THE DESIRABILITY BIAS BEYOND DICHOTOMOUS OUTCOMES

A Thesis
by
CASSANDRA L. SMITH

Submitted to the Graduate School
at Appalachian State University
in partial fulfillment of the requirements for the degree of
PSYCHOLOGY MASTER OF ARTS

May 2019
Department of Psychology

THE DESIRABILITY BIAS BEYOND DICHOTOMOUS OUTCOMES

A Thesis
by
CASSANDRA L. SMITH
May 2019

APPROVED BY:

Andrew R. Smith, Ph.D.
Chairperson, Thesis Committee

Andrew E. Monroe, Ph.D.
Member, Thesis Committee

Christopher J. Holden, Ph.D.
Member, Thesis Committee

Rose-Mary Webb, Ph.D.
Chairperson, Department of Psychology

Michael J. McKenzie, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

Abstract

THE DESIRABILITY BIAS BEYOND DICHOTOMOUS OUTCOMES

Cassandra L. Smith
B.S., Appalachian State University
Certificate in Applied Statistics, University of South Carolina
M.A., Appalachian State University

Chairperson: Andrew R. Smith, Ph.D.

The desirability bias, or wishful thinking, refers to the increased likelihood of predicting an outcome due to a preference for that outcome (Krizan & Windschitl, 2007). Research exploring the desirability bias has been limited to dichotomous decisions (e.g., will Team A or Team B win the game?), but many predictions involve more than two possible outcomes (e.g., will runner A, B, C, or D win the race?). In this study, people were asked to make predictions about upcoming events. Specifically, they were shown a multicolor grid of squares and were asked to predict which color they thought the computer would pick at random. Preference for one color versus the other was manipulated using desirable and undesirable point information. Importantly, half of the participants made predictions when there were two possible outcomes (i.e., two different colors of squares in each grid) while the other half made predictions when there were four possible outcomes (i.e., four different square colors). I found that, overall, participants were more likely to predict that desirable outcomes were more likely than undesirable outcomes—that is, they exhibited wishful thinking. Furthermore, participants were as likely to make wishful predictions when there

were two possible outcomes versus when there were four possible outcomes, meaning that the desirability bias did not change as a function of the number of outcomes. This finding extends previous wishful thinking findings by generalizing the results to situations where there are more than two outcomes.

Acknowledgments

I would like to thank my mentor, Dr. Andrew Smith, for his unconditional and continued support throughout my undergraduate and graduate career, especially in regards to this project. I have been especially lucky to have an advisor who has taught and fostered my ability to think critically, solve problems, and learn in an environment that inspires growth, innovation, and progress. I would also like to thank Dr. Andrew Monroe and Dr. Christopher Holden for their feedback and guidance throughout this thesis and my time at Appalachian State University. Additionally, I would like to thank Dr. Twila Wingrove and Dr. Shawn Bergman for their graduate support and willingness to provide additional opportunities to explore academic and professional avenues. Finally, I would like to thank my friends and family for all of their love and support with which I would not be where or who I am today.

Dedication

I dedicate this thesis to my family – Mom, Richard, and Seymour, as well as to my advisor, Andrew.

Table of Contents

List of Tables

List of Figures

The Desirability Bias Beyond Dichotomous Outcomes

Cassandra L. Smith

Appalachian State University

Abstract

The desirability bias, or wishful thinking, refers to the increased likelihood of predicting an outcome due to a preference for that outcome (Krizan & Windschitl, 2007). Research exploring the desirability bias has been limited to dichotomous decisions (e.g., will Team A or Team B win the game?), but many predictions involve more than two possible outcomes (e.g., will runner A, B, C, or D win the race?). In this study, people were asked to make predictions about upcoming events. Specifically, they were shown a multicolor grid of squares and were asked to predict which color they thought the computer would pick at random. Preference for one color versus the other was manipulated using desirable and undesirable point information. Importantly, half of the participants made predictions when there were two possible outcomes (i.e., two different colors of squares in each grid) while the other half made predictions when there were four possible outcomes (i.e., four different square colors). I found that, overall, participants were more likely to predict that desirable outcomes were more likely than undesirable outcomes—that is, they exhibited wishful thinking. Furthermore, participants were as likely to make wishful predictions when there were two possible outcomes versus when there were four possible outcomes, meaning that the desirability bias did not change as a function of the number of outcomes. This finding extends previous wishful thinking findings by generalizing the results to situations where there are more than two outcomes.

The Desirability Bias Beyond Dichotomous Outcomes

Making judgments and decisions are a normal part of everyday life. Most of these decisions can be relatively inconsequential, like whether to schedule a campus tour at 2 P.M. or 4 P.M. However, many decisions can have lasting consequences, like deciding to apply to a certain college based on the quality of that tour. Optimal decision making is useful for long and short term success. It would be difficult to make the best decision about which graduate programs to apply to if the person applying was unable to accurately judge his or her academic ability and subsequent likelihood of acceptance. Although the ability to make informed judgments about events is important, there are many factors that may affect a person's ability to accurately perceive an event. For example, people's predictions can be influenced by the amount of ambiguity surrounding details of the event (Bier & Connell, 1994), loyalty to a group (Morewedge, Tang, & Larrick, 2018; Tang, Morewedge, Larrick, & Klein, 2017), the number and complexity of possible choices (Schwartz, 2004) and a preference for an outcome (Marks, 1951). Understanding how factors influence the types of judgments people make can help predict success across the breadth of different decisions and judgments. For decades researchers have investigated how preference may influence subsequent judgments (Bar-Hillel & Budescu, 1995; Budescu & Bruderman, 1995; Krizan & Windschitl, 2007; Marks, 1951). The desirability bias—or wishful thinking as it is also called— refers to the increased likelihood of predicting an outcome because of a preference for that outcome (for a review, see Krizan & Windschitl, 2007). Similarly, researchers have also investigated how the number of outcomes influences the decision-making process (Windschitl & Chambers, 2004; Windschitl & Wells, 1998). Something that is currently missing from the described bodies of research is the ability to understand how preference for an outcome changes as a function of the number of different possible outcomes, and what this can mean for biased decision making.

The Desirability Bias

The desirability bias occurs when a person's expectation about the likelihood of an outcome is inflated by his or her desire for that outcome (Krizan & Windschitl, 2007). For example, the desirability bias provides one explanation for why a Carolina Panthers fan might be overly optimistic about the Panthers' chances of winning an upcoming game. There are, of course, numerous valid reasons for optimism; the Panthers might be particularly strong, their opponent might be particularly weak, or they are playing at the Panthers' home field. The desirability bias, however, asserts that the simple desire for the Panthers to win can increase people's expectations about a victory independent of the evidential reasons for the expectation. The desirability bias has been demonstrated across many different domains including political elections (Krizan, Miller, & Johar, 2010; Tappin, van der Leer, & McKay, 2017), sporting events (Babad & Katz, 1991; Massey, Simmons, & Armor, 2011), and games of chance (Crandall, Solomon, & Kellaway, 1955; Irwin, 1953; Irwin & Metzger, 1966; Marks, 1951).

Factors Affecting the Desirability Bias

The desirability bias states that preference for an outcome can change a person's expectation about that outcome, but there are a number of ways of asking a person to express his or her expectations. For example, a Carolina Panthers fan might be asked how likely she thinks the Panthers are to win their upcoming game. Or, the fan might be asked whether or not she thinks the Panthers will win. Although the distinction between asking about the likelihood of the event or whether an event will happen might seem minor, it has been shown to influence the magnitude of wishful thinking. In an experimental test of the influence of judgment type on the desirability bias, Windschitl, Smith, Rose, and Krizan (2010) assigned participants to either make a likelihood judgement about a future event or to make an outcome prediction (e.g., X or Y). They found that outcome predictions reliably led to more wishful thinking than likelihood judgments. This may be because likelihood judgments are usually numerically based, and tend to ground people in the objective likelihood of the event (Windschitl et al., 2010). When making an outcome prediction, people have more flexibility in how a conclusion is formed because it can accommodate different likelihood expectations. For example, if someone evaluates the relevant information and concludes that there is between a 40% and 60% chance that there will be sunshine (e.g., from weather channels, phone apps, or from looking outside), and she is then asked to predict the likelihood of sunshine, she is likely to say a percentage in the 40%-60% range. However, if she is asked whether or not it will be sunny outside (an outcome prediction) she may be more likely to lean towards the upper bound of that judgment (60%) and predict the preferred sunny day. If she knows that she wants to spend the day inside, and does not want the day to be sunny, she may be more likely to lean towards the lower bound of that judgment (40%) and predict a cloudy day. For both decisions, the evaluated likelihoods did not vary depending on the desirability of the outcomes (i.e., they were both somewhere between 40%-60%). However, the outcome predictions did vary as a function of the desirability of the outcomes.

The precision of the information provided also influences the desirability bias (Smith, Smith, Windschitl, & Stuart, 2019). In a recent set of studies, the precision of information given to participants regarding the possible outcomes was either relatively precise (e.g., there is a 60% chance the event will occur) or relatively imprecise (e.g., a visual representation of the likelihood information). In these studies, giving people more precise information increased the magnitude of the desirability bias. It is speculated that this surprising "precision effect" may be due to differences in processing mindsets. When information is precise and easy to process, people do not evaluate it as closely; when the information is imprecise and relatively difficult to process, they more thoroughly process (and use) the information.

An important distinction is the difference between uncertainty and ambiguity/imprecision. Namely, uncertainty refers to the variability in the possible outcomes, whereas ambiguity/imprecision refers to the lack of information about the likelihood of the possible outcomes. For example, the chances of rolling an even number on a fair, 6-sided die would involve high uncertainty because there is an equal chance of the outcome being an odd number and an even number. Although there is high uncertainty, the likelihood information is very unambiguous/precise because there is exactly a 50% chance that the die will land on an even number. Research has shown that situations that are more uncertain yield greater degrees of wishful thinking (Krizan & Windschitl, 2009; Windschitl, et al., 2010).

In many situations, the outcome of an event it relatively certain. For example, when people play the lottery, most understand that they will almost certainly lose. Even though winning is desirable, the certainty of the outcome limits the amount of wishful thinking that might occur. Imagine someone predicting which card will be chosen from a deck containing either six black and four red cards, or nine black and one red card. In the former, uncertainly is higher because there is more variability in the outcome. In the latter, the participant can be fairly sure the chosen card was black. The less certainty in a situation, the greater the chance the individual will rely on guessing, which tends to lean towards the preferred option (Windschitl et al., 2010).

The evidence suggests that the desirability bias *can* be moderated by the likelihood of the outcomes and the information about the likelihood of the outcome. The next group of studies investigated if additional characteristics about the outcomes could diminish bias. In an effort to determine if the desirability bias was simply “cheap talk”, researchers began giving incentives for accurate predictions. In an experimental manipulation of accuracy incentives, Simmons and Massey (2012) promised football fans either $5 or $50 for the correct prediction of who would win an upcoming NFL game. The incentive for predicting the correct outcome did not decrease the desirability bias. This suggests that people truly believe in the predictions that they are making.

