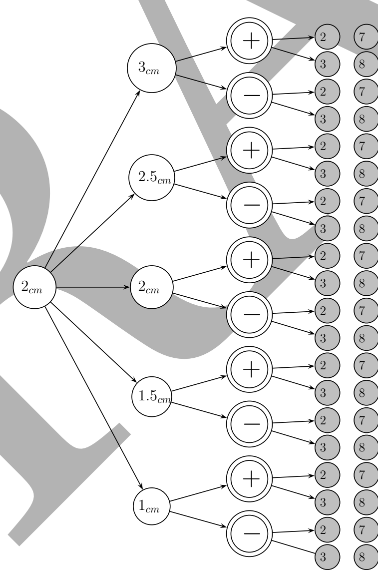


Figure 4: Stimuli design for Experiment 1: A 5x2x2 factorial design was used to construct the Ebbinghaus figures to be presented with five different diameters for the central circle, two different diameters for the external circles to produce an Overestimation or Underestimation effect and two levels of "Number of external circles" per class.

2.2 Experiment 2

On Experiment 2, the detection task remained the same "Are the two central circles the same size?", but this time the two circles that participants had to compare represented the central circle of their own Ebbinghaus figure.

Figure 5 illustrates the stimuli design followed for Experiment 2: There were 5 different possible values for the diameter of the circles being compared, with the signal stimuli being captured by the horizontal lines and the noise stimuli by the five arbitrarily drawn diagonal lines, chosen in an attempt to pair circles which diameters were as close as possible. For each of the pairs created for comparison, there were four possible combinations in terms of the number of external circles being shown. For example, for stimuli of the A class, it could be the case that both figures had 2 or 3 external circles, or that this number could vary between the left and right figures. It is important to note that in order to make the comparison of the central circles challenging, one of the figures was always built to induce an Overestimation effect, leaving the Underestimation effect for the remaining. Thus, 20 different signal and noise stimuli were designed for each class of stimuli. Each of the 80 different pairs for comparison designed for Experiment 2 were presented 8 times, alternating the position of the Overestimation and Underestimation effect inducing figures, leading to a total of 640 trials.

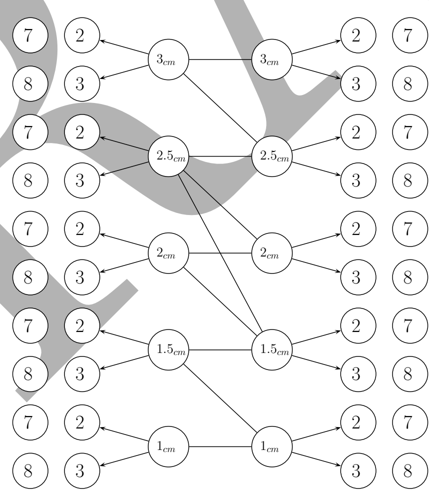


Figure 5: Stimuli design for Experiment 2.

2.3 Materials

The experiment was programmed and executed in **PsychoPy v.12** (Peirce, 2007), using a Mac computer ($59.5cm$ x $34cm$ screen). Experimental sessions were ran inside an isolated room, with participants being sat in a fixed chair $1.10m$ away from the screen.

2.4 Participants

A total of forty-one students (ages 18 to 21) of the Psychology School of the National Autonomous University of Mexico participated in one of the two experiments conducted (20 in Experiment 1 and 21 in Experiment 2). Both experiments were conducted during the same period of time, with participants being assigned to one of these, without them knowing that there was more than one Experiment.

2.5 Procedure

The detection task was conducted in two steps. First, participants had to respond to a binary choice task and then rate their confidence on each response registered within a Confidence scale from 1 to 3, presented on screen to remind participants that 1 meant "Very insecure" and 3 "Very confident". However, the code for the experiment was designed to translate these responses into a larger scale from 1 to 6 depending on their binary response. Values 1 and 6 referred a greater confidence in their previous response (with 1 being "Very sure it was a Noise trial" and 6 being "Very sure it was a Signal trial"), while values in the middle could serve as indicators of greater uncertainty. This translation procedure has been used to lower the risk for fatigue and to guarantee that binary choices registered could always match the values assigned in the confidence scale, (Glanzer y Adams, 1990).

