12-2023

NO EVIDENCE THAT REASONED ANALYSIS IMPAIRS THE ACCURACY OF (OR CONFIDENCE IN) SPORTS FORECASTS

Andrew Langbehn
alangbe1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Andrew Langbehn entitled "NO EVIDENCE THAT REASONED ANALYSIS IMPAIRS THE ACCURACY OF (OR CONFIDENCE IN) SPORTS FORECASTS." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Psychology.

Jeff T. Larsen, Major Professor

We have read this thesis and recommend its acceptance:

Michael A. Olson, Lowell Gaertner

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

NO EVIDENCE THAT REASONED ANALYSIS IMPAIRS THE ACCURACY OF (OR CONFIDENCE IN) SPORTS FORECASTS

A Thesis Presented for the
Master of Arts
Degree
The University of Tennessee, Knoxville

Andrew Thomas Langbehn
December 2023

ACKNOWLEDGEMENTS

This thesis would not have been possible without the encouragement and support of many. Here I would like to extend special thanks to my advisor, Jeff Larsen Ph.D., without whom this thesis would not have been possible. I would also like to thank my thesis committee members, Michael Olson Ph.D. and Lowell Gaertner, as well as Sarah Lamer Ph.D. for their help in the preparation of this thesis. Lastly, I would like to thank my family whose support is the reason I stand here today and the driving force behind all my successes.

ABSTRACT

When making decisions, people can either rely on a gut feeling or engage in reasoned analysis to make a choice. Past research has made competing claims on whether relying on gut feelings or reasoned analysis leads to better decisions. However, these competing claims may be due to the types of decisions being made. Relying on gut feelings has been demonstrated to be superior in judgments about attitudes and leads to greater post-decision satisfaction. However, prior research demonstrating the benefits of gut feelings has used subjective and mostly unverifiable criteria for which to judge the quality of a decision. On the other hand, reasoned analysis has largely been found to be beneficial in other tasks that have an objective, verifiable outcome or criterion for which to judge the quality of a decision. Here, we explore a potential exception to this, sports forecasting. Prior research has demonstrated that relying on gut feelings leads people to make more accurate forecasts about the outcome of sporting events. The first aim of the current research was to replicate this result. However, across the 5 experiments reported here, we see no evidence that relying on a gut feeling increases the accuracy of sports forecasts. The second aim of the current work was to extend prior research to examine how confidently people hold sports forecasts made based on gut feelings and reasoned analysis. We further extend prior work by examining how individual differences affect and interact with the experimental manipulation of how people make decisions. In these additional lines of inquiry, we first saw no evidence supporting the hypothesis that relying on gut feelings increases peoples' confidence in their choices. Second, we observed that individual differences did not predict the accuracy of peoples' forecasts or how confident they were about their forecasts. We conclude that relying on a gut feeling or reasoned analysis is unlikely to affect sports forecasting. Therefore, sports forecasts may be a boundary condition for which gut feelings and reasoned analysis arrive at the same decision and the benefits of relying on gut feelings may be restricted to subjective judgments and individual post-decision evaluations.

TABLE OF CONTENTS

Chapter One Introduction and General Information 1
Gut Feelings and Reasoned Analysis 2
Individual Differences in Reliance on Gut Feelings and Reasoned Analysis 13
Decisional Fit 19
Decision Confidence 21
Chapter Two Study 1A 24
Hypotheses 26
Method 28
Participants 29
Materials and Procedure 30
Forecasts 32
Confidence Ratings 34
Individual Differences 34
Additional Measures 35
Results 35
Manipulation Checks 35
Differential Attrition 36
Forecast Accuracy 38
Individual Differences 40
Forecast Confidence 45
Exploratory Analyses 48
Discussion 52
Chapter Three Study 1B 55
Method 55
Participants 55
Materials and Procedures 56
Results 57
Manipulation Checks 57
Differential Attrition 58
Forecast Accuracy 58
Individual Differences 60
Forecast Confidence 62
Discussion 63
Limitations 66
Chapter Four Study 2A 68
Method 69
Participants 69
Measures and Procedures 70
Forecasts 72
Results 74
Manipulation checks 74
Differential Attrition 74

Forecast Accuracy 75
Forecasts about the Favorite 76
Forecast Confidence 78
Discussion 80
Chapter Five Study 2B 82
Method 82
Participants 82
Measures and Procedures 83
Results 83
Manipulation checks 83
Differential Attrition 84
Accuracy of forecasts 84
Forecasts about the Favorite 85
Forecast Confidence 87
Discussion 88
Chapter Six Study 3 93
Method 94
Participants 94
Survey 95
Results 98
Differential Attrition 98
Manipulation checks 98
Forecast Accuracy 99
Forecasts about the Favorite 101
Confidence 101
Discussion 101
Chapter Seven General Discussion 104
Decision Mode and Forecast Accuracy 104
Population and Sample 105
Manipulation of Decision Mode 106
Differential Attrition 109
Individual Differences 110
Forecast Confidence 113
Conclusions 114
References 117
Appendices 129
Vita 138

LIST OF TABLES

Table 2.1 Study 1A Correlations Between Individual Difference Measures..... 42
Table 2.2 Study 1A Individual Differences and Forecast Accuracy 44
Table 2.3 Study 1A Individual Differences and Forecast Confidence 47
Table 2.4 Study 1A Correlations Between Measures of Fanship and Dependent Measures 51
Table 2.5 Study 1A Correlations Between Forecast Time and Dependent Measures 54
Table 3.1 Study 1B Number of Forecasts by Condition..... 59
Table 3.2 Study 1B Individual Differences and Forecast Accuracy..... 61
Table 3.3 Study 1B Individual Differences and Forecast Confidence..... 64
Table 4.1 Study 2A Individual Differences and Forecasts 77
Table 4.2 Study 2A Individual Differences and Forcacst Confidence 79
Table 5.1 Study 2B Individual Differences and Forecasts 86
Table 5.2 Study 2B Individual Differences and Forecast Confidence..... 89
Table 5.3 Effect of Experimental Manipulation on Reported Reliance on Decision Mode 92
Table 6.1 Study 3 Individual Differences and Decisional Fit..... 100

CHAPTER ONE

Introduction and General Information

Everyday life demands that people are able to make decisions effectively and efficiently. At times, a decision can potentially have life-altering consequences (e.g., whether to attend college) while other decisions are more mundane (e.g., what to have for dinner). However, regardless of the type of decision being made, people want to get them right and considerable literature has been devoted to helping people make better decisions. Many people's intuition, for which there is some empirical support, is that the more they reason analytically about a decision the more likely they are to make the correct or at least the best available choice. In line with this idea, people often engage in drawn out analyses of their decisions, and reasoning has been championed as an effective decision making process (Janis & Mann, 1977). However, engaging in reasoned analysis requires cognitive resources, time, and effort that is not always available when a decision is required (Kahneman & Frederick, 2002). Fortunately, reasoning is not the only way someone can make a decision.

According to dual-process theories of cognition, there are two distinct processes by which a decision can be made. The first, *reasoned analysis*, is the more cognitively demanding of the two. It represents a controlled and deliberate process through which information is carefully weighed until a decision can be made. In contrast, intuition represents a more automatic, quick, and less demanding process (Kahneman & Frederick, 2002). Intuitive processing involves relying upon a first instinct or gut feeling about which option to choose. These gut feelings can be described as an experience of knowing

something without knowing how it is that you came to know it (Epstein, 2010). Because of life's constant demands, and the limited resources (e.g., attention and time) available to process information, people rely on their gut feeling for most of daily life's decisions (Kahneman, 2011). However, whether relying on a gut feeling or reasoned analysis leads to better decision outcomes is under considerable debate.

Gut Feelings and Reasoned Analysis

There is mixed evidence on whether reasoned analysis or relying on a gut feeling leads people to make better decisions (McMackin & Slovic, 2000). In many cases, someone's gut feeling and reasoned analysis may lead them to the same judgment. For example, most people wouldn't put their hand on a hot stove regardless of whether they are relying on a gut feeling or reasoned analysis. However, of interest are when these two modes of thought lead to different outcomes. Take, for example, one of the questions on the Cognitive Reflection Test (CRT) developed by Frederick (2005).

> "A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?"