In addition, Massey et al. (2011) examined whether the desirability bias could be moderated by feedback and experience. In this study, participants were asked to make judgments about the winner of upcoming NFL games throughout a season. Across the 17 weeks, participants were informed of the winner and loser of each game. Despite knowing the actual records of each team, participants were just as optimistic about their preferred team at the end of the season as they were for the first game. This study shows that bias persists even when knowledge is improved.

The desirability bias is the boost in expectations about an outcome when that outcome is preferred. This review has described factors that do and do not influence the magnitude of the desirability bias. Research has shown that while the prediction method and uncertainty of the outcome do moderate bias, accuracy incentives, experience, and feedback do not. One factor that has yet to be investigated is the influence of more than two outcomes on the presence and strength of bias.

Beyond Dichotomous Judgments

Thus far, all of the examples described have shown demonstrations of the desirability bias when there are two possible outcomes (e.g., black or red card, two competing sports teams, and a general election between two party leaders). Research has yet to investigate if, and how, more than two outcomes affects the strength of the desirability bias.

There are two noteworthy reasons why it is useful to understand differences in judgments when there are two or more possible outcomes. First, it makes sense from a practical standpoint. Many real-life decisions involve more than two outcomes. Some of these situations can be low-cost, like people rooting for their own country versus the many other countries during the Olympics. Some of these decisions can have varying degrees of cost, like deciding to bet on red versus black or one of the 38 numbers during a round of roulette. And, some decisions can have lasting and consequential costs, like a home-owner deciding to evacuate based on where a hurricane is likely to hit (e.g., the east coast of North Carolina versus each individual city on the east coast of North Carolina). Understanding how a person perceives the former (two outcomes) versus the latter (more than two outcomes) can help inform policy makers, forecasters, and researchers about how people understand numerical information and probabilities as the number of possible choices/outcomes change.

Second, not only is it important to understand judgments with more than two outcomes in a real-world application, but people tend to engage in non-normative behaviors when more than two outcomes are present. Research has suggested that people have difficulty making a decision when they are overwhelmed by the number of options (Schwartz, 2004). Commonly referred to as ‘overchoice,’ this phenomenon describes the inability to make a decision when there are too many choices, often used in the case of consumer decision-making. Overchoice can refer to the number of options, like choosing a cellular service provider between all options or narrowing it down to a top two. Alternatively, overchoice can reference the complexity of the options (e.g., choosing a cellular service provider based on price, or choosing a cellular service provider based on price, data packages, national coverage, brand recognition, loyalty, and reviews). The complexity of each option may force a person to rely more on alternative factors like availability and preference. Overchoice may lead to less optimal decision-making because of decision fatigue, and may make the person choose the default option, or avoid making a decision at all (Iyengar & Lepper, 2000).

In the case of overchoice, people tend to have a difficult time making a decision as the number and complexity of the options increase. However, in the case of overchoice, it may be difficult to standardize the weight of each of the alternatives because of subjective preference for certain characteristics versus others. In addition, it would be extremely difficult to quantify features that can influence decisions (i.e., intuitive brand preference and recognition about cell phone providers). The next body of research remedies this issue by assigning numerical weights and providing likelihood information about the possible outcomes.

The alternative outcomes effect describes the change in the subjective likelihood of the focal outcome as the distribution of the alternative outcomes change (Windschitl & Wells, 1998). In one demonstration of this effect, Windschitl and Wells (1998) told participants that they would be given 21 raffle tickets for a monetary prize (focal outcome). In addition, the participants were told that there were five other people participating in the raffle, each with a certain number of raffle tickets (alternative outcomes). In one condition, the participant was told that the five competing people held 14, 13, 15, 12, and 13 tickets. In the second condition, the five people held 52, 6, 2, 2, and 5 tickets. The participants were asked how certain they were that their raffle ticket would be chosen on an 11-point verbal scale of certainty (i.e., 0 = impossible to 10 = certain).[1] Two important aspects to note: first, in both conditions the participant making the judgment has 21 raffle tickets. Second, the total number of raffle tickets available (i.e., 81) is the same across conditions. The objective probability of the participant winning both raffles remains the same (i.e., 21/81). Analyses revealed that participants were significantly more certain of their raffle ticket being chosen in the first condition (21-14-13-15-12-13) compared to the second (21-52-6-2-2-5). This effect is thought to occur because people typically do not calculate the objective likelihood of an event occurring when there are more than two outcomes. Instead, they engage in pairwise comparisons by comparing the focal event (in this case, the participant winning the raffle) to each alternative outcome in turn. In the first condition (21-14-13-15-12-13) the focal outcome compares relatively favorably to *all* of the alternatives. In the second condition (21-52-6-2-2-5) the focal outcome does not compare favorably to the first alternative, but compares relatively favorably in the remaining four alternatives. In the first raffle, there are five favorable comparisons, and in the second raffle there are only four. Therefore, people tend to be more confident they will win the first raffle as compared to the second.

Ideally, a decision-maker should lower his or her expectation about the likelihood of an event as more options are added. My probability estimate of the occurrence of one number in a ten-sided die should be lower than for a six-sided die. Although people typically understand this basic rule of probability theory, there are times where this rule is violated. In a study examining the influence of adding low-probability outcomes to a set, Windschitl and Chambers (2004) asked participants to judge the likelihood of winning a number of different raffles. In the first raffle (the baseline version), the participant held 39 tickets and the competitor held 31 tickets. The second raffle introduced “duds” (dud-present version), meaning that the participant held 39 tickets, and of three competitors, one held 31 tickets, one held 2 tickets, and one held 3 tickets. Objectively, the chances of winning the first raffle are greater than the second (i.e., 39/70 versus 39/75). However, participants judged the likelihood of winning the second raffle as significantly higher than the first raffle. The explanation for this dud-alternative effect, like the alternative outcomes effect described above, relies on the fact that participants often engage in pairwise comparisons. In the baseline version (39-31) the focal outcome compares somewhat favorably against the alternative. In the dud-present version (39-31-2-3), the focal outcome compares somewhat favorably against the first alternative, but *very* favorably against the second and third alternatives. These latter two comparisons increase people’s confidence when duds are present, although this violates the objective probability that the focal event will occur.

What is clear from the studies described above is that people do not always evaluate the likelihood of an event accurately or objectively, especially when there are more than two possible outcomes. Because the decisions made are not based solely on probability, people must be using a non-normative strategy to structure their judgments. There are a number of strategies a person might use to evaluate the likelihood of a particular event. Imagine, for example, that John is tasked with predicting whether or not he will win an upcoming raffle.

John is told that he holds 13 raffle tickets. He is also told that Susan holds 10 tickets, Marcus holds 4 tickets, and Jennifer holds 3 tickets. Using a normative strategy, John would count up the number of tickets he holds (13) and divide by the total number of tickets (30) to calculate his objective likelihood. As mentioned earlier, an alternative (and empirically supported) strategy is to make pairwise comparisons by evaluating his chances as compared to all of his competitors—Susan (13 versus 10), Marcus (13 versus 4), and Jennifer (13 versus 3). Finally, he might use a "number of options" strategy and simply focus on the fact that there are four possible outcomes (i.e., four possible people who can win the raffle) and he is one of those four. The most important takeaway is that expectations about the likelihood of an outcome can change depending on the strategy a person uses.

Furthermore, people may use different strategies to solve problems. Although using a likelihood strategy is the most normative, it can sometimes be the most cognitively taxing. Someone with a worse understanding of numbers and probabilities may have a difficult time evaluating their chances as the likelihood information becomes more complex (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Cokely, Ghazal, Galesic, Garcia-Retamero, & Schulz, 2013; Cokely, Ghazal, & Garcia-Retamero, 2014; Newall, 2016; Reyna, Nelson, Han, & Dieckman, 2009; Steen, 1990). Alternatively, using a 'number of options' approach would be the least cognitively taxing because the decision-maker is ignoring the provided numerical information. Using a pairwise comparison approach may be a happy medium between these approaches, such that people pay attention to the likelihood information, but use it in a way to simplify the prediction. Most importantly, the type of strategy a person uses may change as a function of their motivation to come to a certain conclusion (Kunda, 1990).

Current Study

First, people tend to expect and predict a certain outcome when they have a preference for that outcome (i.e., a desirability bias; Marks, 1951; for a review, see Krizan & Windschitl, 2007). Second, people's subjective understanding of the likelihood information can change depending on how the possible outcomes are structured (Windschitl & Chambers, 2004; Windschitl & Wells, 1998). Last, motivation may inform how these two processes interact to influence the bias and accuracy in upcoming predictions (Kunda, 1990). Is decision-making strategy firm throughout a task, or is decision strategy flexible and can change depending on an individual's case by case motivation to conclude a certain outcome?

The current study investigated the relationship between the number of outcomes and the desirability bias. Desirability bias research is clear, but it lacks the incorporation of more than two outcomes. Research investigating the distribution of the possible outcomes has illustrated the non-normative nature of decision making when the number of outcomes increase or the distribution of the outcomes change. The proposed study will bridge the gap in research by investigating both the desirability bias and the number of outcomes.

In this study, participants predicted the outcome of uncertain events. Specifically, they were shown a grid of 300 colored squares and asked to predict which color the computer would choose at random. The objective likelihood of the colors was displayed through the frequency of the colors of the squares. That is, people had a sense of the distribution but did not know the exact proportions. Some of the grids had two colors, meaning there are two possible outcomes (dichotomous condition) and some of the grids had four colors meaning there are four possible outcomes (polychotomous condition). Desirability was manipulated by associating some colors with winning points and some with losing points.

My first prediction was that participants would be sensitive to the proportions of each of the colors. Specifically, the participants would be more likely to predict the color associated with a higher frequency of squares. Second, there would be evidence of a desirability bias; people would be more likely to predict a color when it is desirable (i.e., associated with winning points) than when it is undesirable (i.e., associated with losing points). Third, that there would be a greater desirability bias in situations that are more uncertain. Specifically, people would be more likely to predict the desirable versus undesirable color when there are 50% versus 25% or 75% of the squares present. This is consistent with past research demonstrating a stronger desirability bias with greater uncertainty (Krizan & Windschitl, 2007; Windschitl, et al., 2010). The fourth and final prediction investigated the novel demonstration of both the desirability bias and the alternative outcomes effect. I predicted that the desirability bias would be stronger for those in the polychotomous condition compared to the dichotomous condition. The rationale for this prediction is that there are more strategies available to people when making predictions about events with more than two outcomes (with dichotomous outcomes, a normative strategy and pairwise comparison strategy lead to the same conclusions). This added flexibility when evaluating polychotomous outcomes may increase the influence of desire on people's predictions.

Method

This study was preregistered and all materials, data, and syntax are posted on the Open Science Framework (https://osf.io/m9sjv/). Furthermore, I report how I determined the sample size, all manipulations, and all measures in the study.

