Modeling the Joint Effects of Experiences and Descriptions on Impressions and Choices

A thesis presented to
the faculty of
the College of Arts and Sciences of Ohio University

In partial fulfillment
of the requirements for the degree
Master of Science

Nathaniel D. Phillips

August 2010

© 2010 Nathaniel D. Phillips. All Rights Reserved.

This thesis titled
Modeling the Joint Effects of Experiences and Descriptions on Impressions and Choices

by
NATHANIEL D. PHILLIPS

has been approved for
the Department of Psychology
and the College of Arts and Sciences by

Claudia González-Vallejo
Professor of Psychology

Benjamin M. Ogles
Dean, College of Arts and Sciences

ABSTRACT

PHILLIPS, NATHANIEL D, M.S., August 2010, Psychology

Modeling the Joint Effects of Experiences and Descriptions on Impressions and Choices

(116 pp.)

Director of Thesis: Claudia González-Vallejo

How do individuals form impressions of options when they have access to population-level descriptions and sample-level experiences? In the current study, participants were given descriptions of two options and were then given the opportunity to learn about the options. Results showed that inaccurate descriptions led to poorer performance than accurate descriptions. Additionally, the influence of descriptions was positively related to the perceived credibility of the source of descriptions. Therefore, perceived source credibility amplified the effect of the accuracy of descriptions of options. A mathematical model is proposed to describe how descriptions are combined with experiences in dynamic choice tasks.

Approved: _____

Claudia González-Vallejo

Professor of Psychology

ACKNOWLEDGMENTS

I would like to thank my committee chair and advisor Claudia González-Vallejo for her support and guidance throughout the process of completing the current document.

I would also like to thank Francis Bellezza and Bruce Carlson for their helpful suggestions and comments to help improve the quality of this research.

TABLE OF CONTENTS

	Page
Abstract.....	3
Acknowledgments.....	4
List of Tables	7
List of Figures.....	8
Introduction.....	11
Decisions from Experience and Decisions from Description.....	14
Combining Descriptions with Experience	18
Descriptive Information and Source Credibility	21
The Description Experience Integration Model (DEIM).....	27
Choices as a Function of Impressions.....	32
The Present Study	33
Model Simulations and Hypotheses	36
Simulations 1 to 4 – Description Accuracy x Source Credibility Interaction.....	37
Simulations 5 to 8 – Stimuli Discriminability x Description Accuracy	39
Method	43
Participants.....	43
Procedure	43
Results.....	47
Model Comparisons.....	47
DEIM Parameter Tests.....	49
Description Discounting Parameter w	51
New Experiential Information Discounting Parameter p.....	53
Choice Model.....	55
Summary of Model Fit Analyses	57
Model-Free Comparisons of Impression and Choice Paths.....	58
Source Credibility x Description Accuracy Interaction.....	58
Description Accuracy x Stimuli Discriminability Interaction	61
Summary of Model-Free Analyses	63

Discussion	64
Limitations and Directions for Future Research	65
Conclusion	70
References	113

LIST OF TABLES

	Page
Table 1: Parameters for simulations 1 through 4	71
Table 2: Parameters for simulations 5 through 8	71
Table 3: Parameters for point distributions of options in each experimental condition ...	73
Table 4: Labeling of experimental conditions	74
Table 5: Labels and parameters of 3 models tested	75
Table 6: MSE Data from model validation tests.....	76
Table 7: Parameter estimation results for the description decay weighting parameter w for the full DEIM model	77
Table 8: Parameter estimation results for the new experiential information discounting parameter p for the full DEIM model.	78
Table 9: Descriptive statistics for estimates of impression sensitivity choice parameter c	79
Table 10: Predicted versus actual choices	80
Table 11: Means and standard deviations of choices from best option over all 50 trials for Credibility x Accuracy interaction.....	81
Table 12: Means and standard deviations of choices from best option over all 50 trials for Accuracy x Discriminability interaction.	81

LIST OF FIGURES

	Page
<i>Figure 1:</i> Description information weights as a function of trial and w.....	82
<i>Figure 2:</i> Aggregate results from Simulation 1: Accurate descriptions from a source low in credibility.....	83
<i>Figure 3:</i> Aggregate results from Simulation 2: Accurate descriptions from a source high in credibility.....	84
<i>Figure 4:</i> Aggregate results from Simulation 3: Inaccurate descriptions from a source low in credibility.....	85
<i>Figure 5:</i> Aggregate results from Simulation 4: Inaccurate descriptions from a source high in credibility.....	86
<i>Figure 6:</i> Predicted proportion of choices from the best option as a function of source credibility and description accuracy.	87
<i>Figure 7:</i> Aggregate results from Simulation 5: Accurate descriptions from stimuli that are relatively easy to discriminate.....	88
<i>Figure 8:</i> Aggregate results from Simulation 6: Accurate descriptions from stimuli that are relatively hard to discriminate.....	89
<i>Figure 9:</i> Aggregate results from Simulation 7: Inaccurate descriptions from stimuli that are relatively easy to discriminate.....	90
<i>Figure 10:</i> Aggregate results from Simulation 8: Inaccurate descriptions from stimuli that are relatively hard to discriminate.....	91
<i>Figure 11:</i> Aggregate choices from the best option from simulations 5 through 8.....	92

<i>Figure 12:</i> Screenshot of the beginning of a hypothetical day.	93
<i>Figure 13:</i> Screenshot of a hypothetical participant who had just selected the company Bracken.	94
<i>Figure 14:</i> Distribution of R^2 values for individual participants across all experimental conditions.	95
<i>Figure 15:</i> Distribution of estimates of w for each participant by experimental condition.	96
<i>Figure 16:</i> Mean estimates of w parameter by experimental condition.	97
<i>Figure 17:</i> Distribution of estimates of p for each participant by experimental condition.	98
<i>Figure 18:</i> Mean estimates of p parameter by experimental condition.	99
<i>Figure 19:</i> Distribution of estimates of c for each participant by experimental condition.	100
<i>Figure 20:</i> Mean estimates of c parameter by experimental condition.	101
<i>Figure 21:</i> Predicted proportion of choices from Astride versus observed proportion of choices from Astride.	102
<i>Figure 22:</i> Aggregate results of participant data receiving an accurate description of distributions from a source low in credibility.	103
<i>Figure 23:</i> Aggregate results of participant data receiving an accurate description of distributions from a source high in credibility.	104
<i>Figure 24:</i> Aggregate results of participant data receiving an inaccurate description of distributions from a source low in credibility.	105

<i>Figure 25: Aggregate results of participant data receiving an inaccurate description of distributions from a source high in credibility.....</i>	106
<i>Figure 26: Mean choices from the best option as a function of source credibility and description accuracy.....</i>	107
<i>Figure 27: Aggregate results of participant data receiving an accurate description of distributions that were relatively easy to discriminate.....</i>	108
<i>Figure 28: Aggregate results of participant data receiving an accurate description of distributions that were relatively hard to discriminate.....</i>	109
<i>Figure 29: Aggregate results of participant data receiving an inaccurate description of distributions that were relatively easy to discriminate.....</i>	110
<i>Figure 30: Aggregate results of participant data receiving an inaccurate description of distributions that were relatively hard to discriminate.....</i>	111
<i>Figure 31: Mean choices from the best option as a function of description accuracy and stimuli discriminability</i>	112

INTRODUCTION

Slovic aptly recognized that "Preferences are not simply read off some master list but are constructed on the spot by an adaptive decision maker...decision making is a highly contingent form of information processing, sensitive to task complexity, time pressure, response mode, framing, reference points, and numerous other contextual factors" (1995, p. 369; also see Simon, Krawczyk, Bleicher & Holyoak, 2002). While some preferences are easier to report than others (e.g.; would you rather own a car with or without a working engine?), others are more difficult, especially those made in unfamiliar domains where one has no relevant past experience. For example, imagine that you had never consumed coffee before and were asked whether you would prefer a dark roast or a light roast. Because you had never tried coffee before, you would probably not have a definite preference. Therefore, rather than report an established preference to the question, you would likely construct a tentative preference that is influenced by factors such as generalizations from other similar experiences. Moreover, this preference would be likely to change as you gain experience drinking different types of coffee. Indeed, research by Hoeffler, Ariely & West (2006) has shown that in a preference development task where participants have a goal of determining an ideal concoction of lemonade, early experiences with certain concoctions strongly predicted final preferences. More generally, Hoeffler et al. concluded that sampling decisions, which provide information for preference development, are strongly affected by prior experiences and expectations of future experiences.

Research by Denrell and colleagues (Denrell, 2005; Denrell & March, 2001; Denrell, 2007) has also provided insight into experience-based impression formation. Denrell (2005) used non-motivationally based models of group impression formation to show that a spurious and fragile negative impression of an out-group member can be resistant to change because of one's lack of exposure to out-group members (due to less sampling of the out-group). The reduced sampling can be the result of an adaptive tendency to avoid interactions that one expects will lead to negative outcomes. Thus, veridical negative initial experiences of out-group members lead to less sampling of experiences with out-group members. Paradoxically, it is exactly this lack of future sampling that allows initial impressions to remain negative and resistant to correction. This suggests that a psychologically unbiased observer who ceases to sample due to the reasonable expectation that sampling will result in a negative outcome based on past experiences will retain biased negative impressions.

Other research demonstrates an interplay between impression formation and experiential feedback. Namely, individuals' ability to update impressions online (at the time of stimuli encoding) as a result of experienced feedback/success is supported by a substantial body of research (Betsch, T., Plessner, H., Schwieren, C. & Gutig, R., 2001; Hogarth & Einhorn, 1992). One particularly informative study conducted in this domain is that of Betsch et al. (2001). Betsch et al. (2001) found that individuals can generate implicit value accounts of objects that change over time as a result of feedback. In their experiments, participants were told that their task was to judge the quality of various ads on a computer screen. Along with the ads, various changes in stock prices for 4

hypothetical stocks were shown on the screen, ostensibly as distraction information which participants were instructed to (mostly) ignore. In reality, Betsch et al. were interested in the degree to which these participants would be able to generate an implicit idea as to the final relative value of the stocks. At the end of the task, Betsch et al. found that while participants were not able to explicitly estimate the final cumulative value of each stock individually; however, they were able to report the relative value of each stock with moderate accuracy. Betsch et al. concluded that their participants generated implicit value-accounts for each of the stocks that were updated over the course of the stimuli presentation. Additionally, participants in their experiment were not consciously aware of their accuracy suggesting that the updating process was, at least to some degree, non-conscious (also see Hasher & Zacks, 1984).

While Betsch et al.'s research supported the idea that people use online value accounts to form impressions online, they did not specify the structure of the account or provide a quantitative model. An excellent candidate quantitative model for this hypothetical value account is Hogarth and Einhorn's (1992) belief-updating model. The belief-updating model was designed to model the process by which individuals form and update their beliefs about causal explanations for events over time. Because belief-updating over time is a sequential process, Hogarth and Einhorn predicted that impressions would be sensitive to order effects similar to those in memory. While the authors studied several variations of belief updating processes, the main structure of their model suggested that belief-adjustment is a recursive averaging and adjustment process. They posited that when individuals encounter each piece of information relevant to a

hypothesis, they compute a weighted average of that that information and their pre-existing belief. Hogarth and Einhorn described several variations of this generic model in their paper, but for the purpose of this review one key conclusion will be highlighted. Namely, serial order effects in impression formation are captured by the size of an adjustment weight. When this weight is large, incoming information is weighed more heavily relative to existing impressions and this leads to stronger recency effects. Alternatively, smaller values of the adjustment weight reflect less sensitivity to recent events and larger sensitivity to existing impressions.

These lines of research all carry a central theme that is of primary importance to the present work: when presented with sequential pieces of information, decision-makers can form and update judgments dynamically over time. As such, any situation that provides decision-makers with incremental pieces of decision-relevant information over a period of time will likely cause individuals to use this information to recursively update their prior judgments in some systematic fashion.