In response to this question, many people have a gut feeling that the ball must cost $0.10. However, this question was designed such that the gut feeling response leads people astray and only through reasoned analysis will most people arrive at the correct answer. After reasoning about the question more carefully, it becomes clear that the ball, in fact, would cost only $0.05 ($1.10 = X + ($1.00 + X)). The CRT is illustrative as a case where people's gut feeling and reasoned analysis lead them to different answers and here

reasoned analysis leads people to the correct response while their gut feeling often leads them astray.

The CRT is not the only case in which reasoning leads people to make better decisions. MacGregor et al. (1988) asked participants to estimate different unknown quantities, for example, the amount of mail handled by the U.S. Postal Service. Participants who were first asked to list all the factors they thought were relevant to their decision were more accurate in their estimations than those given no instructions on how to make their judgments. Using a similar paradigm, McMackin and Slovic (2000) later replicated this finding. In their study, those asked to think analytically and write in reasons for their estimates (e.g., the area of the U.S.) before making them, were again more accurate than those given no instructions about how to derive their estimates. As illustrated above, engaging in reasoned analysis leads to more accurate estimates about unfamiliar quantities. However, reasoned analysis does not always improve decision making.

While reasoned analysis increases the accuracy of estimates, in other types of decisions, reasoned analysis seems to impair decision making. For example, Wilson et al. (1993) asked some participants to think about and record the reasons why they liked and disliked different posters. At the end of the study, they allowed their participants to take one of the posters home with them. Those people who first wrote down the reasons why they liked and disliked each poster were more likely to choose the lower quality poster (as rated in a pretesting session) and were less satisfied weeks later in their choice than participants in a control condition. This decrease in satisfaction occurred in both self-

report measures of how much they liked their poster and behavioral measures (e.g., if they hung the poster in their dorm room). Dijksterhuis & van Olden (2006), using a similar research paradigm, asked some participants to engage in reasoned analysis and had others complete a distractor task before making their choice. They further demonstrated that reasoned analysis reduces post-decision satisfaction. One explanation for why participants asked to think about reasons why they liked and disliked each poster were later less satisfied with their choice is that reasoning about attitudes has been shown to temporarily change them (Wilson et al., 1989). When asked to reason about why someone feels positively or negatively toward a stimulus people frequently change their attitudes to be in line with the reasons that they report (Wilson & Hodges, 1992). This is supported by evidence that reasoned analysis about one's preferences leads to lower levels of preference-behavior consistency (Levine et al., 1996; Nordgren & Dijksterhuis, 2009). Therefore, an alternative explanation for the results of Wilson et al. (1993) and Dijksterhuis & van Olden (2006) exists.

While post-decision satisfaction provides indirect evidence that reasoning can impair decision making, it is unclear whether their choice represents an objectively worse decision. It is possible that when asked how satisfied they were with their choice, participants were relying on their gut feeling about how much they liked or disliked their choice (e.g., the poster they chose) and that these studies more simply indicate that reasoning and gut feelings lead to different attitudes about an object. As discussed, this assumption is supported by reasoning-based attitude change (Wilson et al., 1989). For example, in Wilson's (1993) study, participants who were asked to engage in reasoned

analysis about which poster to choose were more likely to choose a humorous as opposed to an artistic poster while those in the control condition liked the artistic poster more. While pretests indicated that, on average, people tended to like the artistic poster more, it is possible that those in the pretesting sessions were relying on a gut feeling to make their ratings. In such a case, this concordance would reflect a difference between attitudes arrived at through reasoned analysis and gut feelings. It is therefore hard to assess the objective quality of their decision (i.e., whether or not they made the correct choice).

There is some evidence that reasoning about a choice can reduce the quality of the decision beyond someone's post-decision feelings about their choice (see Plessner & Czenna, 2011 for a review). For example, Wilson and Schooler (1991) had participants taste test a series of 6 jams. Participants who were asked to reason about the quality of each jam were less likely to rate them in line with expert taste testers' opinions than those in a control group. In a similar study, Wilson and Schooler (1991) asked student-participants to select courses to enroll in for the following semester. Again, they observed that those asked to think about reasons why they should take one course over others were less likely to choose those that were rated highly by other participants in a control condition. Likewise, these participants were also less likely to enroll in courses that were generally well-reviewed. These studies support the assertion that reasoning impairs decision making in affective judgments and provide a more objective criterion for which to compare decisions. However, the criterion by which these decisions are compared is others' judgments. Therefore, the alternative explanation still exists that judgments made based on reasoned analysis are different than those that are made based on a gut feeling

and not objectively of lower quality. These studies help us understand how *decision modes* (i.e., reasoned analysis or gut feelings; Schunk & Betsch, 2006) change our attitudes and provide support that reasoning changes our decision making process, and the outcome. However, without an objective and *verifiable* decision criterion for which to compare a choice against (i.e., one that is based on an observable outcome not contingent on self-report), whether reasoned analysis leads to suboptimal decisions is unclear.

How the quality of a decision is evaluated is a perennial problem in judgment and decision making research. The literature reviewed thus far demonstrating that relying on a gut feeling leads to better decisions has used post-decision satisfaction and others' evaluations as measures of decision quality. However, as noted previously, it is unclear whether these studies are assessing the quality of a decision. A few studies have attempted to address this limitation. For example, in a series of studies, Dijksterhuis (2004) was able to assess the quality of decisions by presenting participants with complex decision problems. In one such study, Dijksterhuis showed participants a series of apartments that varied on a range of positive (e.g., the apartment is fairly large) and negative attributes (e.g., the apartment has an unfriendly landlord). Important for the present research, participants were randomly assigned to either think carefully for 3 minutes or complete a 3-minute *n*-back task in which participants are shown a series of digits and asked to recall the digit *n* places back. The *n*-back task recruits processes involved in working memory (Jonidas et al., 1997) reducing the ability of participants to use reasoned analysis to evaluate the apartment during the task. After thinking carefully for 3 minutes or completing the 3-minute *n*-back task, participants judged each apartment

from extremely negative to extremely positive. Participants asked to think carefully rated the best and worst apartments as more similar than those prevented from using reasoned analysis before making their judgments. In a subsequent study Dijksterhuis (2004) found that those asked to think carefully were also less likely to choose the apartment with the highest number of positive attributes.

In a similar study, Mikels et al. (2011) asked participants to choose between 4 different cars. Participants were shown positive and negative characteristics about each car. One car was described with mostly (75%) positive attributes, one was given mostly negative attributes (75%) and the remaining 2 were given an equal number of positive and negative attributes. Before making a selection, Mikels and colleagues asked some participants to focus on their emotional reactions to the cars while the other half were told to focus on the details of the information provided about them. When each car was given only 4 attributes, participants asked to focus on their emotions and those asked to focus on the details were equally as likely to pick the car with the most positive attributes. However, when each car was given 12 attributes in the same proportions as before, participants asked to focus on their emotions more often picked the car with the most positive attributes compared to those asked to focus on the details (Mikels et al., 2011). Across the two studies, it appears that even against a verifiable criterion gut feelings improved the quality of decision making at least when presented with a decision that includes numerous pieces of information. However, while the paradigm utilized in Dijksterhuis (2004) and Mikels et al. (2011) approach an objective criterion, what represents a positive attribute (e.g., the apartment is located in the center of a city) is still

dependent on an individual's preferences for a particular characteristic, and its relative importance to them. Despite this, the literature reviewed thus far supports the hypothesis that relying on gut feelings improves (and reasoned analysis impairs) decision making at least in some cases.

The effect of gut feelings on decision making likely depends on the type of decision being made. As discussed previously, complex decisions, those in which the options vary on many dimensions, may be one domain where gut feelings improve decision making (Mikels, et al., 2011). Beyond the complexity of the decision, it may also depend on whether someone has an affective reaction about the options in a decision. As noted earlier, McMackin and Slovic (2000) asked some participants to estimate different unknown quantities and sizes. In this case, reasoned analysis outperformed gut feelings. In another task, McMackin and Slovic (2000) asked participants to make affectively charged decisions. In this task they had participants evaluate the quality of different advertisements. Specifically, they asked participants to rate how much other people liked 12 print advertisements. They again told half of the participants to think about and provide reasons for their response before deciding. As with the previous study, the remaining participants were not told how to make their decisions. Unlike their previous experiment, participants asked to think about and provide reasons were less accurate at estimating others' ratings of the advertisements. Here reasoned analysis seems to impair the estimation of how others feel about the same advertisements. Reasoned analysis, according to McMackin and Slovic (2000), impairs decision making in

affectively charged decisions (e.g., rating the quality of advertisements) and improves decision making in objective tasks (e.g., estimates of quantities).