Participants

Based on pilot testing and previous demonstrations of this paradigm, I had a target sample size of 200 participants (i.e., 100 per between subjects condition). I was able to recruit 230 (86.1% Female, M_{age} = 19.19, SD_{age} = 1.43) participants through the Appalachian State University Psychology Subject Pool. The Psychology Subject Pool consists of students enrolled in introductory and intermediate Psychology classes who have elected to enter the Psychology Subject Pool to fulfill an Experiential Learning Credit (ELC) for the course. In addition, participants were given candy for their involvement. Based on the average correlation between repeated measures (r = .395) and a sample size of 230, a sensitivity power analysis indicated a 90% chance of detecting a small-moderate effect (Cohen's d = 0.36) for the critical Outcomes X Desirability interaction (Faul, Erdfelder, Buchner, & Lang, 2009). Appalachian State University's Institutional Review Board (IRB) determined this study to be exempt from IRB oversight (see Appendix A).

Design

The study used a 2 (desirability condition: critical color is desirable vs. critical color is not desirable) x 5 (percentages condition: 25%, 40%, 50%, 60%, 75%) x 2 (outcome condition: dichotomous outcomes vs. polychotomous outcomes) mixed measures design. The outcome condition was manipulated between-subjects, while the percentages condition and desirability condition were manipulated within-subjects.

Procedure

Each participant who met the age and subject pool requirements (18+ and enrolled in an Appalachian State University Psychology course) was invited to take part in the study. After arriving to the lab, the participants were given an informed consent document (see

Appendix B) that briefly stated the purpose of the research study, the risks, an explanation that participation is voluntary, and the contact information of the Principal Investigator. Participants were also informed that participation will take no more than 30 minutes, and will satisfy 1 ELC requirement. After reading the consent form, participants were given the chance to ask any questions. Once answered, participants were instructed to complete the rest of the study on the provided computers.

The participants read instructions about their task. Specifically, they were informed that their task is to predict which color the computer will pick at random from a grid of 300 colored squares. Instructions explained that the participant's prediction will in no way influence which color the computer picks. Before making their prediction, participants received two pieces of information: (1) the desirability point information and (2) the grid information. First, desirability point information informed the participants that one color is associated with either winning or losing points. In this thesis, this color was referred to as the "critical color." For example, a participant might be told that if the computer randomly picks a red square, they will win 100 points. The other colors present were not associated with either winning or losing points (i.e., winning/losing 0 points). This information was available for six seconds. Second, the desirability point information remained on the screen and the grid of 300 colored squares appeared. Participants assigned to the dichotomous outcome condition were shown a grid with two colors and participants assigned to the polychotomous outcome condition were shown a grid with four colors (see Figure 1 for a visual representation of the desirability and grid information). For both outcome conditions, the choice is between the critical color and the non-critical color(s). The experimental program is designed so that for the dichotomous condition, the computer will choose two colors from a bank of four different four-color combinations. The four-color combinations were used for the polychotomous condition. This assures that participants in both outcome conditions will view similar color combinations. The grid and point information were shown for six seconds. The grid and point information then disappeared, and the participants were shown a decision screen where they were asked to predict if the computer would choose the critical color or the non-critical color(s). For the dichotomous condition, participants will predict one of the two colors (i.e., the critical or the non-critical color). For the polychotomous condition, the three remaining non-critical colors were consolidated into one option, meaning they will predict between the critical color and all three non-critical colors, essentially making a dichotomous decision. For a visual representation of the outcome prediction, see Figure 2. Participants were not given feedback as to whether or not their prediction matched what the computer actually picked. After a practice trial where the program explains each step of the experiment, the participants were free to begin.

The participants had the opportunity to win points in one of two ways. First, they can win points if the computer chooses the critical color on a round where the critical color is associated with winning points (they will lose 100 points if for that round the critical color is associated with losing points). Point allocation is independent of the participant's prediction (i.e., if the critical color was associated with losing points, and the participant predicted the non-critical color, if the computer chose the critical color, they would still lose 100 points despite their prediction). Second, they can win 25 points if they make an accurate prediction (i.e., predicting the same color the computer chooses). Participants were told that more points indicated a greater candy prize at the end of the study.

Across each round, two features of the grid changed. First, for half the rounds, the critical color was associated with winning 100 points (making that outcome more desirable) and for the other half the rounds, the critical color was associated with losing 100 points (making that outcome less desirable). Second, the percentage of the critical color was randomized across each trial. The possible percentages include 25%, 40%, 50%, 60%, and 75%.[2] For the dichotomous condition, the second color will comprise the remainder of the squares (e.g., if the percentage of the critical color is 25%, the percentage of the non-critical color is 75%). For the polychotomous condition, the probability of the critical color was identical to its dichotomous counterpart. The only difference is that the *three* non-critical colors comprised the remainder of the squares equally. For example, if the percentage of the critical color is 25%, the percentage of the three non-critical colors would be 25% each, totaling 75%.

There are two different desirability conditions (desirable and undesirable) and five different percentage conditions (25%, 40%, 50%, 60%, and 75%). Therefore, there were 10 possible combinations of the within-subjects desirability and percentages conditions. The order of these 10 combinations were randomized within two blocks of 10 rounds each—for a total of 20 rounds.

After completing the 20 rounds, participants completed individual difference measures (i.e., Life Orientation Test – Revised, Rational Experiential Inventory – 40, and the Abbreviated Numeracy Scale), answered demographic questions (i.e., age and gender), debriefed, credited for their participation in the study, and asked to leave.

Exploratory Measures

In addition to the prediction task (described in the Procedure section), participants completed three exploratory individual difference measures.

Life orientation test – revised. The Life Orientation Test – Revised (LOT-R) was developed to assess variability in generalized optimism (Scheier, Carver, & Bridges, 1994). It is a 10-item scale consisting of six questions assessing optimism/pessimism, and four filler items. The scoring criteria is based on a 5-point Likert scale: 1 = I disagree a lot, 2 = I disagree a little, 3 = I neither agree nor disagree, 4 = I agree a little, and 5 = I agree a lot. The LOT-R is the revised version of the Life Orientation Test (LOT) which contained 12 items (Scheier & Carver, 1992). Higher scores on the LOT-R indicate higher levels of global optimism. Analyses of validity assessed the unidimensional pessimism-optimism spectrum using item response theory (IRT) and found that the LOT-R satisfactorily measured the latent trait of dispositional optimism (Chiesi, Galli, Primi, Borgi, & Bonacchi, 2013). In addition, several studies have confirmed the reliability and validity of the LOT-R using measures of internal consistency and test-retest reliability, construct and predictive validity, and significant correlations to psychological well-being, sense of mastery, and sense of coherence (for a review, Carver, Scheier, Miller, & Fulford, 2009). For a document detailing the instructions, questions, and reverse-scored items, see Appendix C.

Rational experiential inventory. The Rational Experiential Inventory – 40 (REI-40) is used to measure preferences for information processing (Pacini & Epstein, 1999). The REI-40 distinguishes between two different processing methods: Need for Cognition (NFC) which is characterized as a rational style and emphasizes a logical and analytical approach, and Faith in Intuition (FI) which is characterized as an experiential style and emphasizes a pre-conscious, affective, holistic approach. The REI-40 is a 40-item scale with four subscales, two for each of the two processing styles, with each subscale containing 10 questions. The four subscales include rational ability (RA), rational engagement (RE), experiential ability (EA), and experiential engagement (EE). The questions for each subscale include items regarding the ability to think logically, enjoyment of thinking logically, ability to respect one's intuitive feelings, and the enjoyment of relying on this intuition, respectively. The 40 items are scored using 5-point ratings from 1 (definitely not true of myself) to 5 (definitely true of myself). The REI-40 was investigated using confirmatory factor analysis (Björklund & Bäckström, 2008). Analysis confirmed the hypothesized model and indicated that experientiality and rationality were orthogonal traits, and that adequate correlations with other personality traits were reached in order to determine acceptable convergent and discriminant validity. For a document detailing the given instructions questions, and reverse-scored items, see Appendix D.

Abbreviated numeracy scale. Numeracy refers to the ability to use mathematics and statistics in everyday life. In other words, someone's numerical "literacy" (National Numeracy, 2014). Being numerate means being able to reason with numbers in a way that extends beyond an academic setting, and to incorporate numerical reasoning in new, everyday problems. In general, numerate people have the ability to: interpret data and graphs, process information, solve problems, check answers, understand and explain solutions, and make decisions based on logical thinking. Those higher in numeracy have been shown to have superior skills in evaluating medical treatment options and risk, political information, and financial products and investments (Cokely et al., 2012; Cokely et al., 2013; Cokely et al., 2014; Newall, 2016; Reyna et al., 2009; Steen, 1990). Numeracy may refer to both mathematical and statistical numeracy. Statistical numeracy describes real-world reasoning with numerical problems whereas mathematical numeracy describes the ability to visualize and understand graphical or numerical displays of information (Cokely et al., 2018). Statistical numeracy, or practical probabilistic reasoning has the unique ability to predict decision-making in naturalistic, real-world choices such as HIV prevention, cardiovascular risk, medical judgments, policy evaluations, and weather vulnerability (Cokely et al., 2014). In addition, numeracy can also predict resistance to theoretical and logical decision-making errors such as framing effects, sunk costs, inconsistent risk evaluation, overconfidence, and norming (among others; Cokely et al., 2014).

The abbreviated numeracy scale (ANS) was developed to assess numeracy (Weller et al., 2013). The ANS is an 8-item scale that consists of eight different numerical and word problems for the participant to solve. The ANS is superior to other numeracy measures (e.g., the Berlin Numeracy Test; Cokely et al. 2012) because of its broader range of difficulty, generalizability, and prediction of decision-making and risk behavior (Weller et al., 2013). The scoring criteria is dichotomous, meaning that the participant either gets the answer correct or incorrect. Higher scores on the ANS indicate higher levels of numeracy. This scale was chosen because of the probabilistic nature of the experimental paradigm. For a document detailing the questions and answers of the ANS, see Appendix E.

Results

Confirmatory Analyses

For each participant, the number of times he or she predicted that the computer would pick the critical color was measured. The critical color is the color associated with either winning or losing points. Next, I conducted a 2 (outcome: dichotomous vs. polychotomous) x 5 (percentages: 25%, 40%, 50%, 60%, 75%) x 2 (desirability: critical color is desirable [winning 100 points] vs. the critical color is not desirable [losing 100 points]) mixed ANOVA on the number of times people predicted the critical color. The outcome condition was manipulated between-subjects while percentages and desirability were manipulated within-subjects.