Decisions from Experience and Decisions from Description

Decision scientists have distinguished between two types of decisions: description-based, and experienced-based. These types are distinguished on the basis of the type of information used by decision-makers to form impressions of options. In description-based decisions, relevant information is presented verbally and typically depicts an abstract trait or quality of an object. For example, a gamble “A” might be described as follows: “Gamble A has an underlying probability of success equal to a normal distribution with mean 10 and standard deviation 2.” In contrast, an experience-

based decision provides relevant information in the form of sequential, concrete values. For example, the value of gamble A could be conveyed to a person with a presentation of the following successive trials constituting a sample from gamble A: {9.89, 11.22, 10.08}.

There are several reasons why a decision-maker presented with this experiential information is likely to form a systematically different impression of a gamble than one with descriptive information. First, experiential data are necessarily sample-based, and thus are subject to sampling error. This is a very important aspect of experiential data as it means that experiences will likely never provide a complete picture of underlying theoretical qualities. Further, even an unbiased sampling process can produce samples that misrepresent population level parameters to some degree. Second, experiential information may provide more psychological weight than descriptive information. For example, witnessing a single plane crash on television may outweigh any statistical evidence that plane travel is very safe, and lead individuals to form overly negative impressions of plane travel. Indeed, the most successful explanation of this effect is the availability heuristic (Tversky & Kaneman 1974), which states that subjective probabilities are assessed by the ease in which relevant *instances* are brought to mind. If easily accessible relevant instances are informed by experiences, especially recent ones, then it seems likely that experiential information informs probability judgments and impressions.

In partial support of this hypothesis, Hertwig, Barron, Weber and Erev (2004) conducted a series of studies comparing the effects of feedback-based decisions to

description-based decisions. Participants were presented with 6 decision problems that required a choice between two gambles with different expected values. Each of these gambles contained a rare event (an event with a probability of occurrence less than .30) that was critical to the expected value of the gamble. For example, one gamble had a .90 probability of resulting in nothing, and a .10 probability of \$32. Half of the participants, the description group, were given descriptive information (all relevant probabilities and payoffs) about the gambles. The other half of the participants, the experience group, was not told anything about the two gambles. This group was instead presented with the two gambles on a computer and was told to repeatedly sample from either of the two gambles to determine their value. Both groups of participants then selected their gamble of choice and played it for real money. Hertwig et al. found that while the decision group showed an overweighting of rare events (consistent with prospect theory) (Kahneman & Tversky, 1979), the experience group showed an *underweighting* of rare events. For example, in one choice problem participants were presented with one gamble (H) that gave a loss of -3 with probability 1.00 and another (L) that gave a loss of -32 with probability .10 and 0 otherwise. Thus, the expected value of the gamble without the rare event was -3 and the expected value of the gamble with the (highly negative) rare event was -3.2. Results showed that while 64% of the description group selected the gamble with the higher expected value (H), only 28% of the experience group did so. They attributed this finding to individuals' information search strategies, and decisions made as a result of their searches. Namely, participants in the experience-group made only a median of 15 draws from the gambles before making a decision. Because the probability of experiencing a

rare event is positively related to the number of samples one takes, this low median number of samples means that many participants likely never encountered the rare event. Moreover, because the binomial distribution is skewed for small sample sizes, experience-based decision makers are less likely to encounter rare events than expected given their objective likelihood.

Another factor that may have led to this finding is a recency effect due to the sequential updating nature of participants' impression formation. As mentioned before, research by Hogarth and Einhorn (1992) showed that sequential belief-adjustment processes are subject to recency effects similar to those found in memory. Namely, recent events are given higher weights than less recent events. As noted by Hertwig et al. (2004), recency effects result in underweighting of rare events even in large samples because rare events are less likely than common events to be experienced recently. In support of this hypothesis, Hertwig et al. (2004) separated the first half of participants' sequence of draws from their second half. They found that while the first half of draws predicted 59% of participants' final choices on average, the second half predicted 75% of participants' final choices. This suggests that more recent events do indeed have more influence on judgments than less recent events.

Research by Hertwig et al. (2004) and others (e.g.; Barron & Erev, 2003; Fiedler, 2000) have demonstrated convincingly that decisions from experience are psychologically distinct from decisions from description. Yet most research on risky decision-making has focused on decisions from description. A meta-analysis performed by Weber, Shafir and Blais (2004) found that the vast majority of decision-making

studies prior to 2004 dealing with a two out-come risky prospect and a sure thing were conducted only with decisions from description. Moreover, prospect theory (Kahneman & Tversky, 1979), one of the most successful and well-known theories of decision-making under risk, was developed and tested exclusively using description-based studies.

The next theoretical step from recently expanding research contrasting decisions from experience and decisions from description is how decision makers behave when they have access to *both* descriptive and experiential information. To my knowledge, this question has not been addressed directly. The purpose of the current study is to develop, explore, and test a model of impression formation in situations where individuals have access to both descriptive and experiential information.

Combining Descriptions with Experience

A distinct area of research which has addressed processes akin to that surrounding the description / experience gap relates to human's abilities to update probability estimates in a Bayesian manner. Bayes' theorem is the most widely accepted normative model of probability estimation in tasks where probability estimates of events are updated in response to new information. Due to its status as the normative standard, psychologists have conducted extensive research on the extent to which individuals' probability revisions coincide with Bayes theorem. Edwards, Lindman and Phillips (1965) conducted a series of studies addressing this question. The paradigm they used was as follows, participants were presented with options (e.g.; bags of poker chips, bags of marbles) containing different known proportions of target events (e.g.; red chips, black marbles). The experimenter then selected one of the options covertly, and made successive

selections of events with replacement from that option. The goal of the participant was to report his or her estimate of the probability that each of the options was the one selected by the experimenter after each sample. Because the selection of the target option was ostensibly random, participants know the prior probability that each option was the one selected (usually $1 / n$). Furthermore, they know the probability of each event (e.g., a red chip) given each option. Therefore, participants are given all the information necessary to update their probability estimates according to Bayes' theorem.

Edwards and colleagues early experiments concluded that individuals do not update their beliefs in strict accordance with Bayes' theorem. One deviation that individuals had from the normative standard was that they were sensitive to the order in which evidence was presented to them. In one early study addressing the effects of stimuli order, Peterson and DuCharme (1967) presented participants with urns containing different proportions of red and black chips. They then read a sequence of outcomes that had ostensibly been acquired by selecting chips at random from one of the two jars. Participants heard a sequence of stimuli that favored one option in the first half of the study, and then flipped to favor the other option in the second half. Specifically, individuals exhibited a primacy effect, wherein data that occurred early in a stimuli sequence affected individuals judgments more than information presented later in a stimuli sequence (Peterson & DuCharme, 1967). That is, individuals were conservative in their revisions to the subjective probability estimates. This finding is consistent with Asch's hypothesis that individuals form impressions early and distort subsequent evidence to fit that initial impression (Asch, 1946). This account has also been verified by

more recent research (e.g.; Bond, Carlson, Meloy, Russo & Tanner, 2007). Interestingly, this finding does seem at odds with prior research by cognitive modelers (e.g.; Hogarth & Einhorn, 1992; Hertwig et al. 2006) who have found evidence for recency effects in impression formation. This inconsistency in the literature calls for additional research.

Base-rate neglect is another common finding in the literature that is at odds with Bayesian reasoning. Kahneman and Tversky's (1982) taxi cab problem presented participants with a vignette designed to test individuals' sensitivity to base-rate information. Participants were told that two taxi cab companies, blue and green with different numbers of cars operate in a city. Instead of the companies being equally represented, the majority (85%) were known to be green, while the minority (15%) was known to be blue. Participants were told to imagine that an observer had witnessed an accident involving a blue cab, and that the observer had an 80% chance of being correct (in knowing the color of the cab). The participants' task was to determine the probability that the cab was indeed blue given this information. The answer given by participants (approximately 80%) was much higher than the normative Bayesian answer (41%). Tversky and Kahneman explained this result by suggesting that participants made use of the representativeness heuristic, which states that probabilities of events are determined by the degree to which the event is representative of an appropriate mental model. However, this study and others examining the phenomenon of base-rate neglect have been challenged by other researchers (e.g.; Gigerenzer, 1991). Namely, it has been argued that these results are not the result of a truly biased processing system, but rather are byproducts of the presentation format of information in the studies. Gigerenzer (1994)

provided evidence that when statistical information is presented in a frequency format rather than a probability format, participants do not show base-rate neglect and are more in line with the normative Bayesian model. Despite these exceptions, most research has suggested that humans are not perfect Bayesian thinkers. In different tasks, individuals have been shown to estimate probabilities in ways that are systematically different from what Bayes theorem prescribes.

Descriptive Information and Source Credibility

Descriptive information always comes from a source, and all sources are not necessarily equal. In a psychology experiment, participants are given descriptions from experimenters who can appear trustworthy or shady. A person who is planning an upcoming trip and is deciding whether or not it is safe to travel to a certain foreign country can learn crime statistics from a credible newspaper or from a passerby. These differences in sources are likely to result in differences in impression formation and attitude change. Moreover, variance in trustworthiness and credibility of a source is likely related to the robustness of judgments based on information from that source (Kugler, Connolly & Kausel, 2009). Given this consideration, it seems reasonable to suspect that a key factor which would affect description / experience integration is the relative credibility of the source of the descriptive information compared to experiential information.

For example, imagine that a college basketball scout hears rave reviews from trusted colleagues about a high-school player and decides to watch her play several games during a state tournament to assess her abilities. Prior to the tournament, the scout

would likely have a high impression of her based on her described abilities. Now imagine that in the first game of the tournament the player has a terrible game – how does the scout update his impression of her based on this information? He could conclude that the description was inaccurate, disregard it completely, and assign the player a poor score based entirely on her performance in the game. Alternatively, he could conclude that she probably just had a bad game (due to random variation), and leave his high impression of her based on her description intact. After all, he heard from many trusted colleagues that she was a great player and he probably can't make any definitive conclusions based on a single game. But what happens if the scenario changes and instead of the description coming from several trusted colleagues it comes from a phone call from the player's father? In this case, the scout would be much more likely to discard the description quickly in the face of actual experiential information because the player's father is likely to exaggerate his daughter's playing ability. Here, the informational *value* of the description is much lower than that of first-hand experience because there is reason for the scout to suspect its validity, and the scout will likely take this source information into account when integrating the descriptive and experiential information about the player.

In line with this example, social cognition researchers have shown that the credibility of a source does indeed affect the extent to which advice from that source is aggregated with other information. Most notably, the effect of source credibility has been found to interact strongly with elaboration levels under Petty and Cacioppo's elaboration likelihood model (ELM, 1986). The ELM describes persuasion as occurring through either a peripheral or a central route. In the peripheral route, the actual content of a

message is not elaborated using in-depth processing; therefore, people might trust information from a highly credible source regardless of the quality of the information by using source credibility as a cue or heuristic to persuasion. In the central route, information is subjected to high elaboration and is carefully critiqued for its logical merits. Therefore, when arguments are either very strong or very weak, there is no effect of source credibility on persuasion. However, this does not mean that source credibility is totally ignored in all high elaboration conditions. On the contrary, Chaiken and Maheswaran (1994) argued that when the strength of an argument is ambiguous, the perception that a source is high in credibility can lead to a positivity bias in evaluating arguments from that source even in high elaboration conditions. Other research has found that even when source credibility does not affect attitude valence, it can affect the confidence with which individuals hold opinions (Tormala, Brinol & Petty, 2006; Tormala, Brinol & Petty, 2007).

Clearly then, source credibility is a major component of attitude change. As such, any model that attempts to describe the integration of descriptive information from a source and experiential information over time should account for the effects of source credibility. While its qualitative effects have been studied extensively in social cognition research, less has been done on the process by which individuals *quantitatively* aggregate information from a source with experiential information over time. One exception is in research on how individuals aggregate advice from advisors (Budescu & Rantilla, 2000).