The experiment consisted of 640 trials, with each one of these following the next structure:

- **First part: Binary choice Yes/No**

  At the beginning of each trial, the pair of central circles that participants had to compare appeared on screen along with the reminding legends "Do the central circles have the same size?" on top of the screen and the answer keys on the bottom "S = Yes, N = No". After 1.5 seconds, the stimuli disappeared from the screen to prevent habituation, while the reminding legends remained until a response was registered.

- **Second part: Confidence Scale**

  Once participants had registered their binary response to the Yes/No task, the Confidence scale appeared on screen below a the prompt "How certain are you about your response?". The scale shown contained numbers 1, 2 and 3 along with the legends "Very insecure", "More or less sure" and "Very confident".

- **Third part: "Space-bar to continue"**

  Right after a response to the Confidence scale was registered, a third and final screen was shown to indicate the end of the trial, with the legend "Press the space-bar to move on to the next trial". This final screen was included to give participants the chance to take a break before continuing to be exposed to more Ebbinghaus figures, preventing their fatigue.

At the end of the 640 trials, participants were shown the total number of right and wrong responses registered just to provide them with a sense of closure. We never told participants the purpose of the experiment and more importantly, they were never told anything about the two distinct classes of stimuli coexisting within the same task.

3 Results

More than three fourths of participants' response patterns corresponded with what has been reported as part of the mirror effect. In Experiment 1, 17 out of 20 participants presented evidence for the mirror effect in the binary choice task and 18 in the confidence scale. In Experiment 2, 18 participants out of 20 showed evidence for the mirror effect in both tasks. These proportions shown to be statistically significant according to the classical binomial test.

A quick inspection of the data collected by means of data visualization allowed us to detect an abnormality in terms of the pattern of response registered by Participant No.1 in Experiment 2 (Figure 6), who spent eighty trials responding "No", regardless of the type of stimuli (signal or noise) and class (A or B). This participant was excluded from any further data analysis.

To allow for the comparison between our results and those reported in the recognition memory literature, we conducted a step by step replication of the data analysis presented by Glanzer and Adams, (1990):

1. **Comparing d' across classes of stimuli.**

   A t-test was conducted to compare the means of the classical d' estimations obtained for each class of stimuli. Statistically significant differences were found, validating our experimental design: A class signal stimuli were easier to recognize in both Experiments.

2. **Comparing the Hit and False Alarm rates across classes of stimuli.**

   Two separate t-tests were conducted to compare the Hit rates and the False alarm rates registered for each class of stimuli. Statistically signif-

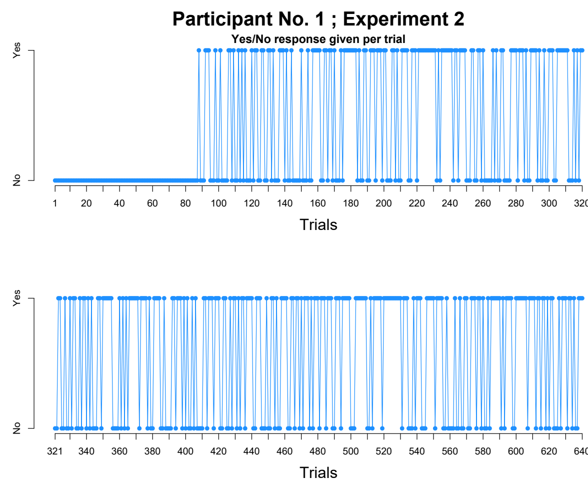


Figure 6: Yes/No responses registered on each of the 640 trials by participant No.1 in Experiment 2. A clear pattern can be observed during the first 80 trials, which suggest that participant's responses were not made according to the instructions or the stimulus presented.

icant differences were found in the same direction as in the recognition memory literature.

3. **Comparing the mean confidence ratings registered across classes of stimuli.**

   Two separate t-tests were conducted to compare separately the mean confidence rates registered for the Signal and Noise trials in each class of stimuli. Statistically significant differences were found in the same direction as reported in the recognition memory literature.

