Dilution Effects in Perceptual Information Integration

Jared M. Hotaling[a] (jhotalin@indiana.edu), **Andrew L. Cohen**[b] (acohen@psych.umass.edu), **Jerome R. Busemeyer**[a] (jbusemey@indiana.edu), **& Richard M. Shiffrin**[a] (shiffrin@indiana.edu)

[a]Department of Psychological & Brain Sciences, Indiana University
1101 E. Tenth St., Bloomington, IN 47408 USA
[b]Department of Psychology, University of Massachusetts
Amherst, MA 01003 USA

Abstract

In cognitive science there is a paradox: Researchers studying decision making have repeatedly shown that people employ simple and often less than optimal strategies when integrating information from multiple sources. However, researchers working in fields such as categorization, memory, and perception have had great success using optimal models to account for information integration. Is this conflict due to the use of different materials and procedures? We test the hypothesis that stimuli requiring more controlled information integration lead to suboptimal performance, while stimuli that lend themselves to more automatic processing produce more optimal integration. We test for one canonical example of sub-optimal information integration, the *dilution effect*, using stimuli more commonly found in perception experiments. Dilution was indeed reliable across several conditions. The largest effects occurred in stimuli manipulated so as to discourage automatic processing. We use the Multi-component Information Accumulation model to explain how stimulus presentation influenced cognitive processing.

Keywords: dilution effect; information integration; models

Introduction

Information integration is the combining of evidence from multiple sources. Many tasks, from speech comprehension to medical decision making, require such integration. Each source of information on its own provides some evidence, but integrating all information yields the best performance. This article investigates the manner in which information is combined. While the literature provides numerous examples of near-optimal information integration, there are just as many examples where it is far from optimal. The types of stimuli and procedures used often determined the pattern of results. Tasks involving quantitative stimuli, like probability judgments (Tversky & Kahneman, 1974), seem to implicate heuristic strategies, more than perceptual tasks, like speech comprehension (Oden & Massaro, 1978). Our research tests the hypothesis that even perceptual information can produce suboptimal integration if it is displayed in a way that discourages automatic processing.

Suboptimal Integration in Decision Making

Many studies of judgment and decision making suggest that information from multiple sources is integrated via simple heuristics. Sometimes these studies produce behavior approaching optimal decision making (Gigerenzer & Todd, 1999), but in many other cases, performance is well short of optimal (Gilovich, Griffin, & Kahneman, 2002). The conjunctive fallacy, unpacking effects, and the dilution effect are just a few of the many common findings that violate normative models of information integration. These deviations from rational behavior are so numerous that it is now common to assume sub-optimal integration as a starting point for theories of decision making.

Optimal Integration in Perceptual Domains

In contrast, there are numerous successful applications of optimal or rational models of information integration in domains such as perception (Oden & Massaro, 1978; Tenenbaum, 1999), categorization (Ashby & Maddox, 1990, 1992; Nosofsky, 1986), and memory (Anderson, 1991; Shiffrin & Steyvers, 1997). Researchers in these more perceptual fields begin with an assumption of optimal integration and only later investigate sub-optimal or heuristic-based strategies.

The Dilution Effect

Although information integration is an object of study by researchers in both decision making and perceptual domains, these fields often seem to operate independently of each other. One reason is the difference in experimental paradigms. Decision making research focuses mainly on linguistic and quantitative stimuli, and is concerned with how individuals use information to form explicit inference or preferences. Perceptual research typically relies on more perceptual stimuli, and concentrates on how the information is produced from external stimulation. Even when words are used as stimuli, as in memory research, the focus of information integration often includes perceptual aspects of the stimuli. The present research aims to bridge this divide through a novel experimental paradigm that combines aspects of each research tradition.

We focus on one example of sub-optimal information integration: *the dilution effect*. This effect refers to a situation where adding null or weak positive evidence to what is already strong positive evidence reduces the overall belief in a hypothesis. The effect has been replicated in numerous studies (LaBella & Koehler, 2004; McKenzie, Lee, & Chen, 2002; Nisbett, Zukier, & Lemley, 1981; Peters & Rothbart, 2000), but Shanteau (1975) gives one of the clearest demonstrations of the dilution effect. In his study, an experimenter drew samples of red (R) and white (W) beads, with replacement, from one of two boxes. Box A was 70% W and 30% R. Box B was 30% W beads and 70% R. The participants did not know from which box the beads were drawn. In one condition, the experimenter drew the sequence WWWRWR from one of the boxes. After every two beads, participants estimated the probability that the beads came from Box A. The mean judgments after WW, WWWR, and WWWRWR were 69.3%, 64.0, and 60.6, respectively. The WW sample provides diagnostic information, information that clearly points to Box A. However the subsequent samples were nondiagnostic; they could have come from either box with equal probability, and should not have changed the estimated likelihood that the entire sequence came from Box A. Yet this non-diagnostic information caused the estimated probability to drop.