In many decision-making contexts, individuals must make decisions that require advice from experts. For example, a novice investor would be unlikely to invest a large

amount of her money in stocks without first consulting experts in the field. Similarly, someone with no medical training who develops excruciating pain in his arm would be better off consulting a doctor for relief than relying on his own ad hoc remedy. While it is uncontroversial that individuals solicit and use advice from experts, what is less obvious is how individuals aggregate advice from multiple experts. Budescu and Rantilla (2000) addressed this question by formulating a descriptive mathematical model of how individuals aggregate advice from experts, make decisions based on that advice, and judge confidence in their final decisions. Based on prior research demonstrating the descriptive validity of averaging models (e.g., Anderson, 1967; Kaplan, 1981), Budescu and Rantilla (2000) derived a model for confidence in the aggregated opinion of a decision-maker obtained by averaging the opinions of multiple judges on the basis of multiple cues, as a function of the features of the task and the characteristics of the experts. Importantly, decision maker's intuitive level of confidence in the aggregate was a monotonically decreasing function of its perceived variance, where the smaller the variance of the final estimate, the higher the decision maker's confidence. Variance in the estimate was a function of several factors, including the imperfect probabilistic relation between a cue and the target event, and the random components related directly to the expert (such as lack of knowledge and expertise), the number of cues assessed by experts, and the fraction of pair-wise overlap in the cues presented to the experts. To test the model, participants were given several hypothetical decision-making scenarios where they had to combine information from several experts to make a probabilistic decision. For example, in one scenario, participants were asked to imagine that they were working

for a large health care organization and needed to determine the probability that a certain patient has a specific condition based on the advice from several doctors. Participants were told that they would receive the following information: the number of doctors giving them estimates, the total amount of information available, how many pieces of information each doctor had access to, along with the amount of overlap in the information that each doctor saw for that case, and finally each doctor's estimate of the probability that the given patient had the disease. Participants then indicated their subjective assessment of the probability that the patient had the disease, followed by their confidence in their estimate on a 7 point Likert-like scale. Budescu and Rantilla (2000) found support for several predictions of their model including the prediction that when experts were equally qualified and informed, decision-makers averaged their forecasts.

What Budescu and Rantilla (2000) did not do is explore how individual judges incorporate advice from experts and their own experiences dynamically over time in a judgment formation domain. Rather, Budescu and Rantilla focused on how judges integrate advice from *multiple* experts at one point in time to form an aggregate judgment. Furthermore, their research focused on how judges form static subjective confidence judgments rather than actual choices between competing options over time. Furthermore, participants were not given information regarding the accuracy or credibility of specific judges. Rather, they were expected to infer the aggregate accuracy and credibility of judges based on the correspondence of multiple judges. Additionally, because Budescu and Rantilla (2000) did not explore the effects of first-hand experiential information on judgments, they were not able to assess the effects of *stimuli* variance on

judgments. Presumably, just as greater variance in expert advice leads to less confidence, greater variance in first-hand experiential information should also lead to less confidence in judgments. That is, as the variance in stimuli output increases, confidence in the true mean stimuli output decreases. For example, compare two individuals who are attempting to determine the mean output of some distribution. One individual sees the following three samples: 1, 5, 10 and the other sees 4, 5, 6. It is predicted that both individuals will judge that the mean value of the distribution equal to 5, but that the second individual will hold this judgment with higher confidence due to the lower variance in stimuli output. Further, this confidence in judgment will lead to confidence in choices, insofar as the decreased variance in stimuli output will make differences in the mean outputs of distributions more apparent.

At this stage, two main suggestions from the literature are relevant to the current study: First, no comprehensive experiment has been conducted to examine the joint effects of descriptions and experiences on judgments. Second, source information (e.g.; credibility) affects the extent to which the information is used in impression formation. Third, averaging models have shown to be successful in describing information from multiple sources. And fourth, the normative Bayesian model of information updating is not always a good descriptive model. In this paper, I develop a mathematical model that incorporates the effects of both descriptive and experiential information on judgments. This model will be applied to a decision task where individuals form impressions of two options while concurrently attempting to maximize point rewards by focusing their attention on the best option. First, I will develop the impression formation model as it

relates to both experiential and descriptive information. Second, I will describe the choice model. I will then present a series of simulations used to explore predictions from the model in an experimental task.

The Description Experience Integration Model (DEIM)

Several models of online impression formation have been proposed in the literature (Hogarth & Einhorn, 1992; Hertwig, Barron, Weber & Erev, 2006; Anderson, 1967; Denrell, 2005; March, 1996). These models describe impression updating as a weighted adjustment process where past impressions of an option are averaged with new information related to that object. Because of the similarity of their experimental task to the one conducted in this study, I chose to use the value-updating model used by Hertwig et al. (2006) to model experiential impression updating. The model is represented as follows:

$$A_{q_t} = \left\{ \begin{array}{ll} A_{q_{t-1}} & \text{if option } q \text{ is not chosen on trial } t \\ \left(1 - \frac{1}{t^p}\right) A_{q_{t-1}} + \frac{1}{t^p} x_{q_t} & \text{if option } q \text{ is chosen on trial } t \end{array} \right\} \quad (1)$$

where, A_{q_t} is the experiential impression at trial t [$0 \leq A_{q_t} \leq 1$], x_{q_t} is the t^{th} piece of evidence for option q [$0 \leq x_{q_t} \leq 1$], and p is the adjustment weight for experiential information [$0 \leq p \leq 1$].

The p parameter captures resistance to new experiential information during averaging. A value of p equal to 1 indicates perfect averaging of all information. As values of p increase, sensitivity to new evidence decreases. A large value of p could indicate stubbornness, perpetual discounting of new information, or a variety of other psychological states. For example, a person who is firm in her convictions that a target

individual is a staunch conservative, may refuse to accept instances of the target expressing more liberal beliefs as diagnostic of the target's true values. Research by Hertwig et al. (2006) on the role of information sampling in risky choice fit this model to data in an experience-only judgment domain and calculated the best fitting value of p to be approximately .30.

A potential problem with using equation 1 in practice is that it is difficult, if not impossible to measure experiential impressions directly when impressions are based on both experiences and descriptions. That is, in situations where individuals are given both a description and experiences related to an option, it is assumed that their verbal impressions of that option are based on a combination of descriptive and experiential information. Moreover, it is assumed that people are unable to introspectively delineate the effects of descriptions and experiences on their judgments (see Nisbett and Wilson, 1977). Because the purpose of this study was to understand impression updating in a joint experience/description task, an alternative version of the experiential updating model was developed to better match the current experimental design. Here, the value of individuals' prior experiential impression of an option ($A_{q_{t-1}}$) is replaced by the true empirical running average of all sample data from that option:

$$A_{q_t} = \left\{ \begin{array}{ll} A_{q_{t-1}} & \text{if option } q \text{ is not chosen on trial } t \\ \left(1 - \frac{1}{t^p}\right) \bar{x}_{q_{t-1}} + \frac{1}{t^p} x_{q_t} & \text{if option } q \text{ is chosen on trial } t \end{array} \right\} \quad (2)$$

where, $\bar{x}_{q_{t-1}}$ is the true empirical running average of all sample data for option q through trial $t-1$. A more thorough discussion of this issue is contained in the discussion section of this paper.

Equation 2 describes online updating of experiential information only. Thus, individuals are assumed to begin with a ‘blank slate’ in reference to the target. To incorporate impressions of descriptive information “D” into this model of experiential information updating, a second model results:

$$I_{q_t} = \left\{ \begin{array}{ll} I_{q_{t-1}} & \text{if option } q \text{ is not chosen on trial } t \\ \left(1 - \frac{1}{tw}\right) A_{q_t} + \frac{1}{tw} D(q) & \text{if option } q \text{ is chosen on trial } t \end{array} \right\} \quad (3)$$

where I_{qt} is the combined impression of option q after the t^{th} trial, $D(q)$ is the description of option q , w is the adjustment weight for descriptive information, and A_{q_t} is the experiential impression of option q after the t^{th} trial (as defined by Equation 2).

An individual’s combined impression therefore, is a weighted average of their impressions based on experiential and descriptive information. In this model, experiential impression formation (Equation 2) is independent of descriptive information. When experiential (A_{q_t}) and descriptive ($D(q)$) information are combined to form a unitary impression (Equation 3) they are entered as separate, independent quantities. This assumption could be violated in situations where the encoding and updating of experiential information is affected by its relation to descriptive information. For example, if a person has descriptive information telling him or her that most people win \$5 when playing a certain gamble, then an experienced outcome of \$3 might feel like a loss compared to a situation where he or she believes that most people only win \$1. Thus, descriptive information could serve as an anchor or reference point by which experiential information is subjectively interpreted in Equation 2. However, for the purposes of this

initial model, encoding and updating of experiential information is assumed to be independent of descriptive information.

The weight for descriptive information is assumed to be a function of the amount of experiential information one has been exposed to (t); conceptually, this is meant to account for the prediction that as time passes and experiential information accumulates the influence of the description of the option on impressions decreases. This prediction is derived from the assumption that the use of the experiential reasoning system is directly proportional to the quantity of experiences a decision maker has access to. This also is supported by research on the availability heuristic which has found that estimates for the probability of an event are directly proportional to the frequency and intensity of examples of that event which can be brought to mind (Tversky & Kahneman, 1974).

The rate of decrease is allowed to vary in this model, as represented by the parameter w . As w decreases, $\frac{1}{t^w}$ approaches one, meaning that descriptive information is weighed more heavily than experiential information. As w increases, descriptive information is quickly washed away by experiential information over time. When w is 0, $\frac{1}{t^w} = 1$, thus impressions are entirely based on descriptive information and are unaffected by experienced outcomes.

One psychological factor that could affect values of w is the perceived credibility of the source of descriptive information. Past research has found that the credibility of a source can significantly affect how a listener uses information from the source (e.g.; Tormala, Brinol & Petty, 2007). Generally, the more credible a source is, the greater the degree to which information from the source will be used in making a decision. This

would be represented by small values of w , and thus large values of $\frac{1}{t^w}$. In contrast, if descriptive information is believed to not be very credible, then experiential information would be more highly weighted in impression formation relative to descriptive information. For example, if a used car salesman told you that a certain car was in great condition, you might question the validity of his recommendation. Consequently, if you take the car for a test drive and find that it shakes when you use the brakes, you would likely conclude that the car does not have great reliability and conclude that the car is not in very good condition based on the experienced information you gained in the test drive. In short, descriptive information that is believed to come from an expert, definitive or trusted source should be weighed more heavily relative to experiential information compared to descriptive information from a weak or uninformed source. This should be reflected in the descriptive information decay parameter w in the DEIM.

It is important to note that w is technically an unbounded parameter. Therefore, an individual who does not weight descriptions at all in his or her judgments would technically have an infinitely large value of w . However, because the complete description weight $\frac{1}{t^w}$ is also a function of t which ranges from 1 to 50 in the current study, moderately smaller values of w (less than 10) can be sufficient to describe categorically fast discounting of descriptions. Consider Figure 1 depicting description weights as a function of w and trials. The vertical axis of this chart depicts description weights $\frac{1}{t^w}$ while the horizontal axis depicts trial numbers. Lines show description weights for different values of w . The chart shows that values of w of 5 or greater

represent near 0 weighting of descriptive information after the first trial. Thus, while w is technically unbounded, it is unlikely for values of w to be larger than 5.

Choices as a Function of Impressions

To model choice behavior using updated impressions, a modified version of Luce's choice rule is used (Yechiam & Busemeyer, 1995):

$$p(q | t) = \frac{e^{s_t^c \cdot I_{q_{t-1}}}}{\sum_{j=1}^n e^{s_t^c \cdot I_{j_{t-1}}}} \quad (4)$$

Where s_t is the cumulative number of prior samples of option q at time t , I_{qt} is the impression of option q at time $t-1$, n is the number of options, and c is the impression sensitivity parameter for choices. According to equation 4, the probability that option q is chosen on trial t is a function of the strength of the impression of option q relative to all other options at trial t . Moreover, choice sensitivity to impressions is allowed to change over time. The c parameter models the choice sensitivity to impressions and can take on any real value. When c is 0, options are chosen randomly. As c increases, individuals are increasingly more likely to select the option with the highest impression as t increases. Psychologically, this could reflect the increasing confidence that individuals have in their impressions as they gain experiential information over time. Alternatively, when c is negative, individuals become increasingly *insensitive* to impressions as trials increase. That is, as trials increase, choices will approach become more random. This could reflect

fatigue or boredom over time. In the current study, participants' confidence will not be directly assessed, but will be inferred from estimates of the parameter c .