Despite the limited scope of research discussed thus far, it is likely that at least in some cases relying on gut feelings improves (or reasoned analysis hinders) the decision making process. What is less clear is why this might be the case. While the mechanisms by which this might occur are largely beyond the purview of the present research, we will review some of them briefly here. Reasoned analysis may impair decision making based on the information that is available during the decision making process. As discussed earlier, reasoning temporarily changes people's attitudes in line with the reasons that are accessible during the decision making process (Wilson et al., 1989; Wilson, 1993; Wilson et al. 1995; Wilson & LaFleur, 1995). What information is available and easier to access may bias the decision making process toward the reasons that someone is able to come up with at the time (Nisbett & Wilson, 1977). Beyond what information is available, the limited capacity of working memory may further limit what reasons can be deliberated upon during the decision making process (Dijksterhuis & Nordgren, 2006). Further, the task of writing down reasons may lead people to rely on reasons that are easier to verbalize. In support of this claim, the task of verbalizing reasons for a judgment has been shown to impair decision making (Schooler & Engstler-Schooler, 1990). Reasoned analysis may also impair decision making due to the improper weighting of information. For example, Levine and colleagues (1996) observed that reasoning increases the variability and inconsistency of reasoners' weighting schemes when making attitude judgments. Finally, reasoned analysis may impair decision making by disrupting an

otherwise adequate decision making process (Halberstadt & Hooton, 2008). Further, Gigerenzer et al. (1999) argue that most judgments can be made relying on simple heuristic cues (e.g., recognition, fluency; for a review see Gigerenzer & Gaissmaier, 2011). Reasoning about a decision then impairs this already sufficient decision making process which has been described as being based on a gut feeling or affective reaction (Halbserstadt & Hooton, 2008). While there are many proposed mechanisms through which reasoned analysis impairs decision making (for a review see Halberstadt & Wilson, 2008) reasoned analysis, at least when making complex decisions in which people are likely to have affective reactions, impairs the quality of decisions.

While there are numerous mechanisms for why decision making may be hindered by reasoned analysis, all of the evidence reviewed thus far that supports the hypothesis that relying on gut feelings improves decision making has used semi-subjective or unverifiable decisions (decision satisfaction, comparison with others' attitude judgments). The first study to demonstrate the benefits of relying on gut feelings in an objective and verifiable decision making paradigm, was conducted by Halberstadt and Levine (1999).

In their study, they asked participants to make forecasts about who would win Sweet 16 basketball games. Sports forecasts is one domain in which gut feelings would be expected to improve decision making. Specifically, people are likely to have affective reactions to which team they think will win a game and there are numerous factors that influence the outcome of a game (e.g., injuries, home-court advantage, and coaching). However, unlike the other types of decisions reviewed thus far, it also provides an objective and verifiable criterion (who won the game and by how many points they won)

for which to compare judgments against. In their study, Halberstadt and Levine (1999) randomly assigned half of their participants to read instructions that told them to go with their gut instinct and to try not to think too much about why a particular team would win each game. The remaining participants engaged in reasoned analysis before making their decision. Specifically, they were asked to think about and provide three reasons why they thought a game would turn out a particular way before making their forecasts. Halberstadt and Levine (1999) observed that participants asked to go with their gut feeling were more accurate at forecasting the outcome of the game than those asked to think about and provide reasons before making their forecasts. To our knowledge, this is the first study to demonstrate that reasoned analysis impairs decision making in a task with an objective and verifiable criterion for which to compare participants' decisions. However, one potential problem with their experiment is the use of a sample size that is fairly small (N = 108) compared to current standards. While Halberstadt and Levine (1999) offer the first test of the benefits of relying on gut feelings in objectively verifiable decision making, their sample size raises concerns about the replicability of their results.

There is some evidence that Halberstadt and Levine's (1999) results do replicate. Plessner et al. (2006) as described briefly in Plessner and Czenna (2007)[1], followed a similar experimental design as Halberstadt and Levine. They asked people to forecast the outcome of FIFA soccer games. Like Halberstadt and Levine, they asked half of their participants to rely on a gut feeling when making their forecasts while the remaining participants were asked to use reasoned analysis to make their forecasts. In one study

[1] In addition to discussing the studies in broad detail in Plessner and Czenna (2007) Plessner and colleagues provided PowerPoint slides from a conference presentation discussing their results.

with a similar sample size (N = 104), participants asked to rely on their gut feelings made more accurate forecasts than those asked to reason about their forecasts (H. Plessner, personal communication, April, 14, 2022).

In a follow up study, the benefit of relying on gut feelings only occurred for those with a high degree of soccer expertise as measured by a soccer-expertise test (N = 80). However, evidence for the effects of reasoned analysis were limited to a fairly narrow set of conditions. In one study, these results were qualified by their involvement with higher level interactions involving the time until the game and whether additional information about the teams was provided to participants. They conclude that relying on gut feelings improves the accuracy of sports predictions, however, only for experts and under certain conditions. Specifically, this effect was only observed when the game was to be played the following week (as opposed to a month later) and there is explicit information available about the game being played (H. Plessner, personal communication, April, 14, 2022). While these results seem to support Halberstadt and Levine (1999) they have not been published in a peer-reviewed journal. Therefore, the first aim of the current work is to provide replication studies in which we ask participants to make forecasts about the outcome of National Football League (NFL) games. Given previous results, we will recruit participants that are likely to have a high degree of expertise and ask them to make forecasts about games that will be played within the next few days. According to Plessner et al., 2006 this is the most likely scenario in which relying on gut feelings increases sports forecast accuracy.

Individual Differences in Reliance on Gut Feelings and Reasoned Analysis

While the research discussed thus far has attempted to manipulate how people make decisions (e.g., asking them to rely on a gut feeling or think about and provide reasons), people differ on the degree to which they prefer to rely on the decision modes (gut feeling, reasoned analysis). Several studies have found that people have stable individual differences in the degree to which they tend to rely on their gut feeling and reasoned analysis when making decisions (Epstein et al., 1996; Sadler-Smith, 2004; Betsch, 2004; Betsch & Iannello, 2010). Therefore, it is possible that people's preferences for how they make a decision may affect the quality of decision making outside of the experimental manipulation of decision mode.

There are several individual differences measures that have been designed to measure people's reliance on the two decision modes. These measures were developed from separate but related theoretical perspectives. Therefore, they use slightly different terms for their underlying constructs. However, they can be understood as measuring related concepts (i.e., the tendency to rely on reasoned analysis vs. gut feelings).

Epstein et al. (1996) introduced the Rational and Experiential Self Inventory (REI; Appendix A) from Cognitive Experiential Self Theory (CEST; Epstein, 1990). CEST posits that there are two information processing systems (rational, experiential). The rational system is rooted in conscious deliberate thought (i.e., reasoned analysis) while the experiential system uses holistic, preconscious, and affect based processing (i.e., gut feelings). Importantly, CEST contends that these processing systems are independent. In line with this assumption the REI has two subscales for which to measure

people's tendency to rely on reasoned analysis (REI-R; e.g., "I have a logical mind", "Using logic usually works well for me in figuring out problems in my life") and gut feelings (REI-E; e.g., "I believe in trusting my hunches", "I tend to use my heart as a guide for my actions).

A similar measure, developed by Scott and Bruce (1995), was developed from a behavioral perspective that people have a tendency to rely on one decision mode as opposed to another when processing information and making decisions. Importantly, this perspective views decision modes as non-independent. Scott and Bruce propose that people have a learned tendency (habit) to use one decision mode or another when faced with a decision. Therefore, they argue against a trait-based perspective and focus on how people have typically made decisions in their past to predict how they will do so in future decision making. This is reflected in the items in their General Decision-Making Style Inventory (GDMS; Scott & Bruce, 1995; Appendix B). The GDMS contains 4 subscales, however of interest here, are the rational (GDMS-R; e.g., "I make decisions in a logical and systematic way", "My decision making requires careful thought") and intuitive (GDMS-I; e.g., "When making a decision, I rely upon my instincts", "When I make a decision, I trust my inner feelings and reactions") subscales. The GDMS seeks to understand how people have made decisions in the past to predict how they will do so in the future. In line with this assumption, experimental evidence suggests that the decision mode people have utilized in the past is able to predict which they will choose for a future decision (Scott & Bruce, 1996).