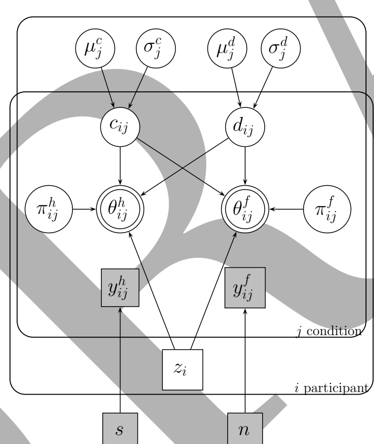
4 Bayesian Cognitive Modeling

We conducted a more detailed analysis by means of the application of Bayesian methods to explore the existence and implications of the Mirror Effect at the individual level, as well as to test the conditions that can be inferred from the data obtained in terms of the compliance of the assumptions typically associated with the SDT framework (such as the relation between the variances of the Signal and Noise distributions and the location of the criteria used to respond to either the binary choice task and the confidence rating scale).

4.1 Latent Mixture Models for Contaminant data

One of the many advantages derived from the application of Bayesian modeling techniques is their ability to test the usually accepted assumption that all data obtained from an experimental task can be explained as the product of high-level mechanisms described by certain statistical models, with the application of latent mixture models to identify which data points are consistent with the model amd distinguish them from contaminant data, (Chávez, M, Villalobos, E., Baroja, J.L & Bouzas, A., 2017; Velázquez, C., Villarreal M., & Bouzas, A., 2019).

In the present study, we developed two latent mixture models to detect contaminant data at two diferent levels. The first of these models worked at the level of the overall responses registered during the whole experiment, while the second took into account every single response registered by each participant. This way, one model allowed us to test if the observed Hits and False alarms rates can be explained by the SDT framework, while the other model tested independence among trials, since stimuli were presented at random (regardless of their class and nature).



$$\mu_j^c \sim \text{Gaussian}(0, 0.7)$$
$$\mu_j^d \sim \text{Gaussian}(0, 1)_{T(0,\infty)}$$
$$\sigma_j^c, \sigma_j^d \sim \text{Uniform}(0, 4)$$
$$d_{ij} \sim \text{Gaussian}(\mu_j^d, \sigma_j^d)$$
$$c_{ij} \sim \text{Gaussian}(\mu_j^c, \sigma_j^c)$$
$$\pi_{ij}^h, \pi_i^f \sim \text{Beta}(1, 1)$$
$$\theta_{ij}^h \leftarrow \begin{cases} \pi_{ij}^h & \text{if } z_i = 0 \\ \phi(\frac{1}{2}d_{ij} - c_{ij}) & \text{if } z_i = 1 \end{cases}$$
$$\theta_{ij}^f \leftarrow \begin{cases} \pi_{ij}^f & \text{if } z_j = 0 \\ \phi(-\frac{1}{2}d_{ij} - c_{ij}) & \text{if } z_j = 1 \end{cases}$$
$$z_i \sim \text{Bernoulli}(0.5)$$
$$y_{ij}^h \sim \text{Binomial}(\theta_{ij}^h, s)$$
$$h_{ij}^f \sim \text{Binomial}(\theta_{ij}^f, s)$$

Figure 7: Latent mixture model designed to detect contaminant data in terms of the overall patterns of Hits and False alarms registered by every participant.

The graphical model for the first latent mixture model is shown in Figure 7, (for an introduction to graphical models, we strongly recommend to read Lee and Wagenmakers, 2013; Lee, 2016). In this model, the total number of Hits (y_{ij}^h) and False alarms (y_{ij}^f) registered for every participant i in the class of stimuli j, are interpreted as the number of succeses registered out of a total number of Signal (s) and Noise (n) trials, according to a binomial process with probabilities (θ^h_{ij}) and (θ^f_{ij}). This underlying probability could arise from two different sources, depending on an individual dicotomic parameter z_j; when $z_i = 1$ these probabilities are drawn from the interaction of the d_{ij} and c_{ij} parameters which correspond to the discriminability and bias measures as described by the SDT framework and which are assumed to be drawn from two distinct class-level normal distributions; when $z_i = 0$ these probabilities come from random parameters π^h_{ij} and π^f_{ij}, with an uniform prior.