Why Faces?

Although the dilution effect has only been explored using traditional judgment and decision making stimuli, it easily lends itself to perceptual stimuli. We use weak and strong evidence from different parts of a face to investigate the effect. For example, imagine you are asked to identify a face captured on a security camera. The top half of the face is relatively clear, but the bottom half is in shadow and harder to see. The top and bottom halves of the face then lend strong and weak evidence to the decision. The primary goal of this research is to determine whether the information from these sources is combined in an optimal fashion, or sub-optimally as exemplified by the dilution effect.

A benefit of using perceptual stimuli is that issues of interpretation and language understanding do not come into play. For example, the conjunction law is violated less if participants interpret “Linda is a bank teller” to mean that she is a bank teller and *not* a feminist (Sides, Osherson, Bonini, & Viale, 2002). People also often misinterpret probabilities, but perform more optimally when information is presented as frequencies (Gigerenzer & Hoffrage, 1995). The present task employs perceptual stimuli, thereby greatly reducing any undesirable influence of language conventions.

Testing Models of the Dilution Effect

In addition to testing for the dilution effect, we evaluate three models of information integration. The *Simple Bayesian* model combines evidence from the two sources of information optimally, according to Bayesian statistical methods, and predicts additive effects. The *Averaging* model calculates a weighted arithmetic mean of the evidence produced by each source, and always predicts dilution. Finally, we use the *Multi-component Information Accumulation* model to explain how information is sampled from multiple sources, and accumulates during deliberation. This model accounts for the behavior we observed in our experiment, and provides insight into how stimulus presentation affects information processing.

The goal of our experiment was to replicate the dilution effect using perceptual stimuli and to determine the role of stimulus presentation on performance. In particular, we tested if images that encouraged more automatic perceptual integration yield reduced dilution effects than images that required more controlled combination of evidence.

In the experiment, participants categorized a test series of faces into two families (Jones or Smith). The test faces were created by morphing together two target faces (representing the patriarch of each family) along a continuum. Different parts of the faces were morphed independently, allowing us to test how individual combined various levels of evidence. In direct analogy to standard work on the dilution effect, the top and bottom halves of a face act as two sources of information. Based on the many studies showing near-optimal combination of perceptual information it would be natural to expect two halves from the Jones side of the morph continuum to produce even stronger responses in favor of Jones. Alternatively, weak evidence might dilute strong evidence to produce a dilution effect.

To investigate factors controlling the size and reliability of the dilution effect, two manipulations differentially encouraged automatic and controlled integration of information. It is fairly common to distinguish automatic and controlled processing, both in theory and empirical research. Most often automatic processing is assumed to be fast and independent of conscious manipulation, and controlled processing is assumed to be slow and conscious. Automatic processing is usually assumed to be more robust, less prone to large errors, less based on heuristics, and closer to optimal than controlled processing. This line of thinking suggests that the dilution effect is less likely when processing is automatic, and more likely when processing is controlled. We use the automatic/controlled language of Schneider & Shiffrin (1977) for convenience sake, rather than to make strong claims that information integration is ever entirely automatic or controlled.

In the present experiment we used conditions that manipulated face images so as to bias processing toward or away from automatic processing. In the Together condition the two half faces are shown atop one another, in a normal configuration. Because identification of faces is over-learned, this should promote automatic processing and produce less dilution. That is, weak evidence, when added to strong evidence from the same category, should increase accuracy. In the Split condition the two half faces were separated horizontally. In the Inverted condition the images were displayed upside-down. Because our perceptual systems have rarely needed to recognize split or inverted faces, each half face might be processed separately, with the results later combined using more deliberate strategies. That is, weak evidence should combine less optimally with strong evidence and produce more dilution.

Method

Participants

Nineteen students from Indiana University (undergraduate and graduate) were paid $16 to participate in this study. All participants had normal or corrected-to-normal vision.