The final measure of someone's tendency to rely on reasoned analysis and gut feelings we consider here is the Preference for Intuition and Deliberation scale constructed by Betsch (PID; 2004; Appendix C) which was developed specifically to assess people's independent preferences for the two decision modes. As with the previous scales the PID is separated into two subscales. The first, the PID-D measures people's tendency to rely on reasoned analysis (termed deliberation in Betsch, 2004) and includes items such as "before making decisions I first think them through" and "when I have a problem, I first analyze the facts and details before I decide". The second, the PID-I measures peoples' tendency to rely on gut feelings (what Betsch called intuition) and includes items such as "I listen carefully to my deepest feelings" and "my feelings play an important role in my decisions". Importantly, the PID is highly correlated with how people report they would approach a decision in hypothetical decision making scenarios (Betsch, 2004).

As the three measures of individual differences measure slightly distinct aspects, it is possible that peoples' responses may vary across them. In line with this assumptions, prior research has shown that the three subscales measuring peoples' tendency to rely on gut feelings are highly correlated (all rs > .79, ps < .05; Cook & Gonzales, 2016). However, responses to the three subscales measuring peoples' tendency to rely on reasoned analysis show a weaker correlation and vary across comparisons. For example, Cook and Gonzales (2016) observed that the REI-R and GDMS-R are moderately correlated (r = .40), the REI-R and PID-D show a strong correlation (r = .67) and the GDMS-R and PID-D show a moderate correlation (r = .48, all ps < .05). Prior research

utilizing the three scales has also found that the PID subscales are slightly negatively correlated (r = -.20, p < .05; Betsch, 2004), the REI subscales (r = -.02, p >.05; Pacini & Epstein, 1999) are not significantly correlated, and the relationship between GDMS subscales range from slightly negative (r = -.25, p < .05) to a non-significant relationship (r = -.08, p > .05) depending on the sample (Scott & Bruce, 1995). Given the differences observed between the three measures we included all three in the present work.

Individual differences in the degree to which people tend to rely on the two decision modes have been shown to correlate with both how people make decisions and the quality of their decisions. For example, people who tend to rely on gut feelings are faster at making a decision (Schunk & Betsch, 2006) and rely more on implicit knowledge (Richetin et al., 2007). This supports the assumption that relying on gut feelings and reasoned analysis are separate processes.

Based on the evidence reviewed previously regarding manipulated decision mode and decision quality, we would also expect individual differences to affect the quality of decisions. Some support for this hypothesis comes from experiments utilizing *ratio-bias* tasks. The ratio-bias task was developed so that engaging in reasoned analysis and gut feelings lead to different outcomes. Specifically, reasoned analysis leads people to the correct answer and gut feelings leads them astray. In the ratio-bias task participants are given two trays of marbles. Each tray contains several red and white marbles. The trays vary on the total number of marbles and the proportion of red and white marbles they contain. The participant's task is to pick the tray that offers them the best chance of randomly drawing a red (winning) as opposed to white (losing) marble. On some trials,

one tray contains, 10 winning red marbles and 90 white marbles (10% chance of drawing a red marble). The other tray contains 1 winning red marble and 9 white marbles (11% chance of drawing a red marble).

On these trials the normatively rational choice is to pick the second tray. It offers a higher *objective probability* of winning (11%). However, Denes-Raj & Epstein (1994) observed that 61% of people pick the first tray which offers a lower objective probability of winning (10%) at least once. However, as noted previously the task is set up such that many peoples' gut feeling is to choose the first tray. Despite tray two offering a lower objective probability of winning, it offers more *chances* to win (i.e., it has more red marbles). The number of potential winners offered by tray two is intuitively appealing. This effect has been repeatedly demonstrated across variations of the ratio-bias task (e.g., Kirkpatrick & Epstein, 1992; Denes-Raj & Epstein, 1994; Walco & Risen, 2017). Importantly for the present work is that people who tend to rely on their gut feelings more are more likely to make this error (Pacini & Epstein, 1999; Gärtner et al., 2022).

However, people who tend to rely on their gut feelings more do not always make lower quality decisions compared to those who tend to rely on reasoned analysis. In one study assessing the tactical decisions of handball players, Raab and Laborder (2011) found that not only did people who tend to rely on gut feelings make faster decisions, they also made tactically better decisions. Therefore, we see a similar pattern of results as those discussed previously in experiments that manipulated participants' decision modes. For affect based, complex decisions (e.g., handball plays) people who tend to rely on

their gut feelings do better while in more simple probability tasks relying on gut feelings hinders performance.

Someone's tendency to rely on gut feelings, as noted above, is related to decision making performance. Therefore, it is likely that someone's tendency to rely on reasoned analysis is also related to decision quality. In line with this prediction, people who have a tendency to rely on reasoned analysis perform better on adult decision-making competence tasks (de Bruin et al., 2007). Likewise, they are more accurate at assessing the probability of different poker hands in a simulated poker task and are less susceptible to anchoring bias (Welsh et al., 2014). Combined these studies indicate that both the tendency to rely on gut feelings and the tendency to rely on reasoned analysis affect the quality of decisions.

In Halberstadt and Levine's (1999) study it is likely that (a) people had preexisting individual differences for the degree to which they tend to rely on a gut feeling and reasoned analysis when making decisions and (b) that these tendencies are likely to have affected the accuracy of their participants' sports forecasts. Following the logic laid out thus far we would expect people who tend to rely on their gut feelings to be more accurate when making their forecasts than those who tend to rely on reasoned analysis. Therefore, the second aim of the current work is to extend Halberstadt and Levine's (1999) research to examine the effect of individual tendences toward relying on gut feelings and reasoned analysis on the accuracy of sports forecasts.

Decisional Fit

Beyond a main effect of an individual's tendency to rely on the two decision modes, it is possible in Halberstadt and Levine (1999) that their participants' tendencies toward the decision modes may have interacted with how they were told to make their forecasts about the outcome of Sweet 16 basketball games. While there has been little research to support this hypothesis in the judgment and decision making literature, the claim that individual characteristics interact with the environment is well supported (see Caplan, 1987 for a review). One related test of this hypothesis in decision making, was conducted by Higgins (2000; 2005). Higgins proposed that people judge their decisions to be better when they experience regulatory fit, when the means by which they pursue a goal matches their values and beliefs. Further, experimental evidence suggests that people assign higher values to chosen products when they experience regulatory fit (Higgins et al., 2003).

In terms of people's tendencies to rely on gut feelings and reasoned analysis when making a decision, Betsch and Kunz (2008) introduced *decisional fit*, the fit between the decision making mode someone tends to rely on (gut feeling, reasoned analysis) and their *applied mode* (the decision mode they use for a given decision). People who typically rely on reasoned analysis experience decisional fit when they use their reasoning to make a decision. By contrast, people who typically rely on their gut feelings experience decisional fit then they rely on a gut feeling to make a decision. That is, people experience decisional fit when the decision mode they tend to use and the one they apply to a given situation match.

In their study, Betsch and Kunz (2008) measured participants' tendencies to rely on the two decision modes using the PID before randomly assigning them to rely on a decision mode. Specifically, participants in the reasoned analysis condition were first asked to list the pros and cons of 2 coffee pots. In the non-reasoned analysis condition participants were told to make their choice spontaneously (akin to deciding based on a gut feeling). Participants who experienced decisional fit when making their choice attached a greater monetary value to the coffee pot. The authors interpret these results as indicating that decisional fit enhances the perceived value of the product. In subsequent studies reported in Betsch and Kunz (2008), decisional fit increases positive attitudes about objects, increases the perceived utility of a product, as well as reduces post-decision regret. Across their studies, decisional fit improved the subjective quality of a decision. While these studies implicate decisional fit as a factor influencing subjective measures of how people feel about products and their decisions, it is not clear whether these results translate to more objective measures of decision quality.

Using the ratio-bias task described earlier, Gärtner et al. (2022) attempted to address this question. In their study, they found that asking people to make their decision based on feelings led them to make more normatively irrational choices (choosing the tray with a lower probability of winning). Similarly, someone's tendency to rely on gut feelings led people to make the normatively irrational choice more often. However, the authors reported no effect of decisional fit on participants' choices.