A situational factor that is predicted to affect confidence in one's relative impressions is the extent to which individuals are able to distinguish between options. If an individual can easily distinguish between two options, then he or she should be confident in their relative value. Conversely, if two options are difficult to distinguish, then individuals will likely distribute his or her samples between the two options to gain more information about each of their distributions.

From signal detection research we know that the ease with which individuals can discriminate between two distributions depends on the discriminability, which is indexed by d' (Macmillan & Creelman, 1991). This index is a function of the variances of the two distributions as well as the differences in their means. As the differences in the means of the two distributions increase, and their respective variances decrease, the discriminability index d' increases indicating easier discrimination. That is, when there is less overlap between two distributions, the range of samples from the two distributions will be less likely to overlap, leading to easier discrimination of the means of the two distributions.

The Present Study

The purpose of the present study is to determine the joint effects of descriptions and experience on judgment and decision making. More specifically, the efficacy of a new model of impression formation will be tested in a repeated choice paradigm. First, a series of simulations were conducted to determine the predicted effects of model

parameters on judgments and decisions. Results from the simulation were combined with findings from past research to derive hypotheses for the study. Second, an experiment was conducted to test experimental hypotheses for both model-specific and model-free instantiations of the hypotheses. Participants in the study were given (potentially misleading) descriptive information about the means of two distributions and were given the goal of obtaining as many points as possible by sampling from the two distributions over several trials.

This task has some key similarities to those used in signal detection studies (Macmillan & Creelman, 1991). In a signal detection task, individuals must determine whether or not a signal was presented on one of multiple trials. Because there is variance associated with the presence of a signal and the absence of a signal (noise), the states are described with two overlapping (typically normal) distributions. Since the distributions overlap, any decision criteria (threshold) chosen by participants is guaranteed to result in both positive and negative errors. The task used in this study departs from signal detection tasks in several fundamental ways. The goal of the participants in the current task is two-fold: they are forming impressions of two distributions, while at the same time attempting to focus their choices on the distribution with the higher mean (to maximize points). In signal detection tasks, individuals do not choose which distribution they are exposed to on each trial as their task is to determine which distribution was drawn from on each trial. As such, they will gain equal amounts of information about both distributions.

Despite these differences, there are significant similarities; in both tasks, individuals are, in some fashion, exposed to stimuli from two different overlapping distributions. While they are being exposed to these stimuli, they will form impressions of the nature of the two distributions. Further, because the distributions overlap, there will inevitably be stimuli from one distribution that are more characteristic of the alternative distribution than they are of the distribution they came from. Thus, variance and uncertainty in the distribution of stimuli can lead to misleading impressions and incorrect discrimination decisions. Of course, the degree to which this occurs depends on the extent to which the distributions overlap. The relationship between d' used in signal detection theory and stimuli used in the current task are as follows: Individuals should correct inaccurate descriptions at a faster pace when d' is high than when d' is low. This is because fewer samples should be needed to detect mean differences in two distributions when the overlap between the two distributions is low compared to when it is high.

Three different manipulations were manipulated in this study: Source Credibility, Stimuli Discriminability, and Description Accuracy. Source credibility refers to the perceived credibility of the source of descriptive information. Stimuli discriminability is defined as the extent to which an observer can discriminate options based on their output (namely, the effect size d of their differences). Finally, description accuracy refers to the veridicality of descriptions of options. Each of these was manipulated in a $2 \times 2 \times 2$ full factorial design. Two primary dependent variables measured were impressions of options after each trial, and choices of options.

Model Simulations and Hypotheses

As a precursor to empirical data collection, a series of simulations were conducted designed to explore predictions from the DEIM. The situation being modeled is one where a participant is presented with two options A and B with uncertain underlying point distributions. The participant's task is to make repeated selections from one of the two distributions in order to accumulate points. To aid the participant, the options are described as having different specific means by a source with either high or low credibility. Namely option A is described as having a distribution with a mean of .60, while option B is described as having a distribution with a mean of .40.

For “inaccurate description” conditions, the descriptions were specifically designed to be the reverse of the veridical means of the two point distributions. This was done to model situations where individuals must combine conflicting information from a description and from experience. In these conditions, option A is normally distributed with a mean of .40, while B is normally distributed with a mean of .60. In “accurate description” conditions, option A has a mean of .60 and option B has a mean of .40

To restrict the range of possible point values between 0 and 1.0, the distributions were truncated at 0 and 1. On each trial, an option was selected according to Equation 4. The impression of the chosen option was then updated in accordance to the DEIM (Equation 3). Each simulation was conducted over 50 choice trials between options A and B. Simulations were conducted using Microsoft Excel software on a Dell PC. The results presented in what follows are the mean impression formation trends collapsed over 500 simulations (participants).

Simulations 1 to 4 – Description Accuracy x Source Credibility Interaction

Four sets of simulations were conducted to determine predictions for the description accuracy by source credibility interaction. Parameter values for each set of simulations are displayed in Table 1. Because the w parameter is intended to weight descriptive information, simulations for low credibility and high credibility sources had values of w set to 1.75 (fast discounting of descriptions) and 0.25 (slow discounting of descriptions) respectively.

Figure 2, Figure 3, Figure 4 and Figure 5 show mean impression and choice trajectories over 50 trials for 500 simulations for Simulations 1, 2, 3 and 4 respectively. The horizontal axis shows the trial number, while the primary (left) vertical axis shows impressions. The secondary (right) vertical axis shows the probability of selecting option A (called Astride in the experiment). Note that in the Accurate Description conditions, A is the best option. Thus, in these conditions, the probability of selecting A tends to increase with experience. In the inaccurate description conditions, A is the worst option. Thus, the probability of selecting A tends to *decrease* with experience. To aid the reader, A will be tagged with an asterix (*) when it is the worst option.

First, consider the effect of source credibility in the accurate description conditions. As seen in Figure 2 and Figure 3, it is clear that source credibility has little to no effect on impressions and choices in these conditions. This is because in accurate description conditions, experiential information is expected to reinforce descriptions. Because this description/experience confirmation occurs regardless of the credibility of the source, source credibility has little effect on impressions and choices.

Next, consider the effect of source credibility in the inaccurate description conditions (Figure 4 and Figure 5). In both of these conditions, experiential information discredited the description on average. Recall that the DEIM predicts that experiential information overcomes descriptions as experiences accumulate; thus, we see that over the course of the 50 trials, impressions converged towards the true values of the options and away from the inaccurate descriptions. In contrast to the effect of source credibility in the accurate description condition, it is clear that in the inaccurate description condition, source credibility had a large effect. Namely, it took significantly longer for inaccurate initial impressions to be corrected in the high source credibility condition than in the low source credibility condition (H1b). Simultaneously, there was a clear difference in choices over time. Namely, participants in the high credibility condition chose option A (the worst option) at a higher rate over the course of the 50 trials than participants in the low credibility condition. This effect on choices was not as strong in the accurate description conditions (H1c). This effect on aggregate choices from the best option are depicted in Figure 6.

Based on these simulations, an interaction is predicted between source credibility and description accuracy. When source credibility is high, participants should place more weight on descriptive information. Consequently, high source credibility is expected to amplify the effects of description accuracy. Under the DEIM, source credibility is expected to affect the description decay parameter w (high source credibility leads to lower values of w , low source credibility leads to higher values of w).

H1a: Estimated values of w in the DEIM will be higher in low source credibility conditions than in high source credibility conditions.

H1b (Impressions): For those given accurate descriptions, there will be no effect of source credibility on impression paths. For those given inaccurate descriptions, there will be better distinction in impressions between the best and worst options over the course of the study for those given descriptions from a source low in credibility than from a source high in credibility.

H1c: (Choices) There will be an interaction between source credibility and description accuracy on aggregate choices across all 50 trials. Participants given descriptions from sources low in credibility will choose the best option at a slightly higher rate when given accurate descriptions versus inaccurate descriptions. Participants given descriptions by sources high in credibility will choose the best option at a much higher rate when given accurate descriptions versus inaccurate descriptions (See Figure 3 for mean predictions).

Simulations 5 to 8 – Stimuli Discriminability $\times$ Description Accuracy

Recall that the parameter c represents the degree to which individuals' sampling decisions correspond to their impressions of the options. As c increases, individuals are more likely to sample from the option that has the highest impression. The main psychological construct that are expected to relate to the parameter c is confidence or trust in one's impressions. That is, if one trusts that his or her impressions are correct, then that person is expected to choose options that he or she has the highest impression of and refrain from choosing options he or she has lower impressions of. In contrast, a person who does not have high confidence in his or her relative impressions of options

may decide that it would be premature to sample from a single option exclusively and choose to be more liberal in his or her choices. In this sense, lack of confidence in one's opinions should spur an information gathering mindset.

In its current form, the model is not equipped to calculate subjective d' values. This is because it does not describe subjective impressions of variability in stimuli – a key component in any discriminability index. Therefore, while it is predicted that changes in d' with regards to the stimuli distributions will affect the parameter c (a claim that will be tested in an experimental study) the model simulation will take this relationship as an assumption and produce data that should be consistent with that assumption:

To explore the interaction between stimuli discriminability and description accuracy, another set of four simulations were generated. Stimuli discriminability was manipulated via the standard deviation of output: lower standard deviations lead to easier discrimination, while higher standard deviations lead to harder discrimination. See Table 2 for stimuli and model parameter values for these simulations.

Figures 7 through 10 present the same dependent variables as simulations 5 through 8. First, consider the effects of stimuli discriminability for accurate descriptions (Simulations 6 and 7 presented in Figure 7 and Figure 8). Mean impression values do not appear to differ substantially between the two conditions. This is to be expected because, as was the case in the previous simulations, when descriptions are accurate, experiences will tend to reinforce initial impressions based on that description. Thus, in both conditions, impressions do not change from their initial values. Stimuli discriminability did however affect choices. In the easy discriminability condition (Simulation 5, Figure

7), participants consistently choose A at a very high rate (above .80) after the 10th trial. This suggests that participants will begin the study in an exploratory sampling mode and then quickly transition to a “point gathering” mode. In the exploratory sampling mode, participants are distributing their choices amongst both options in order to gain more information about their long-term value. Once participants have gained sufficient information to distinguish the two options, they will transition to a “point gathering” mode, where they will focus their choices on the option that is perceived to be the best. In the hard discriminability condition (Simulation 6, Figure 8), participants stay in an exploratory sampling mode longer due to the fact that the experiential information they are seeing does not distinguish the options as well as the easy discriminability condition. As a result, participants choose option A (the best option) at a lower rate, and the increase in this rate over the course of the 50 trials is much slower than in the high discriminability condition.

Next, consider the effects of stimuli discriminability in conditions with inaccurate descriptions (Simulations 7 and 8 in Figure 9 and Figure 10). In these conditions, A is the worst option; thus, it is expected that participants will learn to avoid A as they gain experiences. Comparing the two discriminability conditions, it appears that stimuli discriminability did not have a noticeable effect on the general trajectory of impressions. In both conditions, participants corrected their initial inaccurate impressions at approximately the same rate. Again, this is to be expected because stimuli discriminability does not affect the expectation of experiential information. There does however appear to be a slight effect of discriminability on the variance of impressions at

each trial. In the easy discriminability condition (Simulation 7, Figure 9), the impression lines are very smooth with very little noise. In the hard discriminability condition (Simulation 8, Figure 10), the impression lines appear to contain slightly more noise. This reflects the increased standard deviation of stimuli in this condition.

As in the previous comparison, there was a strong effect of stimuli discriminability on choice. Namely, participants in the easy discriminability condition (Simulation 7, Figure 9) selected the option with the higher impression value at a much higher rate than participants in the hard discriminability condition (Simulation 8, Figure 10). This results in participants making more choices from the best option (B) in the high discriminability condition, than in the low discriminability condition.

Based on these simulations, no interaction is predicted between stimuli discriminability and description accuracy for choices or impressions. However, a main effect is predicted for stimuli discriminability; when stimuli discriminability is high, participants should choose the worst option at a lower rate than when stimuli discriminability is low. Figure 11 presents the predicted choices (from Simulations 7 through 10) for the discriminability x accuracy interaction.

H2a: Estimated values of c in the choice model will be higher in the easy stimuli discriminability conditions than in the hard stimuli discriminability conditions.