Figure 8 shows the values estimated for the z_i parameter on Experiment 1 (left panel) and 2 (right panel) across participants.

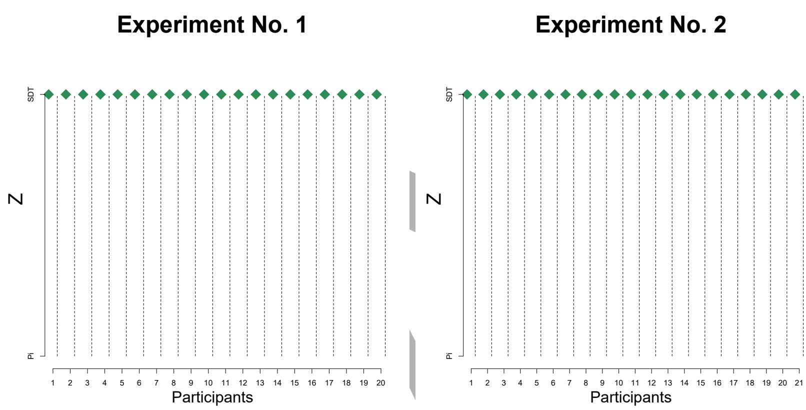


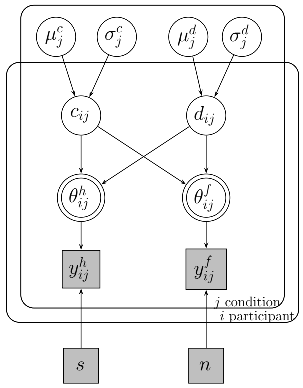
Figure 8: Individual estimations for the dicotomic latent parameter z_i controlling whether the probabilities of making a Hit and a False Alarm could be described by the SDT framework, or as a random probability.

The second Contaminant model is still in progress

4.2 Hierarchical SDT Model

We applied a Hierarchical extension of the SDT model originally developed by Lee and Wagenmakers (2013), where the individual parameters for discrimination and bias were assumed to be drawn from their own group-level normal distributions for each class of stimuli (d_{ij} and c_{ij}, respectively), as shown in the graphical model presented in Figure 9. This model constitutes the base for the great majority of the forthcoming Bayesian models, in terms of both the representation of the underlying processes assumed by SDT and the specific priors used to reflect the assumptions being made for our experimental task.

Just like before, in this hierarchical SDT model the total number of Hits



$$\mu_j^d \sim \text{Gaussian}(0,1)_{T(0,\infty)}$$
$$\mu_j^c \sim \text{Gaussian}(0,0.7)$$
$$\sigma_j^c, \sigma_j^d \sim \text{Uniform}(0,4)$$

$$d_{ij} \sim \text{Gaussian}(\mu_j^d, \sigma_j^d)$$
$$c_{ij} \sim \text{Gaussian}(\mu_j^d, \sigma_j^c)$$

$$\theta_{ij}^h = \phi(\tfrac{1}{2}d_{ij} - c_{ij})$$
$$\theta_{ij}^f = \phi(-\tfrac{1}{2}d_{ij} - c_{ij})$$

$$y_{ij}^h \sim \text{Binomial}(\theta_{ij}^h, s)$$
$$y_{ij}^f \sim \text{Binomial}(\theta_{ij}^f, n)$$

Figure 9: Hierarchical Bayesian extension of the SDT model, in which both d_i and c_i values are assumed to be drawn from their own group-level normal distributions, defined per class.

(y_{ij}^h) and False Alarms (y_{ij}^f) registered for every participant i per class of stimuli j out of a total number of signal (s) and noise (n) trials are described as the result of a binomial process regulated by probabilities θ_{ij}^h and θ_{ij}^h. The SDT framework is portrayed by the way in which d_{ij} (the parameter measuring discriminability) and c_{ij} (the parameter quantifying subjects' bias) interact to produce these probabilities, since both θ_{ij}^h and θ_{ij}^h are assumed ot be equal to the cumulative density function of their respective signal or noise normal distribution above the criterion, (Lee Wagenmakers, 2013; Gescheider, 2013).