Therefore, it is possible that decisional fit may only affect people's subjective feelings about decisions and not affect the accuracy or quality of decisions when

compared to an objective, verifiable criterion. While the ratio-bias task is one domain where decisional fit should be examined further, it differs from Halberstadt and Levine's (1999) study in a few important ways. First, the probabilities offered by the two trays are known whereas in the case of sports forecasting the probability of a team winning is unknown. Second, sports forecasting involves a much more complex decision than that involved in the ratio bias task. There are innumerable factors that influence the outcome of a basketball game while there is only one (the number of marbles of each color) that influences the outcome in the ratio-bias task. As discussed previously, one proposed mechanism for why reasoned analysis impairs decision making is that reasoning is biased by the factors brought into the decision making process. Therefore, if all the factors are known (e.g., the ratio of marbles) reasoning and gut feelings may perform equally. Therefore, a stronger test of the hypothesis that decisional fit improves decision making should be conducted in a paradigm like that of Halberstadt and Levine (1999) which is undertaken here.

Decision Confidence

The research reviewed thus far has demonstrated that whether relying on reasoned analysis or relying on gut feelings leads to better decision making is moderated by the type of task. While the quality of decision making in previous work had been assessed in a myriad of ways (e.g., decision satisfaction and regret, accuracy, and normatively rational vs irrational choices), how these decision modes affect other aspects of decision making has been relatively excluded from analysis. Despite this, there is reason to believe

that people who make decisions based on a gut feeling or their intuition will be more confident in their decision than those who engage in reasoned analysis.

As discussed earlier, intuitive judgments, those made based on a gut feeling, come to mind more easily. They arise from an effortless, fast, and automatic process (Kahneman & Frederick, 2002). Therefore, it is likely that responses made quickly are more likely to reflect an intuitive, as opposed to an analytic, processing mode. Prior research has demonstrated that the speed and ease (fluency) at which an answer is recalled is related to how confidently people hold their answer (Ackerman & Zalmanov, 2012). More specifically, prior research which has examined response latency and confidence has observed that people tend to be more confident in decisions made quickly (Kelley & Lindsay, 1993). Further, reaction times are positively related to how cognitively demanding people feel a decision is to make (Robinson et al., 1997). Unsurprisingly, the more demanding a decision feels the less confident people feel about their decision (Zakay & Tsal, 1993; Zakay et al., 1990; Zakay, 1985). Since reasoned analysis is more cognitively demanding (Kahneman & Frederick, 2002) it follows that relying on a gut feeling would lead people to be more confident in their decisions. This effect could have real world implications. In a study that assessed participants' recall of a mock crime, Robinson et al. (1997) observed that people were more confident in their recollection about the crime later if they perceived their answers to be based on automatic, as opposed to effortful, processing. Similarly, the longer participants took to answer the questions the less confident they were in their answers. Overall, people tend to hold their intuitive judgments with confidence (Bastick, 1982; Simmons & Nelson,

2006). Therefore, it is likely that people who rely on a gut feeling will be more confident in their decisions than those who rely on reasoned analysis.

CHAPTER TWO

Study 1A

Whether someone should rely on a gut feeling or engage in reasoned analysis when making a decision likely depends on the task at hand. Prior research demonstrating that going with a gut feeling improves decision making (e.g., Wilson & Schooler, 1991; Wilson et al., 1993, McMackin & Slovic, 2000) has mostly utilized semi-subjective or unverifiable criteria for which to compare a decision against. In these studies, as McMackin and Slovic argue, it is likely that people have an affective reaction about the alternatives from which to choose. In contrast, the benefits of reasoned analysis have been demonstrated in more objective decision making domains (e.g., Frederick, 2005; MacGregor, 1988; McMackin & Slovic, 2000). However, it is possible that people do not have strong affective reactions in these cases. Therefore, part of the reasons for this disagreement may be due to the type of decision being made. As mentioned previously, sports forecasting represents a domain of decision making in which people are likely to have an affective reaction and provides an objective criterion for which to compare their decision against. Therefore, in the present work we utilized a similar paradigm to Halberstadt and Levine (1999), and had participants make forecasts about upcoming sporting events.

Halberstadt and Levine (1999) randomly assigned participants to either rely on their gut feeling or think analytically about, and provide 3 reasons why, they thought one team, or another would win sweet-16 basketball games. They assessed participants' forecast accuracy in three ways. First, whether they correctly picked who would win the game. Second, the participants' forecasts about the margin of victory (how much they

would win by). Halberstadt and Levine (1999) included a third measure of the quality of participants' forecasts in order to speak to prior research (e.g., Wilson and Schooler, 1991) which assessed the quality of decision based on concordance with experts' judgments. They assessed how close participants' forecasts about a team's margin of victory were to Las Vegas *bookmaker's expectations*.

In betting terms, bookmakers' expectations are called the line scores or betting line. Prior to the start of a game, bookmakers release who they expect will win the game and the number of points the team is expected to win by. For instance, if bookmakers think that Team A will beat Team B by 5 points. Team A is the *favorite* by 5 points and Team B is the *underdog* and expected to lose by 5 points. Therefore, a gambler who places a bet on Team A to win will only receive a payout if Team A *exceeds expectations,* winning the game by more than 5 points. In contrast, a gambler who places a bet on Team B will receive a payout if Team B exceeds expectations either by winning the game outright or losing by less than 5 points. Given their goal (reducing the amount of money paid out to gamblers) and experience, bookmakers' expectations represent the best predictions about who will win a game.

Halberstadt and Levine (1999) observed that participants who used reasoned analysis to make their forecasts were less likely to pick the winning team to win. They were also less accurate at forecasting the number of points a team would win by and were less likely to make forecasts in line with experts' expectations. Combined, their results demonstrate that reasoned analysis impairs forecast accuracy.

Hypotheses

In the current study, we followed Halberstadt and Levine's (1999) procedures as closely as possible with the exception that we had participants make forecasts about National Football League (NFL) football games. We asked half of our participants to go with their gut feelings and the remaining participants to think about and provide 3 reasons why one team or the other would win (reasoned analysis). Based on the results of Halberstadt and Levine (1999) we made the following predictions.

1. Participants asked to make their forecasts based on a gut feeling will, on average, be more accurate at forecasting who will win the game than those asked to think about and provide 3 reasons for their forecast.
2. Participants asked to make their forecasts based on a gut feeling will, on average, be more accurate at forecasting the number of points by which a team will win (margin of victory) than those asked to think about and provide 3 reasons for their forecast.
3. Participants asked to make their forecasts based on a gut feeling will, on average, make forecasts closer to experts' expectations than those asked to think about and provide 3 reasons for their forecast.

As discussed earlier, prior research has found stable individual differences in peoples' tendency to rely on the two decision modes (Epstein, 1999) and that these individual differences predict decision quality (e.g., Raab & Laborder, 2011; Bruin et al., 2007). Beyond a main effect of an individual's tendency to use a decision mode, Betsch & Kuntz, (2008) demonstrated that individual tendencies interacted with their applied

decision mode. Therefore, we also seek to extend previous research by assessing the effect of individual differences on participants' forecast accuracy.

4. We predicted that participants who have a higher tendency for relying on gut feelings would be more accurate at forecasting the outcome of NFL games.
5. We predicted that participants who experience decisional fit when making their forecasts will, on average, make more accurate forecasts.

The current work also extends previous work on the benefit of going with your gut to examine its effect on confidence. Prior work has demonstrated that gut feelings come to mind more quickly and easily than responses formed from reasoned analysis (Kahneman & Frederick, 2002) and fluency is positively associated with confidence (Ackerman & Zalmanov, 2012). This led us to predict that:

6. participants asked to go with their gut feeling will, on average, be more confident about their forecasts than those asked first to think about and provide reasons for their forecasts.

As with accuracy, we predict a main effect of individual differences and an interaction with their applied decision mode for confidence. Therefore, we made the remaining two hypotheses.

7. Participants who tend to rely on their gut feelings when making decisions will be more confident about their forecasts than those who prefer to rely on reasoned analysis.
8. Participants who experience decisional fit when making their forecasts will, on average, be more confident about their forecasts.

Method

Study 1A was conducted during the first round of the 2021-22 NFL playoffs termed the wild card series. The NFL wild card series consists of six games between teams that did well enough to earn a shot at the playoffs but were not the number one team in their respective leagues (see Appendix D for a list of games played in the 2021-22 NFL wild card series). Three of these games are played between teams in the American Football Conference (AFC) and the remaining three between teams in the National Football Conference (NFC). For each conference's wild card games, the 2nd, 3rd, and 4th ranked teams play the 7th, 6th, and 5th respectively with the higher seeded team playing at their home stadium. The researchers chose the wild card games because they provided a good balance between the number of forecasts each participant was asked to make (6), and concerns about participant fatigue.