There was a significant desirability bias, $F(1, 228) = 147.66$, $p < .001$, $\eta_p^2 = .393$. Participants were more likely to predict that the computer would choose the critical color if it was associated with winning points (66%) rather than losing points (41%). There was also a main effect of percentages, $F(4, 912) = 289.59$, $p < .001$, $\eta_p^2 = .559$. As the proportion of the critical color increased, participants were more likely to predict that the computer would randomly select that critical color. And, in line with the previously-discussed findings from the relationship between uncertainty and desirability (Windschitl et al., 2010), there was an interaction between desirability and percentages, $F(4, 912) = 8.53$, $p < .001$, $\eta_p^2 = .036$. Participants exhibited a greater desirability bias when more uncertainty was present (when the color proportions were 50/50) and less of a desirability bias when less uncertainty was present (e.g., when the proportions were 25/75).

Interestingly, there was a main effect of number of outcomes, $F(1, 228) = 7.26$, $p = .008$, $\eta_p^2 = .031$. Participants were slightly more likely to predict the critical color in the polychotomous condition (55.0%) compared to the dichotomous condition (51.4%; see Figure 3).

As a test of the novel hypothesis that there would be a larger desirability bias in the polychotomous condition than the dichotomous condition, I examined the interaction between desirability and number of outcomes. This interaction was *not* significant,

$F(1, 228) = 0.07$, $p = .794$, $\eta_p^2 < .001$. This means that my hypothesis was not supported and people were equally biased in both outcome conditions. In other words, there was not a greater desirability bias in the polychotomous condition compared to the dichotomous condition (see Figure 4).

Finally, the Percentages X Outcomes interaction was not significant, $F(4, 912) = 2.21$, $p = .068$, $\eta_p^2 = .010$, and the Desirability X Percentages X Outcomes interaction was not significant, $F(4, 912) = 1.83$, $p = .122$, $\eta_p^2 = .008$.

In sum, the key result from the study was that, although there was a significant desirability bias, this bias was *not* moderated by the number of outcomes.

Exploratory Analyses

Accuracy in predictions. For each participant, I measured the number of times he or she predicted the color associated with the greater proportion of squares—i.e., made the optimal prediction. For example, regardless of whether a particular color was desirable or undesirable, predicting the color with 60% of the squares in the grid would be the optimal prediction. The 50% percentage condition was omitted from the optimal prediction calculation because no prediction is better or worse than the other (i.e., both option is equally likely). Next, I conducted a 2 (outcome: dichotomous vs. polychotomous) x 4 (percentages: 25%, 40%, 60%, 75%) x 2 (desirability: critical color is desirable [winning 100 points] vs. the critical color is not desirable [losing 100 points]) mixed ANOVA on the number of times people predicted the optimal color. The outcome condition was manipulated between-subjects while percentages and desirability were manipulated within-subjects. Analyses revealed a main effect of percentages $F(3, 684) = 36.70$, $p < .001$, $\eta_p^2 = .139$, such that for the 40% percentages condition, participants made the fewest optimal decisions ($M = 66.2\%$)

compared to the 25% condition (*M* = 80.4%), 60% condition (*M* = 74.2%), and the 75% condition (*M* = 83.8%). All mean differences were significant at the $p < .05$ level. See Figure 5 for a graph representing the optimal decision analyses.

There was a Desirability X Outcomes interaction, $F(1, 228) = 5.20$, $p = .024$, $\eta_p^2 = .022$. Specifically, when the critical color was undesirable, participants in both the dichotomous (*M* = 76.4%) and polychotomous (*M* = 76.6%) conditions made a relatively equal number of optimal decisions. However, when the critical color was desirable, those in the dichotomous condition (*M* = 78.3%) made more optimal decisions than those in the polychotomous condition (*M* = 73.3%). There was also a Percentages X Outcomes interaction, $F(3, 684) = 6.36$, $p < .001$, $\eta_p^2 = .027$, such that for the 25%, 60%, and 75% percentages conditions, optimal decision making for both outcome conditions was relatively equal. The only significant mean difference was between the two outcome conditions in the 40% percentages condition. Those in the dichotomous outcome condition (*M* = 72.1%) made significantly more optimal decisions in the 40% grid compared to the polychotomous outcome condition (*M* = 60.3%). Finally, there was a Desirability X Percentages interaction, $F(3, 684) = 70.47$, $p < .001$, $\eta_p^2 = .236$, such that when the critical color was desirable, participants made more optimal decisions in the 60% and 75% conditions, and when the critical color was undesirable, participants made more optimal decisions in the 25% and 40% conditions.

There was no main effect of desirability, $F(1, 228) = 0.40$, $p = .530$, $\eta_p^2 = .002$ or number of outcomes $F(1, 228) = 1.23$, $p = .268$, $\eta_p^2 = .005$. There was also no Desirability X Percentages X Outcomes interaction, $F(3, 684) = 0.49$, $p = .692$, $\eta_p^2 = .002$.

Overall, people made better decisions when the outcome was more certain. For undesirable outcomes and all grids besides the 40% (i.e., 25%, 60%, 75%), outcome condition did not affect decision making. However, for desirable outcomes and for the 40% grid, decision making was worse for those in the polychotomous compared to the dichotomous condition.

Individual differences. Although not a main concern of the thesis, I was interested in how certain individual differences may relate to each other. For each participant, I calculated a "wishful thinking index." This was calculated by subtracting the percentage of times the participant predicted the computer would pick the undesirable color from the percentage of times the participant predicted the desirable color. Higher numbers indicated a greater degree of wishful thinking. Wishful thinking indexes could range from -1 (predicted the undesirable color every time) to 1 (predicted the desirable color every time). I also created an "optimal decisions index" which measured the proportion of times the participant predicted the "optimal color". Optimal indexes could range from 0 (never predicted the optimal color) to 1 (predicted the optimal color every time). My exploratory analyses investigated correlations using the wishful thinking index, the optimal decisions index, the Abbreviated Numeracy Scale, the Rational-Experiential Inventory (i.e., Rational Ability, Rational Engagement, Experiential Ability, and Experiential Engagement), the Life-Orientation Test (i.e., global optimism), gender, and age.

In order to assess relative associations between individual differences I ran bivariate correlations between the main variables of interest (i.e., wishful thinking, numeracy, and optimal decision making). A correlation table detailing all of the bivariate correlations between individual differences can be found in Table 1. While interesting, variables that were not significantly related to task-generated indices (i.e., the wishful thinking index and the optimal decisions index) are not discussed in detail.

Wishful thinking was negatively related to optimal decision making, $r(228) = -.37$, $p < .001$, and numeracy $r(228) = -.13$, $p = .044$. In addition, numeracy and optimal decision making were positively related $r(228) = .22$, $p < .001$. These associations have important implications. First, they suggest that there are consequences to engaging in a preference-based bias, evidenced by the lower number of optimal predictions. In addition, it offers some insight into *why* people may be more biased by their preferences. The negative relationship between numeracy and wishful thinking suggests that it may be an issue with mathematical and statistical literacy and the inability to accurately judge probabilities and likelihood information. This decreased ability to understand numerical information may lead people to rely more on other factors. Other factors in this case include a preference for one outcome over others. It is also possible, of course, that a third variable accounts for the relationship between wishful thinking and numeracy.

Discussion

This study investigated how the number of possible outcomes influences the desirability bias. As noted earlier, the central hypothesis to this thesis—that the increase in the number of outcomes would increase the desirability bias—was not supported. Participants were as likely to make preference-based biased decisions in the two and four color grid paradigm. Although my hypothesis was not supported, it is worth noting that participants' sensitivity to the proportion of the colors indicated that they understood the task. Participants were also more likely to predict the desirable color, indicating that they exhibited a desirability bias. Finally, this desirability bias was moderated by the proportion of colors, such that bias was the greatest when the outcome was more uncertain. These three findings are consistent with previous research on wishful thinking (for a review see Krizan & Windschitl, 2007).

Implications

Confirmatory analyses. First the results of my study increase the generalizability for desirability bias research. As discussed in the introduction, I am aware of no research investigating the desirability bias where the paradigm compares decisions where there are more than two outcomes. The lack of research in preference-based decision making could have had serious costs if predictions changed as a function of the number and distribution of the outcomes. There have been numerous studies investigating wishful thinking in political elections (Krizan et al., 2010; Tappin et al., 2017). All of these studies asked people to make predictions about the general election, but political decision making can extend far beyond this. Think of presidential primaries, where the number and strength of potential party candidates can change day to day. When aiming to make accurate forecasts about potential front-runners, what if voter prediction strategy changed as frequently as the makeup of the potential candidates? This thesis shows that the influence of desire on expectations about the outcomes is likely to be similar when there are two versus many possible outcomes.

Second, the increase in generalizability helps contextualize past research, and also allows future researchers to continue to use paradigms with two outcomes without concern that the number of outcomes may change the results. If, for example, the number of outcomes moderated wishful thinking, research using only two outcomes would be limited. Now, it is known that the number of outcomes likely does not influence wishful thinking.

Therefore, future research can use existing paradigms without fear of only capturing one aspect of wishful thinking.

The findings of my study also show support for the alternative outcomes effect. Participants in the four-outcome grid were overall more likely to predict the critical color than participants in the two-outcome grid. If the participants were more likely to predict the color when there were more outcomes, it could be because they perceived that color to be more likely than it actually was. This is in line with previous research where researchers found a boost in confidence, expectations, and predictions about a focal outcome when the number of outcomes increased (Windschitl & Chambers, 2004; Windschitl & Wells, 1998). This finding provides additional support that when there are more outcomes, people engage in non-normative prediction strategies. On one hand, if participants were using a likelihood strategy, those who saw the two color grid would predict the critical color just as often as those in the four color grid condition because the percentage of the focal outcome was held constant across outcome conditions. On the other hand, if participants were using a pairwise strategy, people who saw the two color grid would predict the critical color less often as compared to the four color grid condition. This would happen because of the artificial inflation of the critical color's likelihood when the alternatives are considered separately (e.g., making a prediction about a 40/60 grid compared to a 40/20/20/20 grid). Because the number of predictions was higher for those in the four outcome condition, it seems likely that people were using a pairwise strategy when predicting the color the computer would pick.

This study found that the number of outcomes did not influence the amount of wishful thinking. It is possible, of course, that there were features specific to this study that contributed to this finding. For example, figuring out the likelihood of the outcomes is relatively easy because of the grids that were used. Also, for the polychotomous grid, all of the non-critical colors are grouped. Such that for the point information, the grid information, and the outcome prediction, the three non-critical colors were right on top of each other, and the participant was making a prediction between the critical color and *all* of the non-critical colors together. Subsequently, this was a relatively easy task that did not involve a lot of ambiguity or numerical reasoning. If the task was more difficult (e.g., the alternatives were not organized to make relevant comparisons easier) then I might have found different results.