H2b: (Choices) There will be main effects for both description accuracy and stimuli discriminability on aggregate choices across all 50 trials. Participants given accurate descriptions will choose the best option at a higher rate than participants given inaccurate description. Participants given stimuli that are easy to discriminate will

choose the best option at a higher rate than participants given stimuli that are difficult to discriminate.

METHOD

To test these hypotheses, 3 independent variables with 2 levels each were manipulated in a 2 (Source Credibility: High vs. low) x 2 (Accuracy: Accurate vs. Inaccurate) x 2 (Stimuli discriminability index: 1.5 vs. .75) design, resulting in 8 experimental conditions. Two key dependent variables were measured at each of 50 trials: participants' estimations of the average value of the chosen option, and the option participants chose on each trial.

Participants

311 Ohio University students enrolled in Psychology 101 participated in the study for course credit. Data from 7 participants were incomplete due to software malfunctions, leaving full data from 304 participants.

Procedure

Participants came to the lab and were greeted by a research assistant. After completing consent forms, participants were told that they would be engaging in a task where they would be taking on the role of a hiring director at company. They were told that their task will be to hire the best employees they can for that company. Participants were then led into individual cubicles to interact with the experiment via Authorware interactive software.

Participants read a series of slides informing them that they would be bringing in candidates for employment at ABC Software Inc., and that their goal was to bring in the

best employees possible. They were informed that on each day, they would select one of two headhunting companies (named Astride and Bracken) that would send a candidate to the company to be interviewed. When a candidate was brought in, he/she would receive an interview score from 0 to 100. Subsequently, participants would be asked to indicate their best guess as to what the average interview rating was for the company they just consulted. Participants were instructed that a knob at the bottom of their screen would keep an up-to-date measure of the average interview score of all candidates they had brought in over the course of their study. They were reminded that their primary goal in the study was to get this measure as high as possible.

After familiarizing themselves with the study protocol, participants received an email from the company president welcoming them to the company and recapping their task. Additionally, the president explained that the past HR director, Janice, had left a recommendation for the participant regarding how good candidates were, on average, from the two headhunting companies. Depending on the credibility condition, participants read one of two descriptions of Janice from the company president:

Welcome to ABC Software Inc.!

As you know by now, your daily task will be to consult with the two headhunting companies we have contracts with (Astride and Bracken) to bring in applicants.

I know that you don't have any experiences with these two companies, so I've asked Janice, our former HR director to give you some guidance on how well each of these companies have done for us in the past.

LOW CREDIBILITY DESCRIPTION

Janice stopped working with these two headhunting companies several years ago. In that time, the companies have each undergone large changes in how they select prospective employees. As a result, I don't think the past performance of the companies will be predictive of their future performance.

HIGH CREDIBILITY DESCRIPTION

Janice has worked closely with these two headhunting companies for several years. In that time she has kept very detailed notes about the quality of applicants coming from each company. Moreover, she is considered a hiring expert by others in her field.

In the end, your primary goal is to bring in employees with the highest interview scores possible regardless of which headhunting company you use.

Participants were then shown Janice's recommendation:

While both companies have at times provided us with good and bad matches to our company, I have noticed some general tendencies.

I estimate that employees from Astride average scores around 60 while those from Bracken average scores around 40.

Based on these estimations, I recommend you focus your recruiting efforts on Astride.

After participants read this description, they were given a comprehension test to ensure that they read the introduction properly. They were asked 1) Which company did Janice recommend, and 2) Was Janice described as a credible source of information or an un-credible source of information? If participants did not answer these two questions

correctly, they were instructed that they did not read the description carefully and were directed back to the letter from the company president. Participants were not allowed to begin the experiment until they had successfully answered the two comprehension questions. Prior to starting the key trials, participants indicated their initial estimations of the average impression scores from each company. Next, participants started their first day (See Figure 12 for a screenshot).

Participants' first task was to select one of the two companies by clicking on the appropriate box. After making their selection, an icon moved from the box to the 'interview' room located in the center of the screen. After a short delay with the word "Interviewing..." followed by an interview score for that candidate was displayed. This interview score was randomly drawn from the selected company's distribution using Authorware's random number generator. Next, a slider bar appeared below the selected company box where participants were asked to indicate how good they think that employees were from that company on average (See Figure 13).

Because interview scores were restricted to the range $[0, 100]$, scores below 0 were set to 0, and scores greater than 100 were set to 100. Technically, restricting the range of scores changed the true means and standard deviations of the distributions, but not enough to change any of the predictions from the model. Finally, the knob on the bottom of the screen indicating the average score of all candidates participants had selected updated to the new value incorporating the newest candidate selected. After 50 trials, participants were given several free-response questions assessing their understanding of the experimental task. Participants were then debriefed and excused.

Outcomes from companies were based on simple random samples drawn from Normal distributions based on parameters determined by the experimental condition (See Table 3).

RESULTS

An error in experiment programming led to the loss of one experimental condition (Credibility = Low, Discriminability = Easy, Accuracy = Inaccurate). These participants were mistakenly given data for the Discriminability = Hard condition and thus contributed data to the Low/Hard/Inaccurate group. Therefore, subsequent analyses on 2 x 2 interactions were done on subsets of the data to ensure that appropriate comparisons between groups were performed. See Table 4 for the labeling of group conditions.

Model Comparisons

First, I compared the descriptive and predictive validity of three different models for impression data. See Table 5 for a listing of the three models pitted against each other. The first model (Model 1) is the full DEIM model with two parameters w and p . The second model (Model 2) is an experience-only model which ignores descriptions in updating impressions. This model is psychologically unreasonable as descriptions are expected to affect impressions. However, it is a good statistical competitor to the DEIM because it contains one fewer parameter. The third model (Model 3) is a special case of Model 1 with parameters w and p forced to be identical. The reason this model was tested was because the w and p parameters were found to be correlated during the parameter estimation iterations for Model 1. Thus, Model 3 was constructed as a parsimonious competitor for Model 1.

For models 1, 2, and 3, I performed 250 simulations using the following criteria: for each of the seven experimental conditions, half of the participants were randomly assigned to the model building group, and the other half were assigned to the model test group. Using participants in the model building groups, best fitting model parameters were calculated using a least-squares non-linear regression procedure. These model parameters were then used to calculate predicted impressions for participants in the model test group (again, in the same experimental condition). Squared errors between predicted impressions and actual impression were calculated at the trial level for each participant and then averaged to compute a MSE statistic. Thus, each simulation computed unique best fitting model parameter values for each of the 7 experimental condition using half of the participants (the model building group), as well as an MSE statistic calculated from the model-test group. This method of model testing is unique in that it penalizes models that ‘overfit’ to a unique set of data by providing a good fit to data used for parameter estimation purposes but poor predictive validity for new data (Dawes & Corrigan, 1974; Myung, 2000). The key results from these simulations are the distributions of MSE values from the model testing groups (See Table 6).

Results from the simulations showed that the full DEIM model did indeed provide the best predictions for new data compared to the other two models. The mean values of the MSE statistic was lowest for the Full DEIM model (.0144) and highest for the Partial DEIM model (.0164), with the Exp Only model falling in between (.0152). Moreover, the standard deviation of the MSE statistic followed the same ordinal pattern (.0047, .0051, and .0057 respectively).

Two key conclusions are made from these results: first, comparing the mean MSE results from the full DEIM model to the Exp Only model, it can be concluded that the w parameter, which is used to weight descriptive information provides a statistical benefit to predictions as well as representing an important theoretical construct. Second, comparing the mean MSE results of the Full DEIM model to the Partial DEIM model (where the parameters w and p are forced to be identical), demonstrates that the w and p parameters can be statistically differentiated despite their algebraic similarity. This provides evidence for the proposed theoretical distinction between the two model parameters.

Moreover, comparing the standard deviation of the mean MSE statistics shows that the ordinal rank of the average predictive accuracy of the 3 models also captures the ordinal rank of the consistency of their predictive accuracy. That is, there was less variation in the predictive accuracy of the Full DEIM model (.0047) compared to the Exp Only model (.0051) and the Partial DEIM model (.0057). When combined with the previous analyses showing that the mean MSE of the full DEIM model is low compared to other models, this provides strong support for the validity of the model and its parameters.

DEIM Parameter Tests

After the DEIM was determined to be the best model for the experimental data, I proceeded to determine the effects of experimental manipulations on the 2 DEIM parameters (w and p) and the choice model parameter c . A non-linear least-squares regression procedure in SPSS was used to estimate DEIM model parameters (w, p)

simultaneously for each participant. Estimates were calculated by minimizing the following error equation:

$$E = \sum_{t=1}^{50} (\hat{Y}_{jot} - Y_{jot})^2 \quad (5)$$

where, $\hat{Y}_{jot}$ = The predicted impression of option o on trial t using the DEIM (Equation 3), Y_{jot} = The observed impression of option o on trial t. Thus, 50 observations were used to fit a unique estimate of the w and p parameters to each participant. Mean differences between experimental conditions in estimates of parameters w and p were computed in a linear regression model.

Model Fits

R^2 values calculated for each participant according to the equation.

$$R^2 = \frac{\sum_{i=1}^{50} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{50} (y_i - \bar{y})^2} \quad (6)$$

18 out of the 302 (6%) of participants had negative R^2 – indicating that the model was not able to fit their data at all. Closer inspection of the data form these participants suggested that they were not using the response scale properly. For example, one participant gave 10 sequential impressions of “100.” For participants with positive R^2 values, the distribution of R^2 values across experimental conditions is presented in Figure 14 . Across participants, the DEIM provided a good fit to the data. The mean R^2 value was .68 and the standard deviation was .237. The majority of participants had R^2 values larger than .60.

The next set of analyses was conducted strictly on the full DEIM model to determine how the experimental manipulations affected its parameters. Specifically, I

examined how the source credibility, source accuracy, and stimuli discriminability manipulations affected the description discounting parameter w and the new information sensitivity parameter p .

Description Discounting Parameter w

A matrix of histograms of estimates of w is presented in Figure 15. Each histogram presents the distribution of estimates of w for participants in an experimental condition. Recall that according to Hypothesis 1, it is expected that as source credibility *increases*, the discounting rate of descriptive information should *decrease*. Because the w parameter was intended to be a measure of the discounting rate of descriptive information, it was hypothesized that, holding other manipulations constant, participants in the high credibility conditions should have lower values of w than participants in the low credibility conditions. The w parameter was not expected to be affected by the other manipulations (accuracy and discriminability). A key observation from Figure 15 is that, for each experimental condition, the distribution of estimates of w is highly skewed to the right. This is consistent with the expected range of w as it is an unbounded parameter. Next, I compared mean estimates of the w parameter across participants in each experimental condition. Figure 16 and Table 7 present mean values of estimates of w for each experimental condition.

I conducted a series of ANOVAs to determine the effects of experimental manipulations on mean estimates of the w parameter. Prior to analysis, estimates of the w parameter were transformed using the transformation w^{10} order to improve the normality assumption of ANOVA. Because of the loss of Condition 2, a full 3-way ANOVA was

deemed inappropriate to test for main effects and interactions. Therefore, to test the effect of predictions, 3 separate 2-way ANOVAs were conducted on balanced conditions. First, a 2-way ANOVA with Accuracy (Accurate vs. Inaccurate) and Credibility (High vs. Low) as factors was conducted on transformed estimates of w for participants in the hard stimuli discriminability conditions (Conditions 3, 4, 7, 8).

The interaction between source credibility and description accuracy was non-significant: $F(1, 174) = 0.018, p > .05$. The main effect for description accuracy was non-significant: $F(1, 174) = 0.943, p > .05$. A moderately significant main effect of source credibility on estimates of w was found, participants given descriptions from sources low in credibility had higher values of w than those given descriptions from sources high in credibility, $F(1, 174) = 3.797, p = .053, \eta^2 = .021$. A Mann-Whitney non-parametric test of the difference in values of w for the credibility conditions was also conducted and found to be significant: the mean rank of w values in the low credibility condition (Mean rank = 99.74) was significantly higher than the mean rank of w values in the high credibility condition (Mean rank = 75.18), $z = -3.102, p < .05$. Importantly, mean estimates of w were not affected by description accuracy, $F(1, 174) = 0.943, p > .05$. These results support Hypothesis 1; namely, the description discounting rate is affected by the credibility of the source of descriptive information, but not the content of the description.