Stimuli

All of the stimuli used in the experiment were derived from two "target faces" (A and B) selected from the FERET database (Philips, Moon, Rizvi, & Rauss, 2000). After cropping the image to remove hair and head outline, the faces were warped so that their major facial features aligned. Once the faces are aligned, a morph is essentially a linear combination of the grayscale values of the two faces at each pixel. The cropped areas of the 256 × 384 pixels, grayscale images were filled with a sinusoidal grating. Upside-down copies of the two target faces were also made for the Inverted conditions. The four resulting images were used to construct all experimental stimuli.

The experiment began with two short blocks of trials that calibrated morphs levels to the individual. On each trial a half face was presented and participants chose the target that it most closely resembled. The test faces were created by morphing Target A and Target B together along a continuum. Faces favoring A and B were initialized to 94.44% Target A and 5.56% Target A, respectively. A staircase algorithm was used to find top and bottom half face morphs for each target and each orientation that produced an intermediate level of accuracy (approximately 72%). These morphs became the medium (M) strength half faces, while weak (W) and strong (S) morphs were derived by extrapolation. Weak halves use the morph coefficient halfway between the medium morph and 0.5. Strong halves used the morph two thirds of the distance between the medium morph and the target.

Having calibrated all morphs levels, test stimuli were created as follows. For each orientation, the W, M, and S top half faces for Target A were crossed with the W, M, and S bottom half faces for Target A. The same procedure was followed for Target B. As a manipulation check, the W and M half faces for A were also paired with the M and W half faces for B, respectively. Whole faces were presented either in a normal configuration (directly above or below the other half face or background) or horizontally split by 60 pixels. The W, M, and S top and bottom half faces were also presented in isolation with a continuation of the background presented instead of the other half of the face. Pilot testing showed no performance differences between Together and Split half faces, so the latter were omitted. Sample stimuli are shown in Figure 1. This procedure was done separately for upright and inverted faces, yielding 56 test stimuli for each orientation.

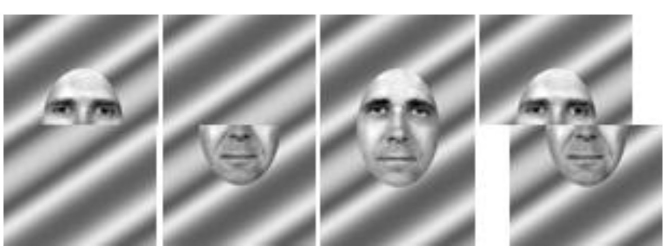

Figure 1: Example test faces.

Procedure

Participants completed two sessions of the experiment on separate days. They were told that they would see a series of faces, each of which belonged to either the Jones or Smith family. They were instructed to use the test face's resemblance to each patriarch to determine the correct family. After several example trials, participants completed two blocks of calibration trials. The first consisted of 72 upright half face trials, interspersed with 48 upright whole face filler trials included to discourage strategies tailored to half faces. Auditory feedback was given after each response, with a high beep for correct and a low beep for incorrect.

After calibration, participants began an integration phase consisting of two blocks of trials in Session 1 and six blocks in Session 2. Each block contained 68 trials. Each test face appeared once per block, with the exceptions of W/W, M/M, and S/S stimuli, which appeared twice. Upright faces appeared in odd numbered blocks. Inverted faces appeared in even numbered blocks.

Each trial began with a test face appearing in one of nine random positions near the middle of the screen. After two seconds the face was masked with one of two scrambled sets of features from the target faces. After 250ms the mask disappeared and the two target faces appeared, one on each side of the screen. Participants chose the family to which the test face belonged. They were then asked, "What is the likelihood that you are correct?", and responded on a 6-point scale from 50% to 100%. A fixed number of points were awarded for each correct choice and the individual with the highest final score received a $20 bonus.

Results

The present analysis focuses on participants' choice proportions, though mean confidence judgments showed a similar pattern of results. We began by removing data from trials in which individuals indicated no confidence in their decision (likelihood judgment of 50%), or responded too fast (less than 150 ms), or too slow (greater than 5 sec). This procedure removed approximately 12% trials, across all participants.

Next, we labeled morphs according to the accuracy they produced on half face trials. That is for example, an individual's half face trials determined which Jones top half morphs were strong, medium, and weak. This relabeling proved unnecessary in most cases because accuracy order matched the physical morph order.

A choice response was considered correct if the test face provided stronger evidence for that target than the alternative. For half faces and most whole faces (i.e. those where top and bottom both favored the same target) this was straightforward. On trials where top and bottom halves favored opposite targets the stronger of the two halves indicated the correct response.