Figure 10 presents the joint and the marginal posterior estimations for the mean of the discriminability and bias parameters on each class of stimuli (A in blue and B in purple), on Experiment 1 (left panel) and Experiment 2 (right panel). According to the hierarchical SDT model, the posterior estimations for the mean of the group-level distributions from which the individual values of d_{ij} in both experiments, are assumed to be drawn from are appear to be fairly separated. This finding confirms the effectiveness of our experimental design while constructing our A and B classes of stimuli.

It is interesting to note that the means of the group-level distributions from which the individual values of c_{ij} are assumed to be drawn from in Experiment 1 seem to be the same across classes of stimuli. This outcome is consistent with the general assumption made within the mirror effect literature about participants not knowing about the coexistence of more than one class of stimuli, being thought to use a single criterion to produce all of their

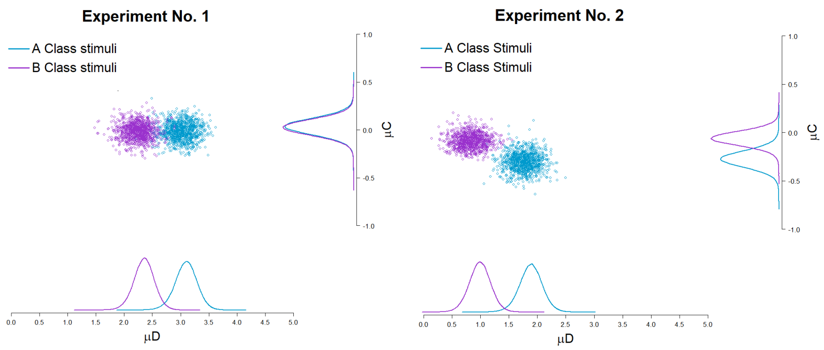
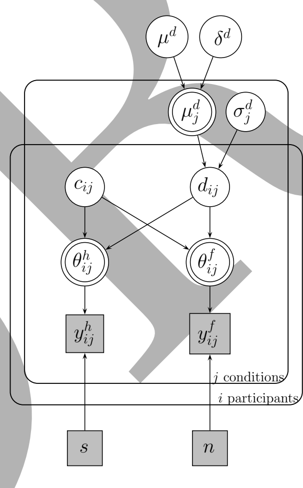


Figure 10: Joint and marginal posteriors for the means of the group-level normal distributions from which individual values of discriminability (d) and bias (c) are assumed to be drawn.

responses. However, this relationship doesn't hold in Experiment 2, where the means of the group-level distributions for the c_{ij} parameters are different. The implications of these findings and their relation with the mirror effect and its implications are further discussed.

4.3 Comparing d' across classes of stimuli



$$\mu^d \sim \text{Gaussian}(0,1)_{T(0,\infty)}$$
$$\delta^d \sim \text{Gaussian}(0,1)_{T(0,\infty)}$$

$$\mu^d_A = \mu^d + \frac{\delta}{2}$$
$$\mu^d_B = \mu^d - \frac{\delta}{2}$$
$$\sigma^d_j \sim \text{Uniform}(0,5)$$

$$d_{ij} \sim \text{Gaussian}(\mu^d_j, \sigma^d_j)$$
$$c_{ij} \sim \text{Gaussian}(0, 0.7)$$

$$\theta^h_{ij} = \phi(\tfrac{1}{2}d_{ij} - c_{ij})$$
$$\theta^f_{ij} = \phi(-\tfrac{1}{2}d_{ij} - c_{ij})$$

$$y^h_{ij} \sim \text{Binomial}(\theta^h_{ij}, s)$$
$$y^f_{ij} \sim \text{Binomial}(\theta^f_{ij}, n)$$

Figure 11: An extension of the hierarchical SDT model to test the differences between the mean of the group-distributions for the d parameter.

We developed a variation of the hierarchical SDT model to test formally the differences found between the mean of the group-level distributions from which the individual values for the discriminability parameter are assumed to be drawn. In this model, shown in Figure 11, we added a μ^d parameter to represent the idea of an underlying class-independent mean value for the d group-level normal distribution and a δ^d parameter to represent the distance between the estimations for the mean value of each the A class and the B class d distribution.