Initial pilot studies not reported here using undergraduate students who participated for course credit did not achieve our intended sample size. Therefore, in the current studies, the researchers recruited participants from the social media platform Reddit. Reddit is a discussion board-based media site that allows users, called Redditors, to organize around topics of interest. Reddit is divided into subreddits based on these topics. Of interest for the current research are those that focus on NFL football teams (e.g., r/buffalobills, r/GreenBayPackers), the NFL, and football more generally (e.g., r/nfl, r/football). Each subreddit has a team of moderators who oversee the discussions for that subreddit. Therefore, prior to posting a brief description of the research and a link to the study, we obtained permission from moderators for each of the subreddits. Some

moderators requested that we post the study advertisement in a specific thread, a subdivision of the subreddit that organizes discussions for the subreddit. These include threads for a particular day or game among others. Therefore, the location of the posts varied between subreddits. However, across subreddits, fans read similar study advertisements. For example, Reddit users who were fans of the Buffalo Bills read:

> "How well can Buffalo Bills fans predict this weekend's winners?
>
> (with chance to win a $25 Amazon gift card)
>
> My students and I are studying how well different teams' fans can pick the winners of the upcoming wild card games. If you have 8-10 minutes to give it a try, please go to http://tiny.utk.edu/picks.
>
> Volunteers will be entered into a drawing to win one of four $25 Amazon gift cards. We'll also post the results of the study online so you'll be able to see how well you did and how well your teams' fans did compared with other fans. (We conducted similar studies last week and are still sorting through the results. You can take part even if you took part last week.)
>
> (Please note that this sub's moderators gave us the OK to make this post.)"

The researchers changed the name of the team's fans to match the organization of the subreddit (e.g., football fans, Chiefs fans). All study materials were delivered to participants through a single link and delivered online via Qualtrics.

Participants

Interested Reddit users, (N = 457) self-selected to take part in the study, affirmed they were over the age of 18, and provided informed consent before beginning the study.

We excluded a total of 28% of participants who failed to meet our inclusion criteria. Specifically, we excluded participants who did not answer two trivia questions about the NFL correctly (11.6%). We also removed participants who took longer than 90 seconds to answer the trivia questions (0.6%). We removed participants if they reported taking a previous version of the study (13.8%), failed to make any forecasts (15.7%), or indicated that they consulted outside sources when making their forecasts (7.0%). Therefore, participants (N = 329) who made at least one forecast and met our inclusion criteria were included in the analyses. Our final sample had more participants in the gut feeling condition (n = 182) than in the reasoned analysis condition (n = 147). This will be addressed further in the results sections.

Overwhelmingly our sample identified as men (95.7%). The remaining participants identified as women (3.9%) or non-binary (0.3%). Our sample predominantly identified as solely White or European American (72.9%) followed by Asian or Asian American (10.6%), Hispanic or Latinx (6.1%), and Black or African American (1.2%). The remaining participants identified with either multiple ethnic identities or other (9.1%). Further, participants were 18 to 69 years of age (M = 29.4, SD = 8.5). As noted in the study advertisement, participants who completed the study were entered into a drawing for a chance to receive 1 of 4 $25 Amazon gift cards.

Materials and Procedure

Interested Reddit users affirmed that they were over the age of 18 before providing informed consent and beginning the study. To encourage participants not to seek outside information to make their forecasts, we told them that the accuracy of their

forecasts would not affect their probability of receiving a gift card. Participants answered two trivia questions about the NFL: “Who was named the NFL’s most valuable player in 2019?” (correct answer: Lamar Jackson, incorrect answers: Adrian Peterson, Barry Sanders, LaDainian Tomlinson) and “What position does the Rams’ Cooper Kupp play? (correct answer: wide receiver, incorrect answers: quarterback, running back, linebacker). Participants completed a demographics questionnaire in which they reported whether they had taken part in a previous version of the study, how much time they spent following NFL football in the past three months, and how closely they follow NFL football. Participants indicated their favorite team before reporting their gender and ethnic identity. Pilot studies indicated that participants in the reasoned analysis condition took longer to complete the study than participants in the gut feeling condition. As a result, there was evidence of differential attrition between the two conditions. Specifically, participants asked to think about and provide reasons for their forecasts were less likely to make forecasts about all the games we asked them to. In an attempt to reduce differential attrition, before we randomly assigned participants to either rely on their gut feeling or reasoned analysis when making their first forecast participants read:

> “Please note that you might be asked to answer several questions by typing a couple short sentences into text boxes when you’re making each of your picks. This might take some time and some people don't like typing out answers on their phones.

> If you do continue past this point, please make every effort to finish the entire survey. It can be difficult to interpret people's results when they quit halfway through."

Similar to the manipulation used by Halberstadt and Levine (1990), participants asked to rely on their gut feeling read the following prompt before making their forecasts:

> "We're interested in how well NFL football fans can predict who will win the upcoming wild card games. We're especially interested in the benefit of intuition in these sorts of judgments. Therefore, we would not like you to think about and analyze your reasons for your predictions. Instead, we'd like you to go with your intuition or "gut feeling". Try to avoid any drawn-out analysis of your decisions, and make your predictions based on your first instinct."

Alternatively, participants in the reasoned analysis condition read:

> "We're interested in how well NFL football fans can predict who will win the upcoming wild card games. We're especially interested in the benefit of reasoning processes in these sorts of predictions. Therefore, in order to prepare yourself for your predictions, we would like you to think about and analyze the reasons why you think one or the other team will win. We'll also ask you to type in the reasons you came up with. Ideally, we'd like you to come up with 3 reasons for each of your predictions."

Forecasts

For each wild card game, participants read the teams that would be playing each other, and which team was playing at home with the home team always listed first. The

instructions varied between conditions. For example, when making their prediction about the game between the Bengals and the Raiders participants in the gut feelings condition read:

> "The Bengals will be hosting the Raiders.
>
> Based on your gut feeling, how do you feel the game will turn out?
>
> (Please do not consult the internet or other resources when making your prediction.)"

At the bottom of the screen, they made their forecast by selecting the team they thought would win the game with the home team always listed first, and the number of points they thought they would win by on a slider scale from 0 to 50 points.

Before making their forecasts, participants in the reasoned analysis condition read:

> "The Bengals will be hosting the Raiders.
>
> At the bottom of this page, we'd like you to predict how you think the game will turn out. In the space below, please provide your reasons why you think that one team will win over the other.
>
> (Please do your best to provide 3 reasons, but please don't consult the internet or other resources when coming up with your reasons or making your prediction.)"

Below the instructions participants entered their reasons why they thought the game would turn out one way or another into a textbox before making their forecast at the bottom of the page using the same questions as participants in the gut feeling condition. We further recorded how long participants took to make a forecast for each game. The

order of the forecasts was fully randomized using the Qualtrics randomizer function for each participant.

Confidence Ratings

After making all 6 of their forecasts participants in both conditions rated how confident they were in each of their forecasts. Specifically, we reminded them who would be playing in each game and the team they forecasted would win. For example, participants in the gut feeling condition read, "You predicted that the Bengals will beat the Raiders. How likely do you *feel* it is that the Bengals will win?" Note that participants in the reasoned analysis condition saw the same prompt with the exception that we asked them "How likely do you *think* it is that the Bengals will win?" Participants responded by moving a slider scale from 50 (it was a complete guess) to 100 (I am completely confident).

Individual Differences

To test our hypotheses regarding individual differences and *decisional fit*, we measured participants' tendency to rely on the two decision making modes (gut feelings, reasoned analysis). Participants completed the three decision-making questionnaires discussed previously. Participants completed, in order, the Preference for Intuition and Deliberation Scale (PID; Appendix C; Betsch, 2004), the Rational Experiential Self Inventory – 40 item (REI-40; Pacini & Epstein, 1999), and the General Decision-Making Style (GDMS; Appendix B) rational and intuitive subscales (Scott & Bruce, 1995). The PID includes two subscales. The preference for intuition (PID-I) and preference for deliberation (PID-D) subscales. The REI-40 includes multiple subscales. However, of

interest for the current study are the rational and experiential subscales, the REI-R and REI-E respectively. Similarly, the GDMS (Appendix B) contains subscales for rational (GDMS-R) and intuitive decision making (GDMS-I). Participants completed the full PID and REI-40, and only the GDMS-I and GDMS-R subscales.