Orientation had almost no effect on accuracy, confirming that calibration successfully equated upright and inverted half face morphs strengths. There were also no significant effects of orientation on the dilution effect, so we present results collapsed across upright and inverted orientation in order to concentrate on evidence level and split. Mean accuracy, collapsed across target, half (top vs. bottom), and orientation is shown in Figure 2. Accuracy tends to increase with evidence strength, providing a coarse check that the stimuli were appropriately calibrated. Accuracy with M/oW faces was below that of even the weak half faces, confirming that these opposite halves were indeed taken as evidence for the alternative category.

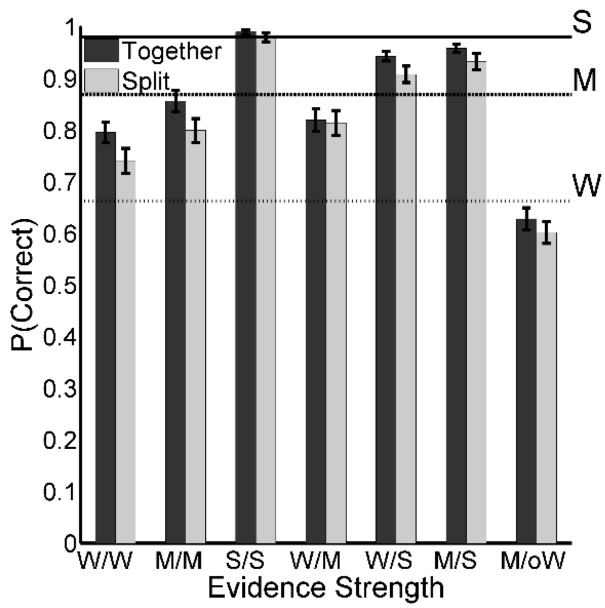

Figure 2: Mean accuracy across evidence levels for whole faces (bars) and half faces (lines).

Our primary research question dealt with how people would combine the two halves of a face and when they might show something akin to the dilution effect. To address this question we compared accuracy with each whole face to that with the stronger half alone. Deviation scores were calculated within individuals by subtracting the mean accuracy given the stronger half face from the response (coded as correct or incorrect) given for whole face. A value greater than 0 indicates additive integration, qualitatively in line with the predictions of a simple Bayesian model. A result less than 0 indicates a dilution effect because additional weak positive evidence decreased accuracy. Figure 3 shows mean deviation scores.

A t-test showed mean deviation scores to be significantly below 0, $t(6767) = 18.61$, $p < .01$. As expected Split faces produced greater dilution effects than Together faces. A 2 (Orientation) x 2 (Split) repeated measures analysis of variance (ANOVA) confirmed this, with a main effect of Split, $F(1,18) = 12.36$, $MSE = .015$, $p < .01$. No other effects were significant. Dilution was greatest for W/S faces, where the difference in top and bottom half strengths was largest. Additive effects were largest in the W/W condition, suggesting that some near-optimal information sampling may have occurred.

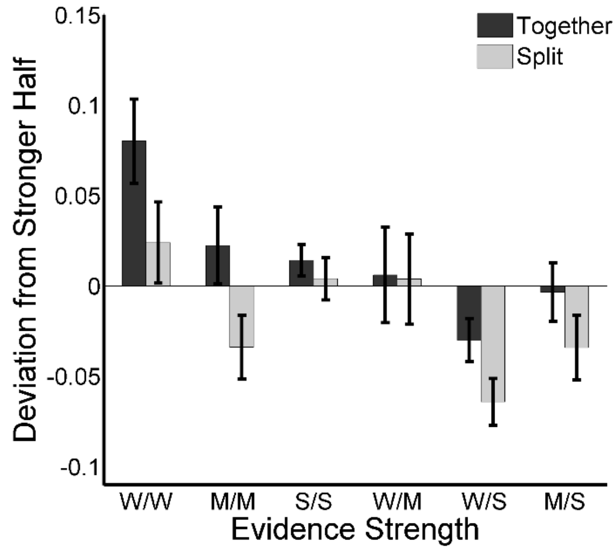

Figure 3: Mean dilution scores across whole face conditions.