Finally, there was a huge effect of wishful thinking ($\eta_p^2 = .393$) in both outcome conditions. Because of this, it is possible that there were ceiling effects. Specifically, it was hypothesized that there would be *more* wishful thinking in the polychotomous condition, meaning that in the desirable conditions, participants would be more likely to predict the desirable color (and conversely less likely to predict the undesirable color). But, if there was already a large wishful thinking effect in the dichotomous condition, it is unlikely that this specific task could identify an increase in bias. Future demonstrations could experiment with different ways of asking for predictions. Research has shown that the desirability bias tends to be elusive when the individual is asked to make a likelihood judgments about the outcomes. Maybe, it is not there is no difference in bias between the outcomes, but the way the prediction was measured left no room to detect a difference.

Exploratory analyses. The exploratory analyses revealed interesting patterns in regards to optimal decision making. First, the effect of percentages indicates that participants were better at making optimal decisions when there was greater certainty. That is, participants made better decisions when there were either 25% or 75% critical squares versus the less certain 40% and 60%. The more uncertain percentage conditions can allow for more flexibility in the prediction because the "correct" answer is less clear. When the provided proportions can reasonably go in either direction when the outcome is more uncertain, people might feel freer to make a guess-like prediction that is consistent with what they desire, which may lead to sub-optimal decision making.

Although the degree of a desirability bias was equal across outcome conditions, the number of optimal decisions was not. When the optimal color was undesirable, those in both outcome conditions made a relatively equal amount of optimal decisions. However, when the critical color was desirable, those in the polychotomous conditions made worse decisions. Follow up analyses indicated that the biggest discrepancy in optimal decisions was in the 40% polychotomous grid. Confirmatory analyses revealed that this was not necessarily due to an increase or decrease in wishful thinking, but may be due to the overall greater likelihood of those in the polychotomous condition predicting the critical color, especially in the 40% condition. What is it about the 40% grid that elicits the biggest differences between the outcome conditions? Previously in this thesis, I described how when there are more than two outcomes, people tend to make pairwise comparisons between the focal outcome and each of the alternatives. Of course, when there are only two outcomes people can certainly still make pairwise comparisons, but these judgments would not differ from an objective likelihood approach. Illustrated in Figure 6, the 40% condition is the only grid where two of the proposed decision strategies (i.e., objective likelihood and pairwise comparisons) would yield different predictions. For example, in the 60% outcome condition, those in the dichotomous condition would compare a focal 60% to an alternative 40%. Conversely, those in the polychotomous condition would compare a focal 60% to alternatives 13.33%, 13.33%, and 13.33%. For both outcome conditions, the "most likely" color would be the critical color. Therefore, whether people use a pairwise or likelihood strategy, they will make the same prediction. However, for the 40% condition, those in the dichotomous condition would compare a focal 40% to an alternative 60%, and those in the polychotomous condition would compare a focal 40% to alternatives 20%, 20%, and 20%. If people use a pairwise strategy in this condition, the “most likely” color in the dichotomous condition would be the non-critical color (the alternative), and for the polychotomous condition, the “most likely” color would be the critical color (the focal). This may help better illustrate the alternative outcomes effect and how it can affect not only prediction strategy, but subsequent accuracy.

When considering the optimal decision making results of this study, it may seem intuitive that a strong desirability bias would automatically indicate an equally strong decrease in accuracy. However, bias by itself does not always lead to worse decision making. If a participant was guessing throughout the task, there would be low levels of *both* bias and accuracy. Conversely, it is possible to be biased but just as accurate. For example, imagine someone predicting the results of a coin flip. If he or she predicts heads every time, that would indicate a bias, but would not necessarily affect accuracy. This is not what my study showed, however. My study showed that biased decision making is also related to poorer decision making (illustrated through the negative correlation between optimal decision making and wishful thinking). The stakes of this paradigm remained relatively low, but the implications of trading good decision making for preference can have consequences ranging from disappointing (losing a bet) to deadly (a doctor choosing a treatment based on subjective preference for that treatment). The exploratory measure of numeracy may possibly explain this inability to make unbiased decisions. The negative relationship between numeracy and wishful thinking means that those who are highly numerate tend to be less biased. This may be because participants who understand probabilities can rely on the numeric information to make decisions, and those who have a difficult time with numeric reasoning may rely on alternative factors (e.g., intuitive preference for an outcome). An interesting next step would be to help improve the numeracy of those making predictions about visual and/or numeric probabilities. This may help to establish a causal link and determine if there are ways to actively debias people.

Limitations and Future Directions

One limitation of this study is the artificial nature of the study. This paradigm was not asking people to predict the outcome of an election (i.e., a real-world task), but instead had participants predicting which square from a grid that a computer would pick at random. That said, the paradigm used allowed me to run a tightly controlled study. In other words, I traded external validity for internal validity.

One benefit of the paradigm was that I could control for factors other than desirability that might influence people's decisions. For example, when making a prediction about an upcoming sports game involving a preferred team, the decision-maker is not only necessarily concerned with accuracy. Predicting a preferred team to win can be the product of multiple factors, including loyalty, team location, rivalries, or identification as fan. Most of the time, people are unwilling to bet against a preferred team (Morewedge et al., 2018). While a real-world example would have been less artificial, it was my goal in this study to isolate the influence of desire, outcome likelihood, and the number of outcomes, something that would be impossible in a real-world prediction task. Furthermore, it would be impossible to randomly assign desire in a real-world task (e.g., it would be impossible to randomly assign half the participants to hope Donald Trump is reelected to the presidency and the other half assigned to hope Trump is not reelected). That said, a future demonstration could remedy this issue by using a slightly more real-world application, but in a context where the uncertainty could be controlled. For example, using a task like the raffle studies that were used in the alternative outcomes research (Windschitl & Wells, 1998).

Another limitation is that this study used one specific method of displaying the likelihood of the outcomes. There are a variety of ways to represent probabilities and uncertainty, ranging from probabilistic and precise (e.g., 40%), to unorganized, visual, and ambiguous (e.g., a scrambled grid of colors representing 40%). Because of the different ways of representing uncertainty, and the moderate correlation between numeracy and wishful thinking, it may be beneficial to explore if other representations of numerical likelihoods would yield an effect where mine did not.

Conclusion

This was the first study to demonstrate that the typical wishful thinking effect found in dichotomous predictions stays true for polychotomous predictions. Although a significant bias was detected, it is fortunate that this bias was not exacerbated by the number of outcomes. The results of this thesis indicate that decision strategy may be more a product of the task and structure of the outcomes, rather than trial by trial situational motivation. It seems as though how people make decisions (e.g., likelihood or pairwise) may be less flexible than originally hypothesized, for better or for worse.

References

Babad, E., & Katz, Y. (1991). Wishful thinking—against all odds. *Journal of Applied Social Psychology*, *21*, 1921–1938.

Bar-Hillel, M., & Budescu, D. V. (1995). The elusive wishful thinking effect. *Thinking & Reasoning*, *1,* 71–104.

Bier, V. M., & Connell, B. L. (1994). Ambiguity seeking in multi-attribute decisions: Effects of optimism and message framing. *Journal of Behavioral Decision Making*, *7,* 169–182.

Björklund, F., & Bäckström, M. (2008). Individual differences in processing styles: validity of the Rational-Experiential Inventory. *Scandinavian Journal of Psychology*, *49,* 439–446.

Budescu, D. V., & Bruderman, M. (1995). The relationship between the illusion of control and the desirability bias. *Journal of Behavioral Decision Making*, *8,* 109–125.

Carver, C. S., Scheier, M. F., Miller, C. J., & Fulford, D. (2009). Optimism. In S. J. Lopez & C. R. Snyder (Eds.), Oxford hand-book of positive psychology (2nd ed., pp. 303-311). New York, NY: Oxford University Press.

Chiesi, F., Galli, S., Primi, C., Borgi, P. I., & Bonacchi, A. (2013). The accuracy of the Life Orientation Test-Revised (LOT-R) in measuring dispositional optimism: Evidence from item response theory analyses. *Journal of Personality Assessment*, *95,* 523–529.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7,* 25–47.

Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R., (2018). Decision making skill: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (2nd eds.), *Cambridge handbook of expertise and expert performance*. New York, NY: Cambridge University Press.

Cokely, E. T., Ghazal, S., Galesic, M., Garcia- Retamero, R., & Schulz, E. (2013). How to measure risk comprehension in educated samples. In R. Garcia-Retamero & M. Galesic (eds.), *Transparent communication of health risks: Overcoming cultural differences* (pp. 29–52). New York: Springer.

Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Measuring numeracy. In B. L. Anderson & J. Schulkin (eds.), *Numerical reasoning in judgments and decision making about health* (pp. 11–38). Cambridge University Press.

Crandall, V. J., Solomon, D., & Kellaway, R. (1955). Expectancy statements and decision times as functions of objective probabilities and reinforcement values. *Journal of Personality, 24,* 192-203.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality, 21*, 329-335.

Irwin, F. W., & Metzger, M. J. (1966). Effects of probabilistic independent outcomes upon predictions. *Psychonomic Science*, *5*, 79-80.

Iyengar, S., & Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995-1006.

Krizan, Z., Miller, J. C., & Johar, O. (2010). Wishful thinking in the 2008 U.S. Presidential election. *Psychological Science, 21*, 140-146.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*, 95–121.

Krizan, Z., & Windschitl, P. D. (2009). Wishful thinking about the future: Does desire impact optimism? *Social and Personality Psychology Compass*, *33*, 227–243.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Marks, R. W. (1951). The effect of probability, desirability, and ‘‘privilege” on the stated expectations of children. *Journal of Personality*, *19*, 332–351.

Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, *22*, 274–281.

Morewedge, C. K., Tang, S., & Larrick, R. P. (2018). Betting your favorite to win: Costly reluctance to hedge desired outcomes. *Management Science, 64*, 997–1014.

National Numeracy. (2014). What is Numeracy? Retrieved from https://www.nationalnumeracy.org.uk/what-numeracy

Newall, P. W. (2016). Downside financial risk is misunderstood. *Judgment and Decision Making, 11*, 416–422.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972-987.

Reyna, V. F., Nelson, W. L., Han P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135*, 943-973.

Scheier, M. F., & Carver, C. S. (1992). Effects of optimism on psychological and physical well-being: Theoretical overview and empirical update. *Cognitive Therapy and Research*, *16*, 201–228.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67,* 1063-1078.

Schwartz, B. (2004). The paradox of choice: Why more is less. New York, NY, US: HarperCollins Publishers.

Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology: General, 141,* 630-634.

Smith, C. L., Smith, A. R., Windschitl, P. D., Stuart, J. (2019). *Wishful predictions from precise versus imprecise information.* Manuscript submitted for publication.

Steen, L. A. (ed.) (1990). *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academies Press.

Tang, S., Morewedge, C. M., Larrick, R. P., & Klein, J. G. (2017). Disloyalty aversion: Greater reluctance to bet against close others than the self. *Organizational Behavior and Human Decision Processes*, *140,* 1–13.

Tappin, B. M., van der Leer, L., & McKay, R. T. (2017). The heart trumps the head: Desirability bias in political belief revision. *Journal of Experimental Psychology: General*, *146*, 1143–1149.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J. and Peters, E. (2013). Development and testing of an Abbreviated Numeracy Scale: A Rasch analysis approach. *Journal of Behavioral Decision-Making*, *26*, 198-212.

Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*, 198-215.

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes*, *111*, 33–47.

Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology*, *75*, 1411–1423.

Table 1

Descriptive statistics and relationships among participant characteristics.

	Mean (SD)	Optimal Index	Numeracy	Optimism	Rational Ability	Rational Engagement	Experiential Ability	Experiential Engagement	Age	Gender
WT Index	0.25(0.31)	-.374**	-.133*	.008	.035	-.02	.144*	.103	-.07	-.069
Optimal Index	0.76(0.17)	-	.219**	-.038	.043	.106	.014	-.062	.064	-.067
Numeracy	3.31(1.59)		-	.085	.233**	.219**	-.02	-.072	-.044	-.157*
Optimism	3.10(0.83)			-	.260**	.115	.305**	.111	.029	.009
Rational Ability	4.50(0.58)				-	.571**	.198**	-.041	-.101	.001
Rational Engagement	4.43(0.58)					-	.065	.009	.005	-.099
Experiential Ability	4.23(0.58)						-	.575**	-.033	.062
Experiential Engagement	4.54(0.52)							-	-.049	.139*
Age	19.19(1.43)								-	-.055
Gender	86% females 13% males									-

Note: * p < .05, ** p < .01, *** p < .001; WT index coded on a -1 – 1 scale; Optimal Index coded on a 0 – 1 scale; Numeracy coded on a 0 – 8 scale; Optimism, Rational Ability, Rational Engagement, Experiential Ability, and Experiential Engagement coded on a 1-5 point scale; All values are r-values. For values listed for a relationship with gender, positive correlations indicate higher levels for females, and negative correlations indicate higher levels for males.

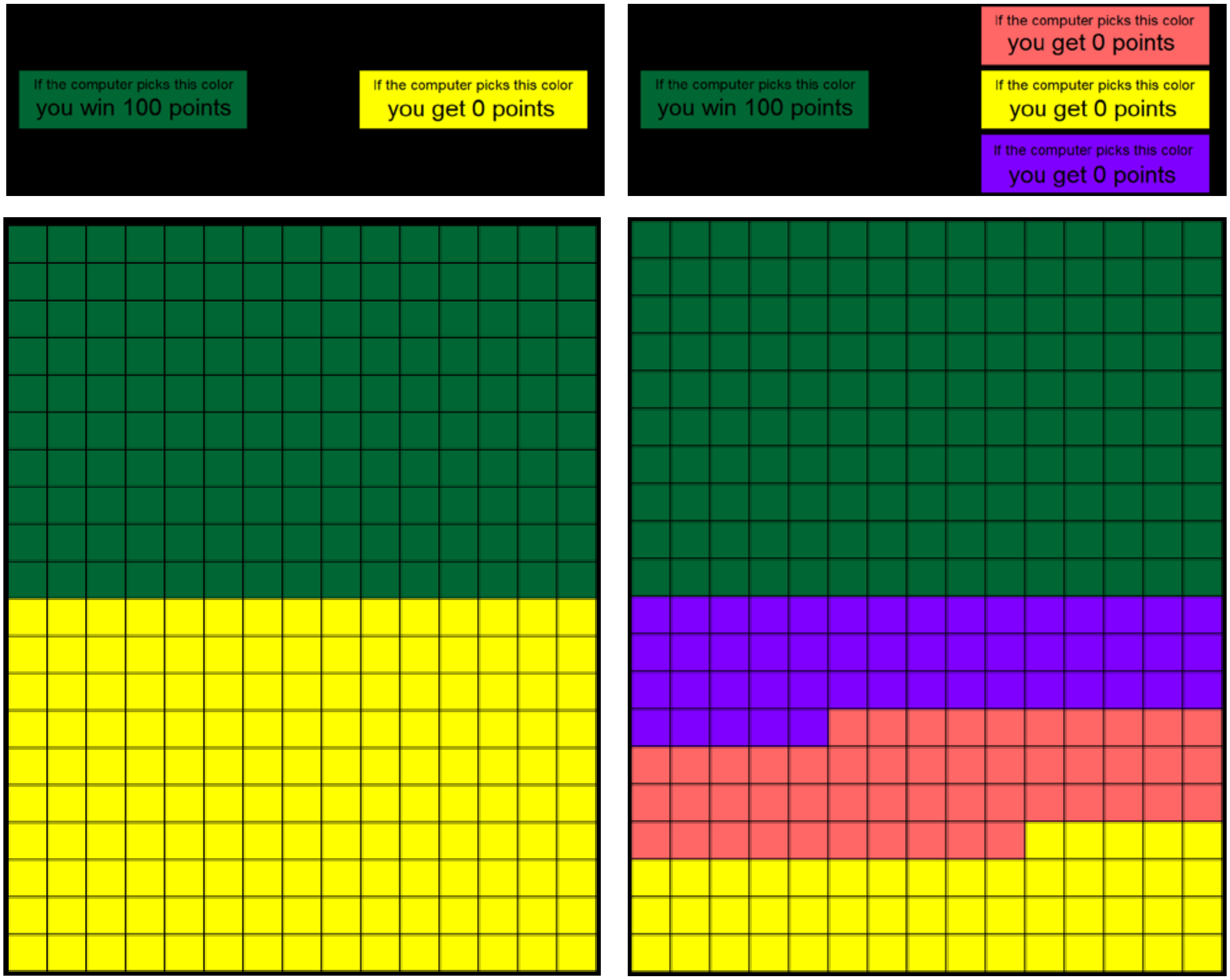

Figure 1. A side-by-side comparison of the dichotomous outcome condition (left) and the polychotomous outcome condition (right). In this trial, the percentage of the critical color is 50%, and the critical color is desirable because it is associated with winning points. The participants used this information to make a prediction on the next screen.

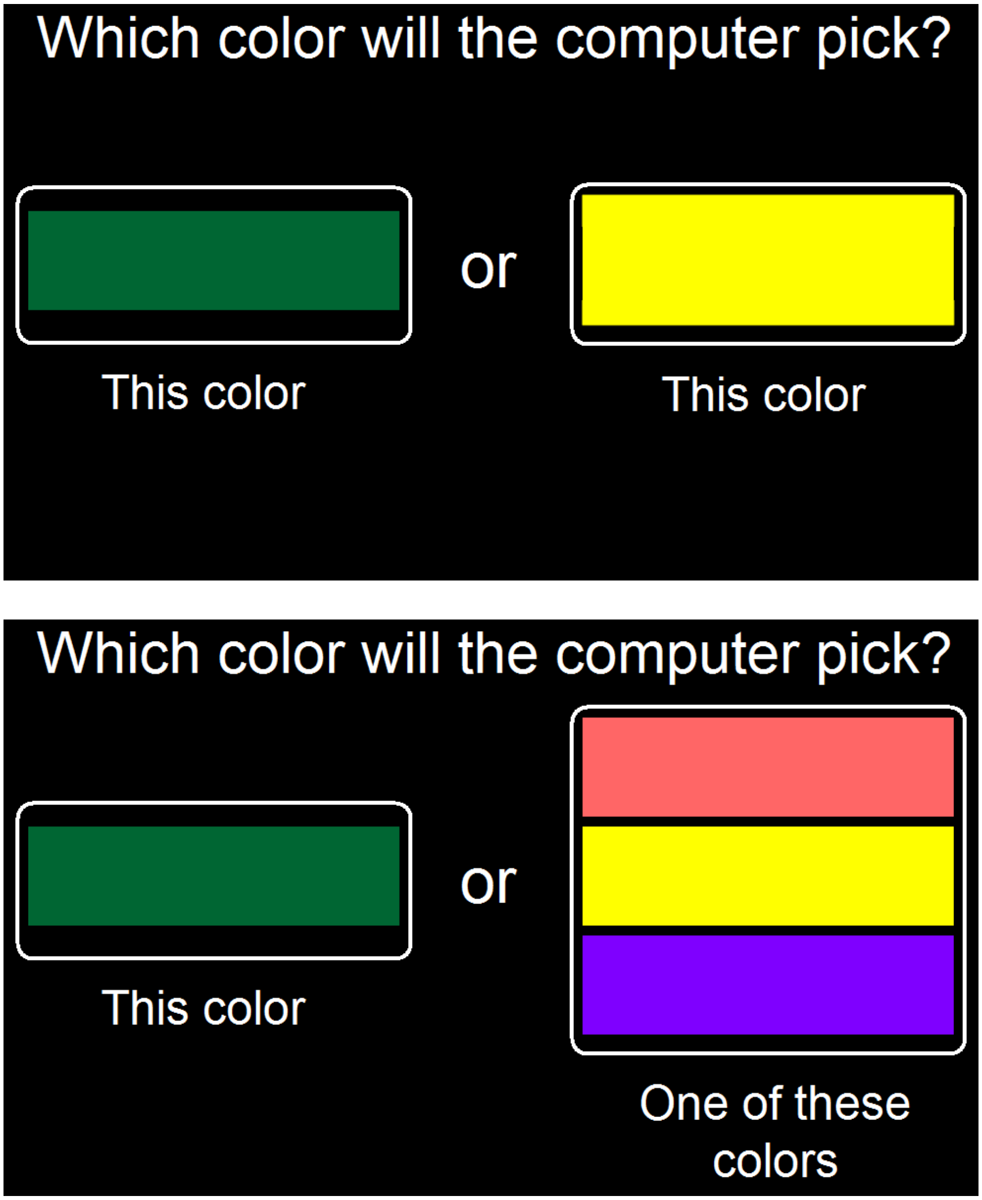

Figure 2. The top pictures shows the outcome prediction that participants in the dichotomous condition made. The bottom pictures shows the outcome prediction that participants in the polychotomous condition made. This information was available until the participants made a prediction.