Next, a separate 2-way ANOVA was conducted to determine the effects of description Accuracy (Accurate vs. Inaccurate) and stimuli Discriminability (Easy vs. Hard) on estimates of w for participants in high source credibility conditions (Conditions

5, 6, 7, 8). The interaction between description accuracy and stimuli discriminability was non-significant: $F(1, 138) = 0.036, p > .05$. The main effect for description accuracy was non-significant, $F(1, 138) = 0.696, p > .05$. Additionally, the main effect for stimuli discriminability was non-significant, $F(1, 138) = 0.995, p > .05$. These results suggest that neither description accuracy nor stimuli discriminability had a significant effect on values of the w parameter. Taken together, these results suggest that the w parameter is psychologically tied to the perceived credibility of the source of descriptions, but not to the actual content of the description (measured by description accuracy), or the structure of the experimental stimuli (measured by stimuli discriminability)

New Experiential Information Discounting Parameter p

The next set of analyses focused on the new experiential information discounting parameter p . Again, larger values of p indicate less sensitivity to new information in experiential impression updating. One participant had an extremely high estimate of p (26.27, 46.73 standard deviations above the mean) and was thus removed from the following analyses. A matrix of histograms of the new experiential information discounting parameter p are contained in Figure 17. There was a positive skew in the distribution of estimates. This was not entirely unexpected as the p parameter is unbounded in the same manner as the w parameter. Data from several participants (34) produced estimates of the p parameter that were negative. This could be the result of one of two factors. One cause could have been responding that strongly violated assumptions of the model. For example, consider participant 48 whose data produced an estimate of p equal to $-.82$. On trial 10, this participant indicated that his/her impression of Astride was

.40. After choosing Astride on trial 11 and receiving an interview score of .29, he/she *increased* his/her impression of Astride to .54. Clearly, this was an unexpected impression change that is more likely due to data mis-entry or random responding than a deficiency in the model.

Next, I compared mean estimates of the p parameter across experimental conditions (See Table 8 and Figure 18). A separate set of 2-way ANOVAs were conducted on values of the p parameter. First, a 2-way ANOVA with Accuracy (Accurate vs. Inaccurate) and Credibility (High vs. Low) as factors was conducted on estimates of p for participants in the hard stimuli discriminability conditions (Conditions 3, 4, 7, 8). The interaction between description accuracy and source credibility was non-significant: $F(1, 175) = 0.716, p > .05$. The main effect for description accuracy was non-significant: $F(1, 175) = 1.318, p > .05$. The main effect for source credibility was non-significant: $F(1, 175) = 0.457, p > .05$.

These results suggest that neither source credibility nor description accuracy affected participants' updating of experiential information. To consider the effects of stimuli discriminability and description accuracy on experiential impression updating, a second 2-way ANOVA with Accuracy (Accurate vs. Inaccurate) and Discriminability (Easy vs. Hard) was conducted on estimates of p (for participants in the high source credibility conditions). The interaction between description accuracy and stimuli discriminability was non-significant, $F(1, 138) = 0.975, p > .05$. The main effect for stimuli discriminability was non-significant, $F(1, 138) = 1.016, p > .05$. The main effect for description accuracy was non-significant, $F(1, 138) = 1.576, p > .06$. These results

suggest that neither description accuracy nor stimuli discriminability affected experiential impression updating.

Choice Model

Estimates for the c parameter were then calculated across all participants for each experimental condition. To estimate the parameter c in the choice model, the following error equation was minimized:

$$E = \sum_{j=i}^n \sum_{t=1}^{50} (\hat{Y}_{jt} - Y_{jt})^2 \quad (7)$$

where,

$\hat{Y}_{jt}$ = The predicted probability of choosing Astride for participant j on trial t using equation 4.

$$Y_{jt} = \begin{cases} 0 & \text{if Bracken is chosen by participant } j \text{ on trial } t \\ 1 & \text{if Astride is chosen by participant } j \text{ on trial } t \end{cases}$$

Values of c were estimated for each participant according to Equation 7. Predicted probabilities of selecting Astride were then calculated according to Equation 4. The primary prediction regarding the choice model was that estimates of the c parameter would be higher in the discriminability = easy condition and lower in the discriminability = hard condition (Hypothesis 2). This is because easy of stimuli discriminability is expected to result in greater sensitivity to impressions for choices. Six participants had extreme values of c (less than -16). Thus, their data were excluded from the present analysis. A matrix of histograms of estimates of c are presented in Figure 19. The majority of estimates of c were between 0 and 1, suggesting that participants were much more exploratory with their choices than expected. The next set of analyses were

conducted on mean estimates of the c parameter across participants in each experimental condition. Figure 20 presents mean estimates of c as a function of stimuli discriminability, source credibility, and description accuracy (See Table 9 for descriptive statistics).

A separate set of 2-way ANOVAs were conducted on values of the c parameter. First, a 2-way ANOVA with Accuracy (Accurate vs. Inaccurate) and Credibility (High vs. Low) as factors was conducted on estimates of p for participants in the hard stimuli discriminability conditions (Conditions 3, 4, 7, 8). The interaction between description accuracy and source credibility was non-significant, $F(1, 183) = 1.287, p > .05$. The main effect for description accuracy was non-significant, $F(1, 183) = 1.233, p > .05$. The main effect from source credibility was non-significant, $F(1, 183) = 2.86, p > .05$.

None of the effects were significant, suggesting that neither description accuracy nor source credibility had a significant effect on impression sensitivity in choices. To consider the effects of stimuli discriminability and description accuracy on experiential impression updating, a second 2-way ANOVA was conducted on estimates of c (for participants in the high source credibility conditions). The interaction between description accuracy and stimuli discriminability was non-significant, $F(1, 149) = 0.643, p > .05$. The main effect of description accuracy was non-significant, $F(1, 149) = 1.35, p > .05$. The main effect of stimuli discriminability was also non-significant, $F(1, 149) = 0.027, p > .05$. Therefore, contrary to predictions, there was no effect of stimuli discriminability on impression sensitivity in choices. Thus, Hypothesis 2a was not supported.

Next I assessed the correspondence between choice predictions and actual choices. For each participant, I calculated the probability of selecting Astride on each trial using Equation 4 and the best fitting value of c for that participant. I then assessed the validity of predictions using two methods. First, I created binary choice predictions by classifying predicted probabilities greater than .50 as a prediction of “Astride” and predicted probabilities less than .50 as a prediction for “Bracken.” I then created a contingency table between predicted and actual choices (See Table 10).

On 4.78% of trials, the predicted probability of selecting Astride was .50, thus no prediction could be made. On trials where a prediction was made (95.22%) the model made the correct prediction on 73.48% of trials, and the incorrect prediction on 26.52% of trials. To further illustrate the validity of choice predictions, I split predicted probabilities of choosing Astride into 20 bins with width 5%. I then calculated the observed proportion of choices in each bin. A bar graph depicting the results from this aggregation process is depicted in Figure 21. A Kendal-Tau ordinal correlation between bins and the observed proportions was significant, Kendal-Tau = .947, $p < .05$.

Summary of Model Fit Analyses

A series of model tests suggested that the DEIM provides a good fit for the experimental data. The DEIM produced high R^2 values in fitting data, and was successful in predicting new data. The model was tested against 2 simpler competing models and was found to be superior. More importantly, Hypothesis 1 was supported: the w parameter that was introduced in the DEIM was found to be psychologically (though experimental manipulation) related to the credibility of the source of descriptive information.

Hypothesis 2, the prediction that stimuli discriminability is related to the choice parameter c was not supported. Next, I will explore the aggregate impression and choice predictions resulting from model simulations.

Model-Free Comparisons of Impression and Choice Paths

Source Credibility x Description Accuracy Interaction

Figure 22, Figure 23, Figure 24, and Figure 25 present impression and choice paths for the 2 x 2 interaction between source credibility and description accuracy. (These figures map onto model simulations in Figure 2, Figure 3, Figure 4 and Figure 5 respectively). For impressions, each figure contains both observed data (in solid lines), and model fitted data (dashed lines). For choice proportions, observed data are in light compound lines, and fitted data are in dark compound lines. Choice data are presented in reference to A (the option with the higher descriptive value). As in the prior model simulations, A is the best option in the accurate description conditions, and is the worst option in the inaccurate description conditions.

First, consider Figure 22 and Figure 23 where participants are given accurate descriptions. Figure 22 presents data in the low source credibility condition and Figure 23 presents data in the high source credibility condition. Visual comparison of these figures suggests that source credibility did not have a large effect on impressions or choices. In both conditions, participants maintained accurate impressions of the two options throughout the 50 trials on average. Next, consider Figure 24 and Figure 25 which present data from conditions where participants are given inaccurate descriptions. An unexpected finding here was that participants were much faster in correcting inaccurate

descriptions than anticipated from model simulations. In fact, by the second trial, participants had already reversed their ordinal preferences from A to B. This pattern is consistent with the high model estimates of the w parameter calculated earlier. However, even though participants reversed their ordinal preferences for the two options quite quickly, there did still appear to be a long term effect of source credibility on impressions. In Figure 24 (the low credibility condition), mean impressions of the two companies were more clearly differentiated than in Figure 25 (the high credibility condition). Supporting of Hypothesis 1b, this suggests that participants given false descriptions from sources high in credibility did not distinguish the alternatives as well as participants given false descriptions from sources low in credibility.

The Accuracy x Credibility interaction will now be considered in terms of aggregate choices from the best option. Recall that according to H1c, an interaction is predicted between source credibility and description accuracy on choices. When source credibility is perceived to be low, a small effect of description accuracy on choices is expected. When source credibility is perceived to be high, a large effect of description accuracy on choices is expected.

Figure 26 presents mean choices from the best option (Astride in the accurate description condition, Bracken in the inaccurate description condition) as a function of source credibility and description accuracy (See Table 11 for descriptive statistics). A two-way ANOVA found a significant interaction between source credibility and source accuracy on choices, $F(1, 187) = 7.12$, $p < .05$, $\eta^2 = .037$. Planned contrasts were conducted to compare mean sampling behavior between credibility conditions within

each accuracy condition. Participants who were given inaccurate descriptions chose the best option more if they were told that the description came from a source low in credibility (Condition 4, $M = 32.57$, $SD = 7.69$) than a source high in credibility (Condition 8, $M = 29.24$, $SD = 9.04$), $t(187) = 2.16$, $p < .05$, $d = 0.40$. That is, these participants were able to disregard the description faster when the source of that description was perceived to be lower in credibility. Participants who were given accurate descriptions chose the best option more if they were told that the description came from a source high in credibility (Condition 7, $M = 37.73$, $SD = 6.97$) than a source low in credibility (Condition 3, $M = 34.81$, $SD = 7.22$), $t(187) = 1.654$, $p = .05$, $d = 0.41$. These analyses support the hypothesis that participants were affected by the credibility of the source of descriptions. Specifically, having a source high in credibility amplified the effect of description accuracy.

A contrast was also conducted to determine whether or not there was an effect of description accuracy for participants who received a description from a source low in credibility. That is, when participants are told that a source is low in credibility, will they be affected by the content of that message? To test this question, a contrast was conducted comparing choice behavior in the “low source credibility” condition between those who received an accurate description (Condition 3, $M = 34.81$, $SD = 7.22$) and those who received an inaccurate description (Condition 4, $M = 32.57$, $SD = 7.69$). Results from this contrast suggested that there was no significant difference between choice behavior in these conditions, $t(187) = -1.43$, $p > .05$, $d = 0.30$. This suggests that participants discarded the description of a source perceived to be low in credibility and

were subsequently unaffected by its content. This is in stark contrast to a comparison between choices of participants who were given descriptions from highly credible sources. For these participants, there was a significant difference between participants given accurate descriptions (Condition 7, $M = 37.73$, $SD = 6.97$) and those given inaccurate descriptions (Condition 8, $M = 29.24$, $SD = 9.04$), $t(187) = 4.87$, $p < .05$, $d = 1.07$.