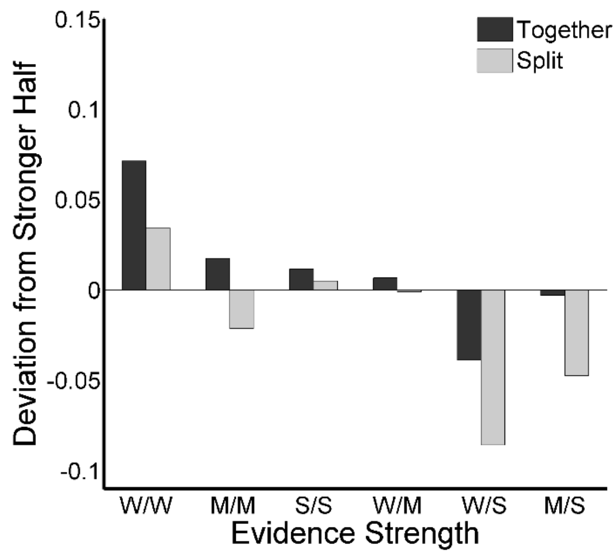

Figure 4: Predicted deviation scores for the McIA model.

Discussion

These findings are rather surprising given that the dilution effect had not previously been observed in a perceptual context. The bulk of the existing literature suggested that performance would probably resemble that of a near-optimal integration process, but we found that participants were often less accurate with two pieces of diagnostic evidence than one. Clearly sub-optimal information integration is not limited to the numerical or linguistic stimuli found in traditional judgment and decision making research. Additionally, our results provide insight into how people processed information in the task. As predicted, dilution was greater when automatic perceptual integration of top and bottom halves was made more difficult by splitting the face. Surprisingly, the dilution effect was present to an equal extent in both Inverted and Upright orientations. Since, inversion was meant to interfere with strategies tailored to upright faces, this may suggest that participants did not treat the stimuli as they would normal faces. In contrast, the split manipulation may have operated at a lower level where splitting disrupted general purpose whole-object automatic processing in either orientation.

Models of Perceptual Dilution

The obtained pattern of results poses problems for two of the candidate models introduced earlier. The Simple Bayesian model posits that information from top and bottom halves is combined optimally. Consequently, the model predicts additive effects in all conditions, except M/oW, and cannot account for the large dilution effects observed when a W half was paired with a S half.

The Averaging model, on the other hand, assumes that individuals always integrate top and bottom halves of face by taking the average of the evidence produced by each. The model predicts dilution effects whenever top and bottom evidence strengths are unequal, but it predicts deviations scores near 0 for conditions where top and bottom are of the same strength. Thus, the large additive effects in the W/W condition cannot be explained through averaging alone.

As an alternative we propose the *Multi-component Information Accumulation* model (McIA). This model represents information integration as a process of accumulating evidence to a decision threshold, θ. According to the model, on each trial a participant repeatedly samples information from one of three sources of evidence: the top half, the bottom half, or the whole face. Each sample provides evidence causing preference to move toward one of two decision bounds. These boundaries represent the amount of preference required to make each response. At one moment a sample may favor the Jones response, causing the preference state to take a step toward the Jones boundary. However, the next sample may favor Smith, causing the preference state to step away from the Jones boundary and toward the Smith boundary. In this manner preference evolves as a noisy random walk process until a decision threshold for one response is reached. The model is thus capable of making predictions for both accuracy and response times.

The probability of sampling whole face evidence is a free parameter, α, representing the likelihood that the perceptual system would automatically combine the top and bottom halves into a single whole face. Since splitting the halves apart increased the size of the dilution effect, α, was estimated separately for Split and Together faces. The probabilities of sampling the top or the bottom half were then each $(1 - \alpha)/2$. The probability of stepping toward the correct decision boundary after a sample is given by the rate parameter, δ. Since the rate of evidence accumulation should vary with stimulus strength, six rate parameters were estimated. These corresponded to W, M, and S morphs for both top and bottom half faces. For half face trials there is only one δ to sample at each moment. However, on whole face trials one of three sources of evidence is sampled at each moment. For example, if a stimulus was comprised of a W top and S bottom, the three sources would be $\delta_{Weak\ Top}$, $\delta_{Strong\ Bottom}$, and whole face evidence produced by automatically integrating the two halves. If the whole face evidence is sampled, an evidence accumulation rate is calculated as the Bayesian optimal combination of top and bottom rates, assuming independence. This represents the idea that automatic perceptual integration of top and bottom halves produces additional, perhaps configural, evidence for the correct response. The value of θ proved relatively unimportant for fitting choices, and was arbitrarily set to 10. In the future we plan to use the McIA model to simultaneously fit choices and response times, which will allow for better estimation of θ.