Figure 12 shows the posterior densities for the individual d estimations conducted for all participants in Experiment 1 (left panel) and 2 (right panel), for each class of stimuli. From these graphs it can be seen that estimates made for the A class stimuli (in blue) are consistently higher than those made for the B class (purple), both in terms of the maximum density point (marked with a black dot) as in the overall length of each posterior distributions.

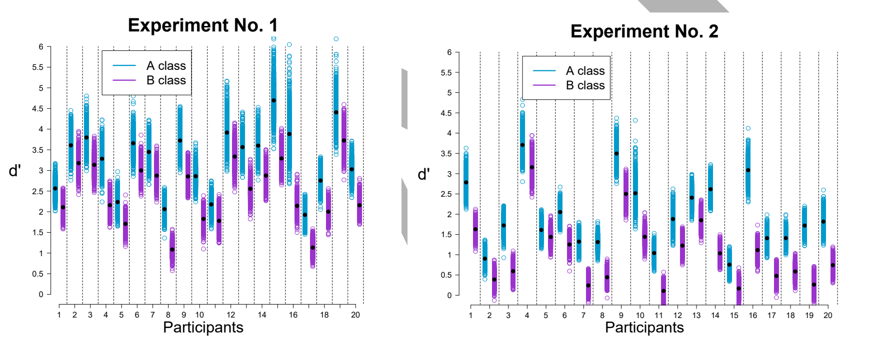


Figure 12: Posterior densities estimated for each participant on Experiment 1 (left) and 2 (right) in terms of the values of their individual d_{ij} estimates under each class of stimuli, as seen from above. The marked dot signals the point of maximum density within each posterior distribution.

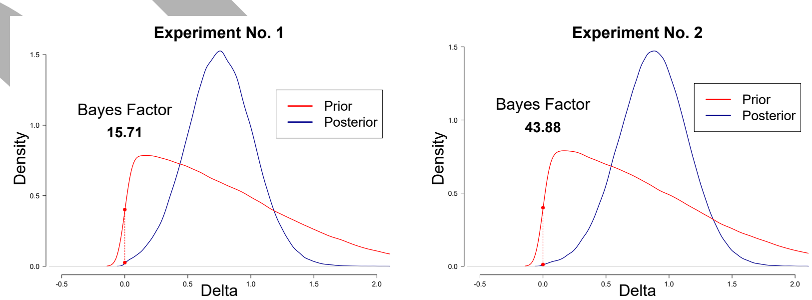
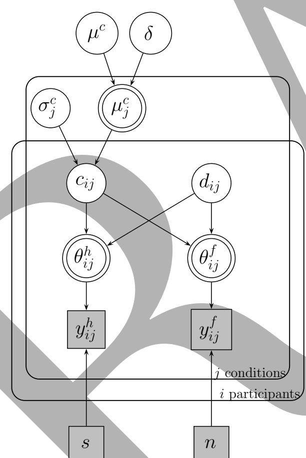


Figure 13: Bayes factor between the prior and posterior distributions of the δ^d parameter representing the difference between the means of the d group-level distributions.

Figure 13 illustrates the results obtained after computing the Bayes factor between the prior and posterior distributions for the δ^d parameter at the 0 value as the "no difference point", for Experiments 1 (left) and 2 (right). It is interesting to note that the Bayes factor computed for Experiment 2 is much larger than that in Experiment 1, indicating a much greater difference in terms of how easy it was for participants to detect trials where the circles for comparison had the same size for the A and B class.