Taken together, these results support the aggregate predictions (H1b and H1c) made from *a priori* simulations of the model regarding the effects of source credibility and description accuracy on impressions and choices.

Description Accuracy x Stimuli Discriminability Interaction

Mean impression and choice data for conditions 5, 6, 7 and 8 are presented in Figure 27, Figure 28, Figure 29 and Figure 30. As in the previous graphs, for impressions, observed data are in solid lines and model fitted data are in dashed lines. For choice proportions, observed data are in light compound lines, and fitted data are in dark compound lines.

Consider Figure 27 and Figure 28 which display data for participants given accurate descriptions for stimuli that are relatively easy and hard to discriminate respectively. These Figures map onto model simulations 5 and 6. Visual inspection of the impression and choice paths for these conditions suggest that stimuli discriminability did not have a large effect on behavior in these conditions. Participants began by matching their impressions to the description and held those impressions on average across all 50 trials. The choice paths appeared very similar in the two conditions as well.

Now consider Figure 29 and Figure 30 which display mean impression and choice data for participants given inaccurate descriptions for stimuli that are relatively easy and hard to discriminate respectively. The effect of stimuli discriminability appears to be larger in these conditions than in the accurate description conditions. That is, with regards to impressions, the average difference between the best (B) and the worst (A) option is greater in the easy discriminability condition than in the hard discriminability condition. With regards to choices, the rate of decrease in choices from the worst option (Astride) appears to be slightly faster in the easy discriminability condition (Figure 29) than in the hard discriminability condition (Figure 30)

Another set of analyses were conducted regarding the interaction between description accuracy and stimuli discriminability on aggregate choices. Recall that Hypothesis 2b predicted two main effects: increased description accuracy and increased ease of stimuli discrimination will lead to more choices from the best option. Again, because of the loss of one of the experimental conditions, the following analyses only apply to participants in the high credibility conditions. Figure 31 presents mean choices from the best option (A in the accurate description condition and B in the inaccurate description condition) as a function of description accuracy and stimuli discriminability (see Table 12 for descriptive statistics).

An unexpected significant interaction between description accuracy and stimuli discriminability on decisions was found, $F(1,152) = 4.11$, $p < .05$, $\eta^2 = .026$. To understand the nature of the interaction, planned comparisons were conducted to establish the effect of description accuracy within each level of stimuli discriminability. For

participants in the easy discriminability condition, there was a significant effect of description accuracy on decisions: those given accurate description (Condition 5, $M = 38.71$, $SD = 7.12$) chose the best option significantly more than those given inaccurate descriptions (Condition 6, $M = 35.17$, $SD = 7.24$), $t(152) = 2.03$, $p < .05$, $d = 0.33$. For participants in the hard discriminability condition, there was a significant effect of description accuracy on decisions: participants given accurate descriptions (Condition 7, $M = 37.73$, $SD = 6.97$) chose the best option more than participants given inaccurate descriptions (Condition 8, $M = 29.24$, $SD = 9.04$), $t(152) = 4.95$, $p < .05$, $d = 1.07$. Taken together, these results suggest that the effect of description accuracy depended on the discriminability of the stimuli: when stimuli were relatively easy to discriminate, the effect of description accuracy was lower than when stimuli were relatively hard to discriminate.

Summary of Model-Free Analyses

Model free analyses of impression and choice paths supported key predictions from model simulations (H1b, H1c, H2b). An interaction on aggregate choices was found between source credibility and description accuracy on impressions and choices: participants were better off getting descriptions from sources high in credibility when descriptions were accurate, but were worse off when descriptions were inaccurate (H1b and H1c). Two main effects for stimuli discriminability and description accuracy were found on aggregate choices (as predicted by H2b); however these effects were qualified by a significant interaction. Namely, the effect of description accuracy on aggregate

choices was smaller when stimuli were easier to discriminate than when they were harder to discriminate.

DISCUSSION

The purpose of this study was to develop, explore, and test a mathematical model (DEIM) of dynamic judgments. This model was derived based on past research on two different fields: decisions from experience and decisions from description. Additionally, it is suggested that this conceptualization may be a more parsimonious explanation of other research findings that have been couched in terms of dual-process theories of cognition (e.g.; cognitive-experiential self theory). The model was designed to make predictions of judgment situations where individuals have access to both descriptive and experiential information. Descriptive information is assumed to come from a source that may vary in its perceived credibility, which affects the extent to which information from that source is weighed in judgments.

Two simpler competing models (each with only one parameter) were pitted against the DEIM (a two-parameter model): an experience-only model which ignores descriptions and a simplified version of the DEIM with only 1 weighting parameter. The DEIM was found to be the better model. Two tests of the models using both data-fitting and cross-validation methods were used to demonstrate that the increased complexity of the DEIM is justified by A) its increased ability to explain variance in model building data sets, and B) its ability to make better predictions of new data in model validation data sets. The fact that the DEIM is a more complicated model than the two competing

models but has better predictive success, suggests that its parameters represent important psychological constructs.

Hypothesis 1a stated that source credibility would affect the w parameter in the DEIM. This prediction had implications for impressions (H1b) and choices (H1c). Each of these predictions were supported. Hypothesis 2 stated that as stimuli discriminability increases (or in other words, as the stimuli become easier to discriminate), the value of the choice confidence parameter c would increase. This model-specific prediction was not supported in model fit analyses. However, aggregate choice predictions based on model simulations of the effect of the c parameter were partially supported in observed aggregate choice data (H2b). This suggests that the predicted effect of stimuli discriminability was theoretically sound, but were not adequately captured in the choice model used in the current study.

Limitations and Directions for Future Research

An unfortunate setback in this study was the loss of one of the experimental conditions. This loss precluded the possibility of detecting a three-way interaction between the three independent variables. While no specific predictions were made regarding this interaction, it would have been informative to see if a three-way interaction between source credibility, description accuracy, and stimuli discriminability was detected. Moreover, if such an interaction truly exists, then the two-way effects found in this study may be moderated by levels of the third variable.

Despite the fact that ordinal model predictions were supported in this study, the small effect size of source credibility was unexpected. In model simulations performed

prior to empirical data collection, it was predicted that inaccurate information from a source high in credibility would cause judges to maintain inaccurate impressions in the face of contradictory information. However, mean empirical results from this study found that individuals were much more sensitive to experiential information (even in the high credibility, inaccurate description group) than expected. This could have happened for many potential reasons. One possibility is that judges simply forgot the descriptions given to them prior to the critical judgment task. Indeed, if a majority of participants truly forgot the description, then neither the content of the description or the credibility of the source would have affected their judgments. However, post-experimental debriefing questions suggest that this was not a problem for this study. The final 69 participants in the study were given a supplementary form asking them to recall Janice's estimation. 68 of the 69 participants correctly recalled that Janice estimated Astride candidates to score 60 on average and Bracken candidates to score 40 on average. Thus, presuming that these 69 participants were not systematically different from other participants, it is highly unlikely that forgetting was a potential cause for the lack of reliance on descriptive information.

Another possibility is that judges are inherently extremely sensitive to recent events and will quickly discard descriptive information without regards to the credibility of its source. In other words, people's judgments are based on what they see and not what people tell them. This explanation is consistent with literature on the availability heuristic which states that judgments are based on the ease with which instances come to mind. Indeed, it is likely a thorn in the side of statisticians and psychologists that no amount of

statistical data can convince some people that it's safer to travel by plane than by car. Therefore, if experiential information is more cognitively accessible than descriptive information in a dynamic judgment task, judgments may be more heavily influenced by experiential information than descriptive information regardless of the credibility of the description's source. However, research on persuasion would suggest that source credibility should have had a large influence on judgments in this task because the strength of the argument given by Janice was ambiguous. Chaiken and Maheswaran, (1994) argued that when the strength of an argument is ambiguous, judgments are based on heuristic processing (specifically, high decision weights on descriptions from sources high in credibility). This conflict in results may be due to difference between the task in the current study and those used by Chaiken and Maheswaran (1994). Chaiken and Maheswaran had participants give attitude judgments towards consumer products (telephone answering machines) described by 8 different attributes. Thus, the decision task used by Chaiken and Maheswaran was certainly much more complex than the one given to participants in the current study where options were described on a single attribute (average interview score). It may be the case that when options are defined by a single attribute, and stimuli information are presented using the same scale as the final judgment, participants believe they are capable of generating their own judgments and do not find it necessary to rely on descriptions.

A separate possibility is that participants interpreted the description from Janice as being only relevant to past data (when she was working at the company) and less relevant to the present state of the two companies. That is, the dynamic nature of the task

may have, at least implicitly, suggested to participants that the quality of candidates from each company was changing over time. To account for this possibility, future studies would need to clearly specify that descriptions given are relevant to both past and future states of the stimuli.

There are a number of theoretical and methodological assumptions behind the DEIM which may present problems for making inferences to other domains. Firstly, it was assumed that the perceived credibility of the source of description does not change over time. That is, individuals who are told a source is credible are assumed to maintain that credibility assessment even in the face of information that contradicts the source's description. This assumption may be false. It is possible that individuals do actively reassess the perceived credibility of a source as a result of new sampling information. If so, then the greater reliance on experiential information over time, relative to descriptive information, predicted by the model may be due to decreased trust in the description rather than increased availability of sampling information. This possibility could be tested using different models that allow for the w weight to change dynamically over time.

A modeling limitation of this study was that experiential impressions were not directly assessed. That is, when participants were asked to give their best estimates of average interview scores, it was assumed that they were providing estimates based on both descriptions and past sampling information (corresponding to Equation 3). Because experience impressions are used in the DEIM but were not measured, an assumption was made that participants were updating their experiences using an explicitly defined anchoring-and-adjustment procedure (Equation 2). This assumption may cause the reader

to question why the both experiential (corresponding to Equation 2) and combined impressions (corresponding to Equation 3) were not assessed separately in the current study. The reason for this is that it was assumed that participants are not be able to cognitively partition the influence of descriptions and sampling experiences on their impressions (see Nisbett and Wilson, 1977). Thus, their responses are presumed to be the result of both types of inputs. Moreover, even if individuals did have some ability to perform this introspective task, having participants give two numerical estimates for the average value of candidates from each company may have led to problematic demand characteristics. That is, asking participants to provide two different estimates based on two different criteria (experience only vs. experience and description), may influence them to conform to the *hypothesis* that the two processes exist and are co-occurring.

In the current analyses, I attempted to rectify this modeling problem by replacing participants' prior experiential impression ($A_{q_{t-1}}$) with the prior empirical sample mean $\bar{x}_{q_{t-1}}$. Thus, I assumed that while participants may idiosyncratically change their experiential impressions of options when presented with new information (as a function of the weight $\frac{1}{t^p}$), they are able to, perhaps implicitly, cognitively maintain the true empirical running average of output from an option. This strong assumption may be false, however, there is some research to suggest that it is not entirely implausible. Recall that Betsch et al.'s (2001) research on individuals' online processing of numerical information found that people can maintain implicit impressions of stocks that reflect their true empirical mean value. If their research translates to participants in the current study, then the modeling assumption I made via necessity may not be entirely implausible.

Another potential solution to this problem in future studies would be to include additional experimental conditions where individuals do not receive a description of the options. Impression data from these individuals would be purely experiential in nature (corresponding to Equation 2), and could be compared to data from individuals who do receive descriptions. Data from individuals in ‘no description’ conditions would also serve as a good test of the psychological interpretation the DEIM. Namely, these data should either be poorly fit by the DEIM or at the very least result in arbitrarily high values of the description discounting parameter w .

CONCLUSION

This study provided evidence for the efficacy of a new model of impression formation. This model successfully described impression data in an experimental task where participants have access to descriptive information and experiential sampling information. The model predicts that the credibility of sources of descriptions as well as the discriminability of stimuli are important predictors of impression change in dynamic situations.