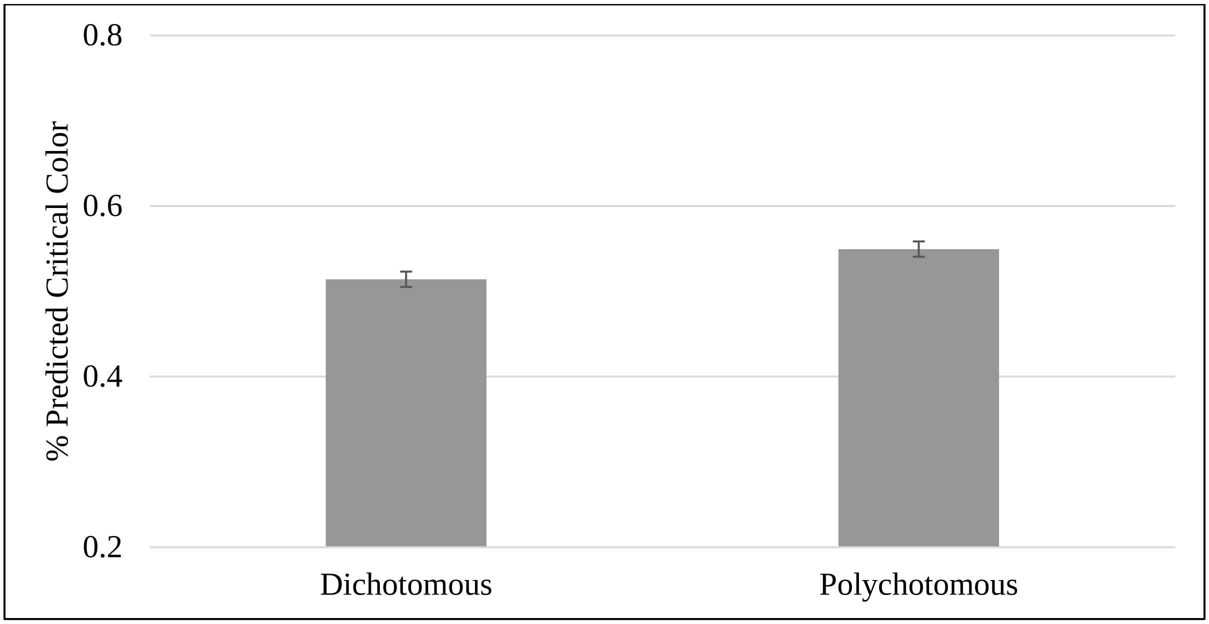

Figure 3. The number of trials where participants predicted the critical color across the two outcome conditions.

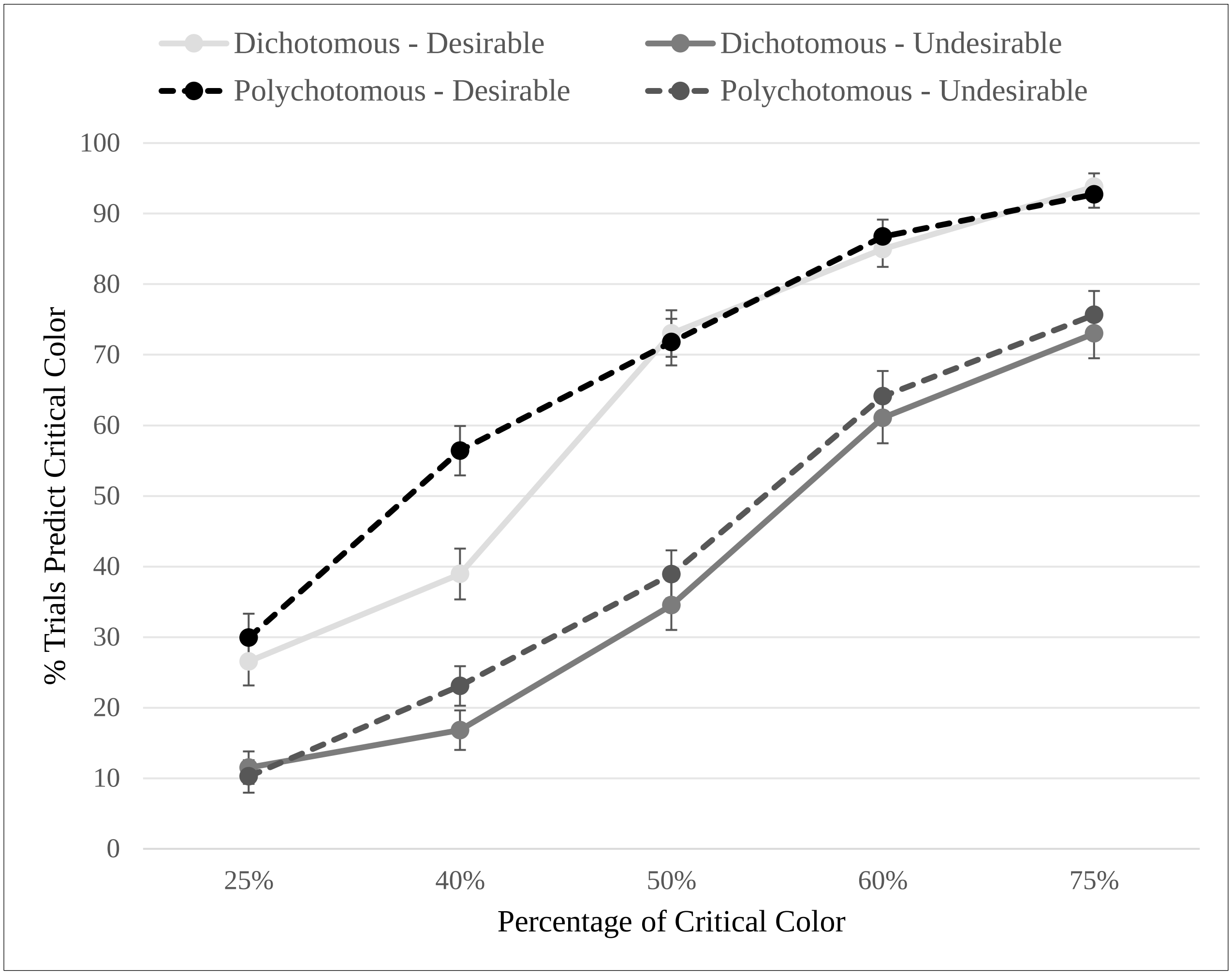

Figure 4. The percentage of trials where participants predicted the critical color across the different percentage conditions for each of the desirability and outcome conditions.

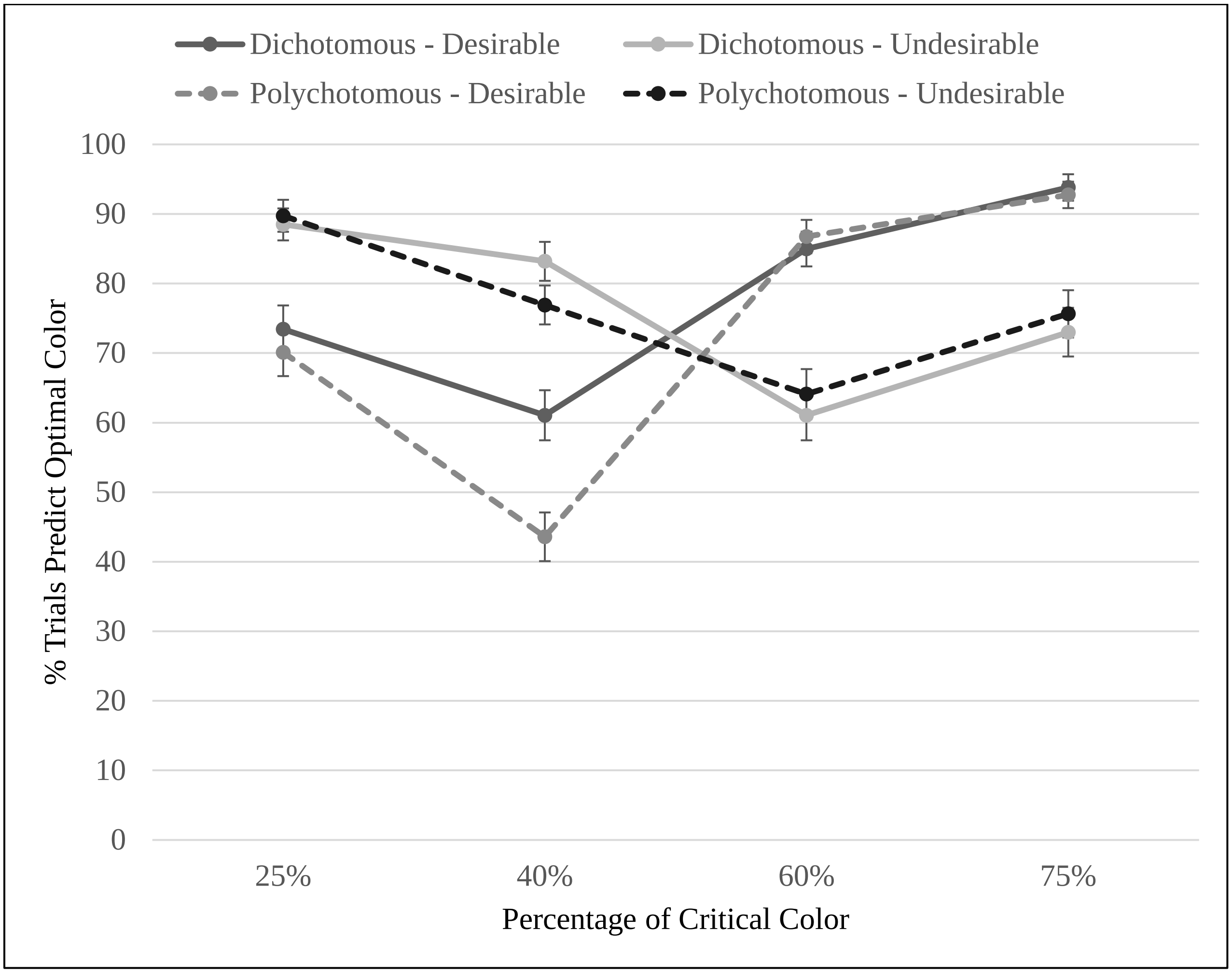

Figure 5. The percentage of trials where participants predicted the optimal color across the different percentage conditions for each of the desirability and outcome conditions.

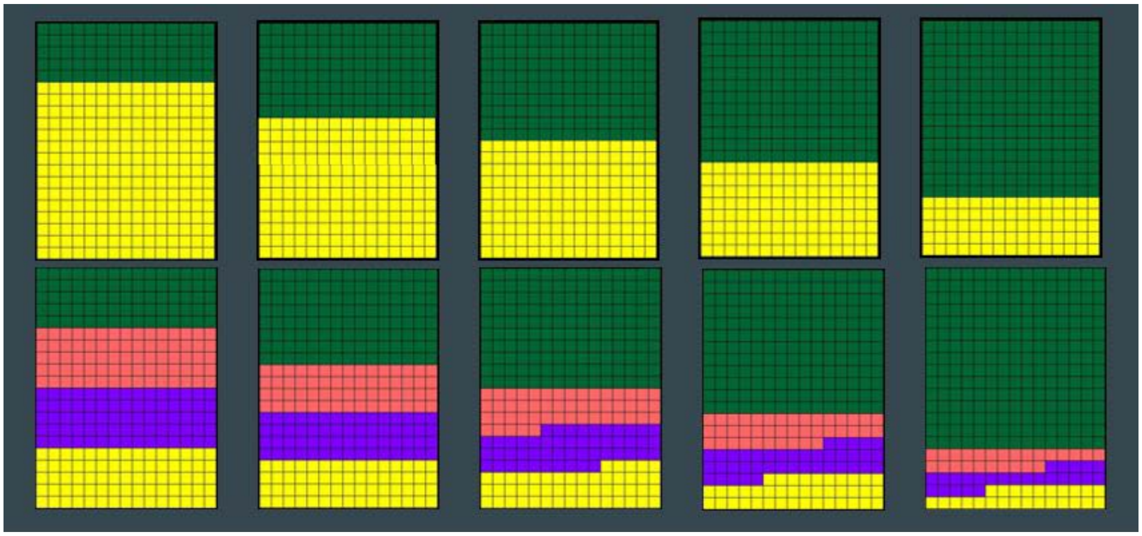

Figure 6. A back-to-back graphic of the different percentages conditions. The top row consists of the grids from the dichotomous condition and the bottom row consists of the grids from the polychotomous condition. From left to right, the percentage condition is: 25%, 40%, 50%, 60%, and 75%.