Additional Measures

At the end of the study, we asked participants how well they know NFL football, if they consulted the internet or an NFL playoff bracket, or if they received help when making their forecasts. As a manipulation check, we asked participants in both conditions "When making your predictions, how much did you rely upon your intuition or 'gut feelings'?" and "When making your predictions, how carefully did you think about reasons why one or the other of the two teams would do better?" on scales from 0 (not at all) to 4 (extremely). Finally, participants reported how difficult they found the task before receiving a debriefing and their unique completion code to look up their results after the study concluded. Participants interested in entering the Amazon gift card drawing or who wished to receive an email when the results were tabulated were sent to a separate survey to enter their contact information.

Results

Manipulation Checks

We first assessed the efficacy of our manipulation. We ran independent samples *t*-tests to assess whether our manipulation affected participants' reliance on their gut feeling and reasoned analysis when making their forecasts. As expected, participants asked to go with their gut feeling reported relying on their gut feelings more (M = 2.42,

SD = 1.05) than those asked to provide three reasons (*M* = 2.06, *SD* = .96), *t*(226) = 2.63, *p* = .0091, *d* = 0.35. On the other hand, participants asked to provide reasons (*M* = 2.82, *SD* = .86) did not think more carefully about the reasons why one team would win than participants asked to go with their gut (*M* = 2.64, *SD* = .94), *t*(226) = -1.45, *p* = .15, *d* = -0.19. Combined, our manipulation had only a small effect on how participants made their forecasts.

However, it is worth noting that this manipulation check may not be an accurate reflection of how participants made their forecasts in the current study. It is possible that responses to these questions overestimate the amount that participants used the two decisions modes. For one, participants may have been responding in line with how we asked them to make their forecasts and not how they made their forecasts (i.e., demand characteristics). It is also possible that participants were unaware of how they made their forecasts or did not draw the same distinctions between the two decision modes that we did. In such a case their responses to our manipulation check questions may under or overrepresent the degree to which they made their forecasts in line with how we asked them to.

Differential Attrition

In previous pilot studies not reported here, we had observed that participants asked to think about and provide reasons for their forecasts were less likely to make all the forecasts we asked them to. Before testing our hypotheses, we assessed whether there was differential attrition in the current study. We ran a *t*-test comparing the number of forecasts participants in the two conditions made with people who met the inclusion

criteria noted above with the exception that we included participants who failed to make any forecasts. Participants (n = 360) were assigned approximately equally to think about and provide reasons for their forecasts (n = 176) and to go with their gut feeling (n = 184). Of those asked to provide reasons, 16.5% failed to make any forecasts compared to only 1% of participants who were asked to go with their gut feeling. Further, fewer participants asked to provide reasons for their forecasts made all 6 forecasts (72%) compared to those asked to go with their gut feeling (97%). Overall, participants asked to think about and provide 3 reasons before making a forecast (M = 4.57, SD = 2.42) made less forecasts than those asked to rely on their gut feeling (M = 5.86, SD = .82), $t(358) = 6.84$, $p < .0001$, $d = 0.72$. Therefore, we observed evidence of differential attrition in the current study.

Given that participants asked to think about and provide reasons for their forecasts made fewer forecasts than those asked to go with their gut feeling, it increases the likelihood that there were systematic differences between the two groups. The task of thinking about and writing in reasons for why one team would win over another is more difficult than the task of forecasting the outcome of the games based on a gut feeling and took longer. The differential attrition we observed may have introduced a confounding variable. Specifically, it is possible that participants in the two conditions varied in *fanship* (how big of a fan someone is). Participants who were asked to think about and provide reasons for their forecasts and provided forecasts for all 6 games may be bigger fans of NFL football than participants in the gut feelings condition who made all 6 forecasts. Fanship is likely to be positively correlated with how accurate someone is at

forecasting the outcome of NFL games. We address this concern in exploratory analyses reported later. However, it is also possible that there are other systematic differences between the two groups that we are unable to address with the data we collected. Therefore, the results reported here should be taken in context of the differential attrition we observed.

Forecast Accuracy

Prior to running the analyses, we computed participants' average accuracy scores for each of the three accuracy measures utilized in Halberstadt and Levine (1999). The average number of correct picks was assessed by taking the number of forecasts for which the participant forecasted the correct team would win divided by the number of forecasts they made. For example, a participant who correctly forecasted the winning team for 2 of their 6 forecasts received a score of 0.33. In order to provide a more direct replication of Halberstadt and Levine, we first ran *t*-tests comparing participants in the two groups to test our hypothesis regarding main effects of experimental condition on the 3 measures of accuracy. See Appendix D for a summary of participants' responses across games.

To test Hypothesis 1, that participants in the gut feeling condition were more likely to pick the winning team, we ran a *t*-test comparing participants in the reasoned analysis and gut feeling conditions on the average number of correct forecasts. Contrary to our prediction, participants asked to go with their gut feeling (M = .75, SD = .16) did not pick the correct team more often than those asked to rely on their reasons (M = .74, SD = .19), $t(327) = -0.54$, $p = .59$.

We then assessed the effect of experimental condition on participants' accuracy about the margin of victory for each game. We calculated an average *misprediction score* for each participant as done in Halberstadt and Levine (1999). If a participant picked the winning team, we took the difference between the margin of victory they forecasted and the number of points the team won by. For example, during the NFL wild card series, the Bengals beat the Raiders by 7 points. If a participant forecasted the Bengals would beat the Raiders by 10 points, the forecast was given a misprediction score of 3 (10 – 7 = 3). Likewise, if a participant picked the losing team to win, we summed their margin of victory forecast with the real margin of victory. For example, if a participant forecasted that the Raiders would win by 10 their forecast has a misprediction score of 17 (10 + 7 = 17). As with the number of correct picks we then averaged their misprediction score by summing their scores for each forecast divided by the number of forecasts they made. We conducted a *t*-test to test Hypothesis 2, that participants asked to go with their gut feeling would be more accurate at forecasting the number of points a team would win by. Contrary to our prediction, participants asked to go with their gut feeling (M = 12.57, SD = 2.98) were not more accurate than those asked to rely on their reasons (M = 13.01, SD = 4.29), $t(327) = 1.09$ $p = .28$. Together, we observed that participants asked to go with their gut feeling were not more accurate in their forecasts than those asked to think about and provide reasons why one team would win over the other.

We then turned to the number of points participants' forecasts deviated from experts' expectations. To test Hypothesis 3, that participants in the gut feeling condition were more likely to forecast a team's margin of victory in line with experts' expectations,

we calculated the average number of points their forecasts deviated from experts' expectations. Difference scores were calculated in the same way as participants' *misprediction scores* with the exception that we used the line score as the comparison instead of the outcome of the game. Contrary to our prediction, participants asked to rely on their gut feeling (M = 5.77, SD = 3.69) did not make forecasts closer to experts' expectations as compared to those asked to think about and provide reasons for their forecasts (M = 6.21, SD = 3.88), $t(327) = 1.03$, $p = .30$.

Combined, our analyses failed to support our hypotheses, and the results observed in Halberstadt and Levine (1999), that making forecasts based on gut feelings makes people more accurate at picking who would win and the number of points they would win by. We further did not see any difference in how closely participants' forecasts were to experts' expectations. These findings cast doubt on the conclusion by Halberstadt and Levine (1999) that participants who go with their gut are more accurate at forecasting the outcome of sporting events than those asked to think about and provide reasons for their forecasts.

Individual Differences

We predicted in Hypothesis 4 that participants who tend to rely on their gut feelings would be more accurate, and in Hypothesis 5 that those who experience decisional fit would be more accurate than those who did not have decisional fit when making their forecasts. We measured participants' tendency to rely on their gut feelings and reasoned analysis with three individual difference measures (PID, REI, and GDMS).

To account for missing responses on the individual difference measures we calculated participants' average score for each of the 6 individual difference subscales.