4.4 Comparing c across classes of stimuli

We developed a similar extension of the plain hierarchical SDT model to test for the differences in the mean of the group-level c distributions per class of stimuli, (Figure 14). Similarly, Figure 15 shows the individual posterior densities for the c parameter obtained for all participants across classes of stimuli, on Experiment 1 (left) and 2 (right). Figure 16 presents the results of the Bayes factor computed between the prior and the posterior distributions for the δ^c parameter at 0.



$$\mu^c \sim \text{Gaussian}(0, 0.7)$$
$$\delta \sim \text{Gaussian}(0, 1)$$
$$\mu^c_A \leftarrow \mu^c + \tfrac{\delta}{2}$$
$$\mu^c_B \leftarrow \mu^c - \tfrac{\delta}{2}$$
$$\sigma^c_j \sim \text{Uniform}(0, 5)$$
$$c_{ij} \sim \text{Gaussian}(\mu^c_j, \sigma^c_j)$$
$$d_{ij} \sim \text{Gaussian}(0, 1)_{T(0,\infty)}$$
$$\theta^h_{ij} \leftarrow \phi(\tfrac{1}{2}d_{ij} - c_{ij})$$
$$\theta^f_{ij} \leftarrow \phi(-\tfrac{1}{2}d_{ij} - c_{ij})$$
$$y^h_{ij} \sim \text{Binomial}(\theta^h_{ij}, s)$$
$$y^f_{ij} \sim \text{Binomial}(\theta^f_{ij}, n)$$

Figure 14: An extension of the hierarchical SDT model to test the differences between the mean of the group-distributions for the c parameter.

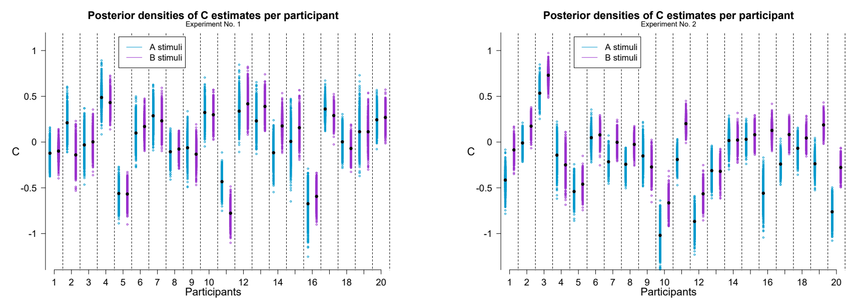


Figure 15: Posterior densities estimated for each participant on Experiment 1 (left) and 2 (right) in terms of the values of their individual c_{ij} estimates under each class of stimuli, as seen from above. The marked dot signals the point of maximum density within each posterior distribution.

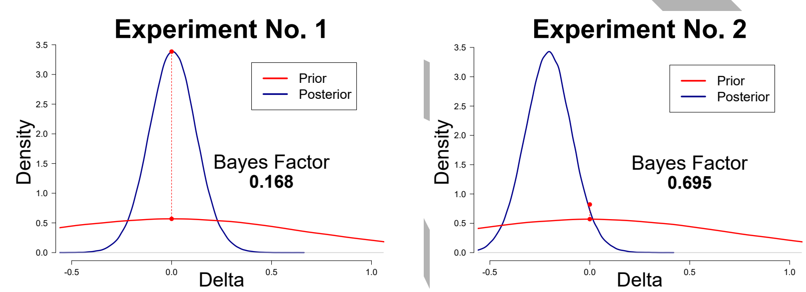


Figure 16: Bayes factor between the prior and posterior distributions of the δ^c parameter representing the difference between the means of the c group-level distributions.

4.5 Comparing the Hit and False Alarm rates as reported in the Mirror Effect

5 Discussion

6 Conclusion

7 Acknowledgements

References

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallum, E., & Bonnel, A.-M. (2002). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226-235.

Boutis, K., Pecaric, M., Seeto, B., & Pusic, M. (2010). Using signal detection theory to model changes in serial learning of radiological image interpretation. *Adv in Health Sci Educ*, *15*, 647-658.

Brown, H. D., Kosslyn, S. M., Breiter, H. C., Baer, L., & Jenike, M. A. (1994). Can patients with obsessive-compulsive disorder discriminate between percepts and mental images? a signal detection analysis. *Journal of Abnormal Psychology*, *103*(3), 445-454.

Carlos, V., Manuel, V., & Arturo, B. (2019). Velocity estimation in reinforcement learning. *Computational Brain Behavior*(2), 95-108.