Appendix A

To: Cassandra Smith
Psychology
CAMPUS EMAIL

From: Monica Molina, IRB Associate Administrator
Date: 4/05/2018
RE: Notice of IRB Exemption

STUDY #: 18-0234
STUDY TITLE: Making Predictions

Exemption Category: (2) Anonymous Educational Tests; Surveys, Interviews or Observations

This study involves minimal risk and meets the exemption category cited above. In accordance with 45 CFR 46.101(b) and University policy and procedures, the research activities described in the study materials are exempt from further IRB review.

All approved documents for this study, including consent forms, can be accessed by logging into IRBIS. Use the following directions to access approved study documents.

1. Log into IRBIS
2. Click "Home" on the top toolbar
3. Click "My Studies" under the heading "All My Studies"
4. Click on the IRB number for the study you wish to access
5. Click on the reference ID for your submission
6. Click "Attachments" on the left-hand side toolbar
7. Click on the appropriate documents you wish to download

Study Change: Proposed changes to the study require further IRB review when the change involves:

- an external funding source,
- the potential for a conflict of interest,
- a change in location of the research (i.e., country, school system, off site location),
- the contact information for the Principal Investigator,
- the addition of non-Appalachian State University faculty, staff, or students to the research team, or
- the basis for the determination of exemption. Standard Operating Procedure #9 cites examples of changes which affect the basis of the determination of exemption on page 3.

Investigator Responsibilities: All individuals engaged in research with human participants are responsible for compliance with University policies and procedures, and IRB determinations. The Principal Investigator (PI), or Faculty Advisor if the PI is a student, is ultimately responsible for ensuring the protection of research participants; conducting sound ethical research that complies with federal regulations, University policy and procedures; and maintaining study records. The PI should review the IRB's list of PI responsibilities.

To Close the Study: When research procedures with human participants are completed, please send the Request for Closure of IRB Review form to irb@appstate.edu.

If you have any questions, please contact the Research Protections Office at (828) 262-2692.

Appendix B

Consent to Participate in Research
Information to Consider About this Research

Making Predictions

Principal Investigator: Cassandra L. Smith - smithcl13@appstate.edu
Faculty Advisor: Andrew R. Smith – smithar3@appstate.edu
Department: Psychology

You are being invited to take part in a research study about how people make decisions. If you take part in this study, you were one of about 600 people to do so. By doing this study we hope to learn how people make decisions about uncertain events. In this study, you were asked to complete a task where you can make different types of predictions. Also, you were asked questions about your personality, age, and gender. Participation will take no more than 30 minutes.

Unfortunately, you cannot volunteer for this study if are under 18 years of age. To the best of our knowledge, the risk of harm for participating in this research study is no more than you would experience in everyday life. Your identity cannot be connected to your data. Therefore, all of your responses in this study will remain confidential. There may be no personal benefit from your participation but the information gained by doing this research will help us understand factors that influence people's decisions. In turn, this may help us design practices to improve people's decision making.

You will not be paid for your time, but your participation in this study you will earn you 1 ELC via the SONA system. Your participation in this study is completely voluntary. You can decide to stop at any time for any reason and you many skip any question you would prefer not to answer. You will receive no penalty for stopping this study early. In order to fulfill your ELC requirement, there are research and non-research alternatives to participating in this study. For example, one non-research option is to read an article and write a 1-2 page paper summarizing the article and your reaction to it. This would be worth 1 ELC. Additionally, there are other studies you may participate in to meet this requirement. More information about this option can be found at: psych.appstate.edu/research. You may also wish to consult your professor to see if other non-research options are available.

The people conducting this study were available to answer any questions concerning this research, now or in the future. You may contact the Principal Investigator at 828-262-2272 or smithcl13@appstate.edu. You may also contact the faculty advisor at smithar3@appstate.edu. If you have questions about your rights as someone taking part in research, contact the Appalachian Institutional Review Board Administrator at 828-262-2692 (days), through email at irb@appstate.edu or at Appalachian State University, Office of Research and Sponsored Programs, IRB Administrator, Boone, NC 28608.

Appalachian State University's Institutional Review Board (IRB) has determined this study to be exempt from IRB oversight.

By continuing to the survey, I acknowledge that I am at least 18 years old, have read the above information, and provide my consent to participate under the terms above.

Appendix C

Life Orientation Test (LOT)—Revised

Please be as honest and accurate as you can throughout. Try not to let your response to one statement influence your responses to other statements. There are no "correct" or "incorrect" answers. Answer according to your own feelings, rather than how you think "most people" would answer.

1	2	3	4	5
I disagree a lot	I disagree a little	I neither agree nor disagree	I agree a little	I agree a lot

1. In uncertain times, I usually expect the best. ______________
2. It's easy for me to relax. ______________
3. If something can go wrong for me, it will. ______________
4. I'm always optimistic about my future. ______________
5. I enjoy my friends a lot. ______________
6. It's important for me to keep busy. ______________
7. I hardly ever expect things to go my way. ______________
8. I don't get upset too easily. ______________
9. I rarely count on good things happening to me. ______________
10. Overall, I expect more good things to happen to me than bad. ______________

Appendix D

Rational-Experiential Inventory (REI)—40

Test Format: The items are rated using a 5-point scale ranging from 1 (definitely not true of myself) to 5 (definitely true of myself).

Items

Rationality scale
I try to avoid situations that require thinking in depth about something. (re–)
I'm not that good at figuring out complicated problems. (ra–)
I enjoy intellectual challenges. (re)
I am not very good at solving problems that require careful logical analysis. (ra–)
I don't like to have to do a lot of thinking. (re–)
I enjoy solving problems that require hard thinking. (re)
Thinking is not my idea of an enjoyable activity. (re–)
I am not a very analytical thinker. (ra–)
Reasoning things out carefully is not one of my strong points. (ra–)
I prefer complex problems to simple problems. (re)
Thinking hard and for a long time about something gives me little satisfaction. (re–)
I don't reason well under pressure. (ra–)
I am much better at figuring things out logically than most people. (ra)
I have a logical mind. (ra)
I enjoy thinking in abstract term. (re)
I have no problem thinking things through carefully. (ra)
Using logic usually works well for me in figuring out problems in my life. (ra)
Knowing the answer without having to understand the reasoning behind it is good enough for me. (re–)
I usually have clear, explainable reasons for my decisions. (ra)
Learning new ways to think would be very appealing to me. (re)

Experientiality scale
I like to rely on my intuitive impressions. (ee)
I don't have a very good sense of intuition. (ea–)
Using my gut feelings usually works well for me in figuring out problems in my life. (ea)
I believe in trusting my hunches. (ea)
Intuition can be a very useful way to solve problems. (ee)
I often go by my instincts when deciding on a course of action. (ee)
I trust my initial feelings about people. (ea)
When it comes to trusting people, I can usually rely on my gut feelings. (ea)
If I were to rely on my gut feelings, I would often make mistakes. (ea–)
I don't like situations in which I have to rely on intuition. (ee–)
I think there are times when one should rely on one's intuition. (ee)
I think it is foolish to make important decisions based on feelings. (ee–)
I don't think it is a good idea to rely on one's intuition for important decisions. (ee–)

I generally don't depend on my feelings to help me make decisions. (ee–)
I hardly ever go wrong when I listen to my deepest gut feelings to find an answer. (ea)
I would not want to depend on anyone who described himself or herself as intuitive. (ee–)
My snap judgments are probably not as good as most people's. (ea–)
I tend to use my heart as a guide for my actions. (ee)
I can usually feel when a person is right or wrong, even if I can't explain how I know. (ea)
I suspect my hunches are inaccurate as often as they are accurate. (ea–)

Note. The name of the subscale to which each item belongs appears in parentheses, ee = Experiential Engagement; ea = Experiential Ability; re = Rational Engagement; ra = Rational Ability. A minus sign (—) with a scale name denotes reverse scoring.

Appendix E

1. Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?
Answer: **500**

2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?
Answer: **10** people

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?
Answer: **.10%**

4. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?
Answer: **100** people

5. If the chance of getting a disease is 20 out of 100, this would be the same as having a **20**% chance of getting the disease.

6. Suppose your friend just had a mammogram. The doctor knows from previous studies that, of 100 women like her, 10 have tumors and 90 do not. Of the 10 who do have tumors, the mammogram correctly finds 9 with tumors and incorrectly says that 1 does not have a tumor. Of the 90 women without tumors, the mammogram correctly finds 80 without tumors and incorrectly says that 10 have tumors. The table below summarizes this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

	Tested Positive	Tested Negative	Totals
Actually has a tumor	9	1	10
Does not have a tumor	10	80	90
Totals	19	81	100

Answer: **9** out of **19**

7. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?
Answer: **5** cents

8. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
Answer: **47** days

Footnotes

[1] In addition, ninety-six different participants were asked to provided certainty estimates on an 11-point numerical measure of certainty (i.e., 0% - 100%). The results of the analyses using the two different outcome measures yielded the same conclusion.

[2] I decided on this distribution, as opposed to the commonly used 30%, 40%, 50%, 60%, 70% (for a review, see Krizan and Windschitl, 2007). This experiment will manipulate dichotomous versus polychotomous by changing the number of each of the colors on the 300-square grid. I needed frequencies that could be divided evenly by two for the dichotomous condition, and frequencies that could be further divided by three for the three non-critical colors in the polychotomous condition.

Vita

Cassandra Louise Smith was born in Alexandria, VA, to Frances and Jeffrey Smith. She graduated from Thomas A. Edison High School in June, 2012. The following autumn, she entered Appalachian State University to study psychology, and in May, 2016, she graduated summa cum laude with a Bachelor of Science degree in Psychology. In the 2016-2017 academic year, she obtained a Certificate in Applied Statistics from the University of South Carolina in Columbia, SC. In the fall of 2017, she enrolled into the Master of Arts Experimental Psychology program at Appalachian State University under the advisement of Dr. Andrew R. Smith, researching judgment and decision making and social cognition. She obtained her M.A. in Psychology in May, 2019.

Ms. Smith is a Graduate Student Ambassador for Appalachian State University and a member of the Society for Judgment and Decision Making. After graduation, she plans to attend The Ohio State University in Columbus, OH, to pursue a Ph.D. in Decision Psychology.