The best fitting parameters of the McIA model are given in Table 1. Deviation scores based on the model's prediction are shown in Figure 4. The model does a remarkable job of capturing the basic qualitative patterns in the data. It produces dilution effects because deliberation is sometimes driven by the evidence in the weaker half, producing more errors than with the stronger half alone. In the W/S condition this produces very large dilution effects because, for example, $\delta_{Weak\ Bottom}$ is much smaller than $\delta_{Strong\ Top}$. However, unlike the Averaging model, the McIA model does not always predict dilution. Instead it posits that on some trials the perceptual system automatically combines the top and bottom halves into a configural whole, yielding high accuracy. This explains the additive effects for W/W, as well as the difference between Split and Together conditions. According to the model whole face evidence was sampled 63% of time for Together faces, but only 37% of time for Split faces. This supports our hypothesis that separating the top and bottom halves of face encourages more controlled, less optimal strategies.

Table 1: Best Fitting Drift Rate and Attention Parameters of the McIA Model.

$\delta_{Weak\ Top}$	0.522
$\delta_{Medium\ Top}$	0.549
$\delta_{Strong\ Top}$	0.586
$\delta_{Weak\ Bottom}$	0.518
$\delta_{Medium\ Bottom}$	0.525
$\delta_{Strong\ Bottom}$	0.594
$\alpha_{Together}$	0.627
α_{Split}	0.365

Conclusion

The present results represent a synthesis of two divergent trends in the extant literature. We used the stimuli and procedures of a perceptual categorization study to investigate a central decision making phenomenon. Unlike in many previous studies using perceptual stimuli, we found widespread and reliable sub-optimal integration, in the form of the dilution effect. Informative differences in the size of this effect were also found. The Together condition, which encouraged automatic face processing, yielded relatively little dilution compared to the Split condition, which encourages more controlled integration.

Note that we do not see processing mode as a binary concept, but rather a continuum between the extremes of fully automatic and fully controlled integration. To the degree that deviation scores were higher for Together conditions than for Split conditions, we posit a greater degree of automatic integration. The McIA model instantiates this idea through a random walk process with three sources of evidence, the top half alone, the bottom half alone, and the whole face, which represents instances where the perceptual system automatically combines the evidence from the two halves. The model explains how processing was modulated by stimulus presentation. Since Together faces were more naturalistic stimuli, participants were able to sample whole face information more often, yielding greater accuracy.

We also found interesting differences in the size of the dilution effect across levels of evidence strength. For conditions where the top and bottom halves were very unequal, significant dilution was observed. The McIA model produces this result by switching attention between top and bottom halves as it repeatedly samples information. Over time, this effectively averages the evidence in each half. In contrast, additive effects were observed in several conditions where top and bottom strengths were equal. The McIA model also predicts this result because averaging the evidence strengths of these two halves (as described above), produces deviation scores near 0. However, when whole face evidence is sampled, the probability of stepping toward the correct boundary is the Bayesian optimal combination of the two half face δ values. These whole face samples push accuracy above that of the stronger half alone.

These results pose a serious challenge to the idea that integration of perceptual information is always well described by rational models. The prevalence of dilution effects for even the most natural of stimuli suggests that there is still more work to be done to fully bridge the span between optimal integration in perceptual and sub-optimal integration in judgment and decision making. This work is a first step toward determining the conditions under which sub-optimal information integration is to be expected. In future work we plan to extend this experimental paradigm to investigate other paradoxical phenomena, such as the conjunctive fallacy, the disjunction effect, and availability effects.

References

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

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 598-612.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 50-71.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*, 684-704.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Oxford: Oxford University Press.

LaBella, C., & Koehler, D. J. (2004). Dilution and confirmation of probability judgments based on nondiagnostic evidence. *Memory & Cognition*, *32*(7), 1076-1089.

McKenzie, C. R., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: change in belief afer hearing two sides of a dispute. *Journal of Behavioral Decision Making*, *15*, 1-18.

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, *13*, 248-277.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39-57.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172-191.

Peters, E., & Rothbart, M. (2000). Typicality can create, elimate, and reverse the dilution effect. *Personality and Social Psychology Bulletin*, *26*(2), 177-187.

Philips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE transactions on pattern analysis and machine intelligence*, *22*(10), 1090-1104

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica, 39*, 83-89.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM--retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*(2), 145-166.

Sides, A., Osherson, D., Bonini, N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory and Cognition, 30*(2), 191-198.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kerns, S. A. Solla & D. A. Cohn (Eds.), *Advances in Neural Information Processing Systems* (Vol. 11, pp. 59-68). Cambridge, MA: The MIT Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.