Chávez, M., Villalobos, E., Baroja, J., & Bouzas, A. (2017). Hierarchical bayesian modeling of intertemporal choice. *Judgment and Decision Making*, *12*(1), 19-28.

Coren, S. (1971). A size-contrast illusion without physical size differences. *American Journal of Psychology*, *84*(4), 565-566.

Coren, S., & Miller, J. (1974). Size contrast as a function of figural similarity. *Perception Psychophysics*, *16*(2), 355-357.

de Fockert, J., Davidoff, J., Fagot, J., PArron, C., & Goldstein, J. (2007). More accurate size contrast judgments in the ebbinghaus illusion in a remote culture. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 738-742.

Gescheider, G. A. (2013). *Psychophysics: The fundamentals* (3rd Edition ed.; . Psychology Press, Ed.).

Girgus, J. S., Coren, S., & Agdern, M. (1972). The interrelationship between the ebbinghaus and delboeuf illusions. *Journal of Experimental Psychology*, *95*(2), 453-455.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology*, *16*(1), 5-16.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*(3), 546-567.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of the Experimental Analysis of Behavior: Human Learning and Memory*, *12*(1), 21-31.

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review*, *16*(3), 431-455.

Glanzer, M., Kim, K., & Adams, J. K. (1998). Response distribution as an explanation of the mirror effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*(3), 633-644.

Gordon, S. K., & Clark, W. C. (1974). Application of signal detection theory to prose recall and recogintion in elderly and young adults. *Journal of Gerontology*, *29*(1), 64-72.

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Association for Psychological Science*, *23*(1), 3-10.

Grossberg, J. M., & Grant, B. F. (1978). Clinical psychophysics: Applications of ratio scaling and signal detection methods to research on pain, fear, drugs and medical decision making. *Psychological Bulletin*, *85*(5), 1154-1176.

Jr., L. O. H., Hammond, K. R., Lusk, C. M., & Mross, E. F. (1992). The application of signal detection theory to weather forecasting behavior. *American Metereological Society*, *103*, 863-883.

Killeen, P. R. (2014). Signal detection theory. *Enclyclopedia of Theory in Psychology*.

Lee, M. D. (2016). *Bayesian methods in cognitive modeling.* (unpublished chapter)

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course* (1st Edition ed.; C. U. Press, Ed.).

Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the ebbinghaus illusion. *Journal of Experimental Psychology*, *89*(1), 147-151.

McNicol, D. (2005). What are statistical decisions? In *A primer of signal detection theory* (p. 1-17). Oxford: Oxford University Press.

Naliboff, B. D., & Cohen, M. J. (1981). Signal detection and threshold measures for chronic back pain patients, chronic illness patients, and cohort controls to radiant heat stimuli. *Journal of Abnormal Psychology*, *90*(3), 271-274.

Nuechterlein, K. H. (1983). Signal detection in vigilance tasks and behavioral attributes among offspring of schizophrenic mothers and among hyperactive children. *Journal of Abnormal Psychology*, *92*(1), 4-28.

Peirce, J. W. (2007). Psychopy-psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13.

Pessoa, L., Japee, S., & Ungerleider, L. G. (2005). Visual awareness and the detection of ferful faces. *Emotion*, *5*(2), 243-247.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *IRE Trans. on Information Theory*, *12*(4), 171-212.

Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the ebbinghaus illusion (titchener circles). *Perception*, *34*, 847-856.

Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for sigmal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 985-999.

Stainslaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments and Computers*, *31*(1), 137-149.

Swets, J. A. (1973, 12). The relative operating characterstic in psychology. *Science*, *182*, 990-1000.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological science in the public interest*, *1*(1), 1-26.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*(5), 301-340.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*(6), 401–409.
Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*(3), 523-535.
Wallis, T. S. A., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis and Prevention*, *39*, 1177-1185.
Westermann, S., & Lincoln, T. M. (2010). Using signal detection theory to test the impact of negative emotion on sub-clinical paranoia. *Journal of Behavior Therapy and Experimental Psychiatry*, 96-101.
Wickens, T. D. (2002). The signal-detection model. In *Elementary signal detection theory* (p. 3-16). New York: Oxford University Press.
Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262-276.
Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(2), 304-309.