Table 1

<i>Parameters for Simulations 1 Through 4</i>					
		Simulation 1	Simulation 2	Simulation 3	Simulation 4
		(Accurate, Low Cred)	(Accurate, High Cred)	(Inaccurate, Low Cred)	(Inaccurate, High Cred)
Stimuli Parameters	D(A)	0.60	0.60	0.60	0.60
	D(B)	0.40	0.40	0.40	0.40
	μ_A	0.60	0.60	0.40	0.40
	μ_B	0.40	0.40	0.60	0.60
	σ_A	0.20	0.20	0.20	0.20
	σ_B	0.20	0.20	0.20	0.20
<i>Model Parameters</i>	b	0.34	0.34	0.34	0.34
	c	1.00	1.00	1.00	1.00
	w	1.75	0.25	1.75	0.25

Table 2

<i>Parameters for Simulations 5 Through 8</i>					
		Simulation 5	Simulation 6	Simulation 7	Simulation 8
		(Accurate, Easy)	(Accurate, Hard)	(Inaccurate, Easy)	(Inaccurate, Hard)
Stimuli Parameters	D(A)	0.60	0.60	0.60	0.60
	D(B)	0.40	0.40	0.40	0.40
	μ_A	0.60	0.60	0.40	0.40
	μ_B	0.40	0.40	0.60	0.60
	σ_A	0.13	0.26	0.13	0.26
	σ_B	0.13	0.26	0.13	0.26
<i>Model Parameters</i>	b	0.34	0.34	0.34	0.34
	c	1.50	0.50	1.50	0.50
	w	1.00	1.00	1.00	1.00

Table 3

Parameters for point distributions of options in each experimental condition.

Accuracy	Discriminability	Astride		Bracken		Effect size
Condition	Condition					(of A vs. B)
		μ	σ	μ	σ	d
Accurate	Easy	60	13.33	40	13.33	1.50
Accurate	Difficult	60	26.67	40	36.67	0.75
Inaccurate	Easy	40	13.33	60	13.33	-1.50
Inaccurate	Difficult	40	26.67	60	26.67	-0.75

Table 4

Labeling of experimental conditions. Participants in Condition 2 were mistakenly given data for Condition 4.

Credibility	Discriminability	Accuracy	Condition #	N
Low	Easy	Accurate	1	34
Low	Easy	Inaccurate	2 (Lost)	X
Low	Hard	Accurate	3	36
Low	Hard	Inaccurate	4	76
High	Easy	Accurate	5	42
High	Easy	Inaccurate	6	36
High	Hard	Accurate	7	42
High	Hard	Inaccurate	8	38

Table 5

Labels and parameters of 3 models tested

Model	Description	Parameters
1	Full DEIM (Equation 3)	w, p
2	Exp Only (Equation 2)	p
3	Constrained DEIM	$w = p$

Table 6

MSE data from model validation tests. Lower minimum, maximum, mean, and median MSE values indicate better models. Lower standard deviation of MSE indicates more consistent predictions.

Model	Minimum MSE	Max MSE	Mean MSE	Median MSE	Standard Deviation of MSE
1 (Full DEIM)	0.0048	0.0272	0.0144	0.0148	0.0047
2 (Exp Only)	0.0047	0.0292	0.0152	0.0158	0.0051
3 (Partial DEIM)	0.0037	0.0334	0.0164	0.0161	0.0057

Table 7

Parameter estimation results for the description decay weighting parameter w for the full DEIM model.

Condition	Credibility	Stimuli	Accuracy	Mean	Median	SD
Discriminability						
1	Low	Easy	Accurate	8.12	1.48	10.55
2	Low	Easy	Inaccurate	-	-	-
3	Low	Hard	Accurate	5.28	1.23	7.46
4	Low	Hard	Inaccurate	5.67	1.89	7.66
5	High	Easy	Accurate	4.65	1.15	6.57
6	High	Easy	Inaccurate	4.84	1.30	6.53
7	High	Hard	Accurate	3.78	0.43	6.59
8	High	Hard	Inaccurate	4.23	0.94	7.02

Table 8

Parameter estimation results for the new experiential information discounting parameter p for the full DEIM model.

Condition	Credibility	Stimuli	Accuracy	Mean	Median	SD
Discriminability						
1	Low	Easy	Accurate	0.14	0.15	0.31
2	Low	Easy	Inaccurate	-	-	-
3	Low	Hard	Accurate	0.35	0.32	0.39
4	Low	Hard	Inaccurate	0.39	0.41	0.29
5	High	Easy	Accurate	0.36	0.29	0.64
6	High	Easy	Inaccurate	0.52	0.37	0.89
7	High	Hard	Accurate	0.58	0.42	0.80
8	High	Hard	Inaccurate	0.38	0.34	0.40

Table 9

Descriptive statistics for estimates of impression sensitivity choice parameter c .

Condition	Credibility	Stimuli	Accuracy	Mean	Median	SD
Discriminability						
1	Low	Easy	Accurate	0.65	0.46	0.93
2	Low	Easy	Inaccurate	-	-	-
3	Low	Hard	Accurate	0.41	0.43	0.37
4	Low	Hard	Inaccurate	0.42	0.40	0.85
5	High	Easy	Accurate	0.67	0.51	0.91
6	High	Easy	Inaccurate	0.61	0.52	0.69
7	High	Hard	Accurate	0.83	0.49	1.47
8	High	Hard	Inaccurate	0.50	0.51	0.91

Table 10

Predicted versus actual choices.

		Actual Choice		
		Bracken	Astride	Total
Predicted Choice	Bracken	5336 (74.05%)	1870 (25.95%)	7206
	Astride	1528 (27.24%)	4081 (72.76%)	5609
	No prediction	370 (57.45%)	274 (42.55%)	644
Total		7234 (53.75%)	6225 (46.25%)	13459

Table 11

Means and Standard Deviations of Choices From Best Option Over All 50 Trials For Credibility x Accuracy Interaction.

Credibility	Accuracy	Condition	N	Mean	SD
Low	Inaccurate	4	76	32.57	7.69
	Accurate	3	36	34.81	7.22
High	Inaccurate	8	38	29.24	9.04
	Accurate	7	41	37.73	6.97

Table 12

Means and Standard Deviations of Choices From Best Option Over All 50 Trials For Accuracy x Discrimination Interaction.

Accuracy	Discrimination	Condition	N	Mean	SD
Inaccurate	Easy	6	36	35.17	7.24
	Hard	8	38	29.24	9.04
Accurate	Easy	5	41	38.71	7.12
	Hard	7	41	37.73	6.97

Figure 1: Description information weights as a function of trial and w .

Figure 2: Aggregate results from Simulation 1: Accurate descriptions from a source low in credibility.

Figure 3: Aggregate results from Simulation 2: Accurate descriptions from a source high in credibility.

Figure 4: Aggregate results from Simulation 3: Inaccurate descriptions from a source low in credibility.

Figure 5: Aggregate results from Simulation 4: Inaccurate descriptions from a source high in credibility.

Figure 6: Predicted proportion of choices from the best option as a function of source credibility and description accuracy.

Figure 7: Aggregate results from Simulation 5: Accurate descriptions from stimuli that are relatively easy to discriminate.

Figure 8: Aggregate results from Simulation 6: Accurate descriptions from stimuli that are relatively hard to discriminate.

Figure 9: Aggregate results from Simulation 7: Inaccurate descriptions from stimuli that are relatively easy to discriminate.

Figure 10: Aggregate results from Simulation 8: Inaccurate descriptions from stimuli that are relatively hard to discriminate.

Figure 11: Aggregate choices from the best option from simulations 5 through 8.

Figure 12: Screenshot of the beginning of a hypothetical day.

Presentation Window

Bracken

Based on *everything* you know about Bracken, how good do you think that employees are from Bracken on average?

Very Negative
0

45

Very Positive
100

OK

Bracken
Interview Room

Score for This Candidate
53

Average score of all candidates you've selected

0

23

100

Figure 13: Screenshot of a hypothetical participant who had just selected the company Bracken.

Figure 14: Distribution of R2 values for individual participants across all experimental conditions.

Figure 15: Distribution of estimates of w for each participant by experimental condition.

Figure 16: Mean estimates of w parameter by experimental condition.

Figure 17: Distribution of estimates of p for each participant by experimental condition.

Figure 18: Mean estimates of p parameter by experimental condition.

Figure 19: Distribution of estimates of c for each participant by experimental condition.

Figure 20: Mean estimates of c parameter by experimental condition.

Figure 21: Predicted proportion of choices from Astride versus observed proportion of choices from Astride.

Figure 22: Aggregate results of participant data receiving an accurate description of distributions from a source low in credibility.

Figure 23: Aggregate results of participant data receiving an accurate description of distributions from a source high in credibility.

Figure 24: Aggregate results of participant data receiving an inaccurate description of distributions from a source low in credibility.

Figure 25: Aggregate results of participant data receiving an inaccurate description of distributions from a source high in credibility.

Figure 26: Mean choices from the best option as a function of source credibility and description accuracy.

Figure 27: Aggregate results of participant data receiving an accurate description of distributions that were relatively easy to discriminate.

Figure 28: Aggregate results of participant data receiving an accurate description of distributions that were relatively hard to discriminate.

Figure 29: Aggregate results of participant data receiving an inaccurate description of distributions that were relatively easy to discriminate.

Figure 30: Aggregate results of participant data receiving an inaccurate description of distributions that were relatively hard to discriminate.

Figure 31: Mean choices from the best option as a function of description accuracy and stimuli discriminability

REFERENCES

- Anderson, N. (1967). Averaging model analysis of set-size effect in impression formation. *Journal of Experimental Psychology*, 75, 158-165.
- Asch, S. (1946). Forming impressions of personality. *The Journal of Abnormal Psychology and Social Psychology*, 41(3), 258-290.
- Barron, G. & Erev, I. (2003). Small feed-back based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16, 215-233.
- Betsch, T., Plessner, H., Schwieren, C., & Guitg, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin*, 27, 242-253.
- Bond, S. D., Carlson, K. A., Meloy, M. G., Russo, J. E., & Tanner, R. (2007). Information distortion in the evaluation of a single option. *Journal of Organizational Behavior and Human Decision Processes*, 102(2), 240-254.
- Chaiken, S., Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460-473.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81(2), 95-106.
- Denrell, Jerker. & March, James G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, 12, 523-538.

- Denrell, Jerker. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, 112, 951-978.
- Denrell, Jerker. (2007). Adaptive learning and risk taking. *Psychological Review*, 114, 177-187.
- Edwards, W., Lindman, H. & Phillips, L. D. (1965). Emerging technologies for making decisions. T. M. Newcomb, ed. *New Directions in Psychology*, Vol. 2. Holt, Rinehart and Winston, New York, 261-325.
- Fiedler, Klaus. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*. 107, 659-676.
- Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, 39, 1372-1388.
- Hertwig, R., Barron, G., Weber, E. & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534-539.
- Hoeffler, S., Ariely, D. & West, P. (2006). Path dependent preferences: The role of early experience and biased search in preference development. *Organizational Behavior and Human Decision Processes*, 101(2), 215-229.
- Hogarth, R., & Einhorn, H. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1-55.
- Kahneman, D & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 313-327.
- Kahneman, D. & Tversky, A. (1982) The psychology of preferences. *Scientific American*, 246, 160-173.

- Kaplan, M. (1981). Amount of information and polarity of attraction. *Bulletin of the Psychonomic Society*, 18, 23-26.
- Kugler, T., Connolly, T., & Kausel, E. E. (2009). The effect of consequential thinking on Trust Game behavior. *Journal of Behavioral Decision Making*, 22(2), 101-119.
- Macmillan, A. N. & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.
- March, James. (1996). Learning to be risk averse. *Psychological Review*, 103, 309-319.
- Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190-204.
- Nisbett, R. & Wilson, T. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 1977.
- Peterson, C. & DuCharme, W. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology*, 73(1), 61-65.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.
- Simon, D., Krawczyk, D. C., Bleicher, A., & Holyoak, K. K. (2002). The transience of constructed preferences. *Journal of Behavioral Decision Making*, 21, 1-14.
- Slovic, P. (1995). The construction of preference. *American Psychologist*, 50, 364-371.
- Tormala, Z. L, Brinol, P. & Petty, R. (2007). Multiple roles for source credibility under high elaboration. *Social Cognition*, 25(4), 536-552.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

- Tversky, A., & Kahneman, D. (1982). Evidential Impact of Base Rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Weber, E. U., Shafir, S., & Blais, A. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*(2), 430-445.
- Yechiam, E. & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*(3), 387-402.