We first assessed whether the relationship between the individual difference measures' subscales was consistent with prior research. In line with Cook and Gonzales (2016) we observed that the correlations between the three subscales used to measure participants' tendencies for making decisions based on gut feelings were highly correlated, but those assessing preferences for reasoned analysis were only moderately correlated (Table 2.1). We further examined the relationship between peoples' tendency to rely on gut feelings and reasoned analysis within scales. We observed the expected negative relationship between subscales in the REI and GDMS (Table 2.1; Pacini & Epstein, 1999; Scott & Bruce, 1995). In contrast, however, we observed a non-significant relationship between subscales for the PID (Table 2.1). Together, the results supported our assumption that there are slight differences between the three scales used to measure individual differences in reliance on gut feelings and reasoned analysis.

Based on this evidence, we assessed our hypotheses regarding individual differences separately for the 6 subscales measuring people's preferred decision making mode and participants' three accuracy scores. We submitted the data to cross-classified mixed models in R (R Core Team, 2021) including participants tendency to rely on a decision mode, condition, and the condition by tendency interaction as fixed effects and game and

Table 2.1 Study 1A Correlations Between Individual Difference Measures

	Reliance on gut feelings			Reliance on reasoned analysis		
	PID	REI	GDMS	PID	REI	GDMS
Reliance on gut feelings						
PID		.67**	.61**	.02		
REI			.72**		-.16*	
GDMS						-.22**
Reliance on reasoned analysis						
PID					.34**	.43**
REI						.47**
GDMS						

Note: Reported correlations are bivariate correlations between individual difference measure subscales.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

participant number as random intercepts[2]. Condition was contrast coded (gut feelings (-1) vs. reasoned analysis (1)) and participants' scores on the individual difference measures were mean centered prior to model fitting. If participant's tendency to rely on the two decision modes moderates the relationship between experimental condition and accuracy, their individual difference subscale score by condition interaction term would significantly predict accuracy.

We first examined whether participants' tendency to rely on reasoned analysis, as measured by the REI, affected their accuracy. There was a main effect of how much someone tends to rely on reasoned analysis on both participants misprediction scores, b = -1.01, se = 0.35, $t(231)$ = -2.87, p = .0045 and the number of points their forecasts deviated from the experts, b = -1.46, se = 0.44, $t(231)$ = -3.29 p = .0012. Reasoners were more accurate at forecasting the number of points a team would win by and made forecasts closer to experts' expectations (Table 2.2). However, these significant effects were not replicated in the remaining analyses in the current study (Table 2.2) and were inconsistent in subsequent studies reported later.

Scores on the REI-R moderated the relationship between experimental condition and participants' misprediction scores. In line with our prediction, participants who preferred to use reasoned analysis made more accurate forecasts when they were asked to first think about and provide 3 reasons before making their forecast $t(231)$ = -2.02, p =

[2] We initially included a random slope for the condition by preference interaction across the games participants made forecasts about, however, we experienced singularity issues. Therefore, we simplified the models and only included trial and participant number as random intercepts for the models including the PID-I and REI-R. Models with the PID-R, REI-E, and GDMS also did not have the random intercept for participant number as the models had singularity issues. Degrees of freedom for all mixed models reported are estimated using the Satterthwaite approach (Satterthwaite, 1946).

Table 2.2 Study 1A Individual Differences and Forecast Accuracy

		Correct Pick			Misprediction Score			Deviation from Experts		
	Predictor	*b*	*se*	*z*	*b*	*se*	*t*	*b*	*se*	*t*
PID-I										
	(Intercept)	1.47	0.52	2.82**	12.62	3.36	3.76**	5.87	0.28	20.92***
	Condition	-0.08	0.07	-1.15	0.24	0.18	1.39	0.31	0.22	1.38
	Intuition	-0.13	0.12	-1.09	0.14	0.32	0.45	0.10	0.41	0.25
	Intuition*Condition	0.09	0.12	0.80	-0.03	0.32	-0.08	-0.63	0.41	-1.56
PID-D										
	(Intercept)	1.47	0.52	2.82**	12.62	3.36	3.76**	5.89	0.28	20.98***
	Condition	-0.07	0.07	-1.11	0.24	0.18	1.37	0.32	0.22	1.44
	Reasoning	-0.02	0.12	-0.14	0.06	0.32	0.84	-0.55	0.40	-1.37
	Reasoning*Condition	-0.01	0.12	-0.07	-0.20	0.32	-0.64	-0.51	0.40	-1.26
REI-E										
	(Intercept)	1.50	0.52	2.90**	12.65	3.30	3.84**	5.85	0.32	18.24***
	Condition	-0.08	0.07	-1.17	0.29	0.20	1.46	0.42	0.25	1.66
	Intuition	-0.19	0.14	-1.43	0.06	0.37	0.17	0.45	0.47	0.95
	Intuition*Condition	-0.06	0.14	-0.46	0.06	0.37	0.16	-0.34	0.47	-0.73
REI-R										
	(Intercept)	1.49	0.52	2.87**	12.71	3.31	3.85**	5.88	0.31	18.96***
	Condition	-0.10	0.07	-1.45	0.36	0.19	0.06	0.53	0.24	2.20*
	Reasoning	0.15	0.13	1.15	-1.01	0.35	-2.87**	-1.46	0.44	-3.29**
	Reasoning*Condition	0.12	0.13	0.91	-0.71	0.35	-2.02*	-0.41	0.44	0.36
GDMS-I										
	(Intercept)	1.47	0.52	2.81**	12.71	3.32	3.83*	5.78	0.34	16.95***
	Condition	-0.08	0.07	0.29	0.24	0.20	1.20	0.43	0.26	1.65
	Intuition	-0.07	0.10	0.48	0.24	0.27	0.89	0.63	0.35	1.81
	Intuition*Condition	-0.02	0.10	0.84	0.28	0.27	0.94	-0.10	0.35	-0.29
GDMS-R										
	(Intercept)	1.48	0.53	2.81**	12.75	3.32	3.84*	5.79	0.34	16.88***
	Condition	-0.09	0.07	-1.27	0.29	0.20	1.40	0.47	0.26	1.79
	Reasoning	0.21	0.12	1.85	-0.52	0.32	-1.63	-0.53	0.41	-1.27
	Reasoning*Condition	-0.01	0.12	-0.04	-0.49	0.32	-1.51	-0.15	0.41	-.036

Note: The results are from separate cross-classified mixed models assessing the effect of individual differences on the three dependent measures.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

However, this finding was not supported by subsequent analyses examining moderation by participants' tendency to rely on reasoned analysis as measured by the PID, $t(275) = -0.64$, $p = .53$, $b = -0.20$, $se = .32$, and the GDMS $t(216) = -1.51$, $p = .13$, $b = -0.49$, $se = .32$. Nor did subsequent studies provide further support that participants reliance on gut feelings moderated the relationship between experimental condition and accuracy. Likewise, participant's tendency to rely on gut feelings when making decisions did not moderate the relationship between experimental condition and misprediction scores (REI $b = .061$, $se = .37$, $t(227) = .16$, $p = .87$, PID $b = -0.03$, $se = .32$, $t(275) = -0.08$, $p = .93$, GDMS $b = 0.26$, $se = .27$, $t(216) = 0.95$, $p = .35$ (Table 2.2)

Forecast Confidence

Despite observing no differences in terms of accuracy, it is possible that our manipulation may still have affected participants' confidence in their forecasts. Regardless of accuracy, people are still likely to feel more confident about their forecasts which are based on a gut feeling. Therefore, we turned to test hypothesis 6 which stated that participants in the gut feeling condition would feel more confident about their forecasts than participants in the reasoned analysis condition.

We first removed participants who failed to make at least one confidence judgment (1.8%). As with our analyses regarding forecast accuracy, we first calculated participants' average confidence ratings for the games they made forecasts about.[3] Contrary to our prediction participants asked to make their forecasts based on their gut

[3] There was a data collection error for the first 6% of participants for the game between the Chiefs and Steelers. For these participants, they saw the team they forecasted would lose the game and were asked how confident they were that the team would win the game. Therefore, we removed these confidence judgments before analyzing participants' average confidence.

feeling condition were not more confident about their forecasts (M = 73.76, SD = 7.95) than those asked to think about and provide 3 reasons for their forecasts (M = 74.11, SD = 9.97), $t(302) = -0.34$, $p = .73$.

We then assessed whether individual differences (Hypothesis 7) and decisional fit affected participants' confidence in their forecasts (Hypothesis 8). We submitted the data to cross-classified mixed models including condition, preference, and the condition by preference interaction as fixed effects and participant and game as random effects[4] (Table 2.3). We ran separate models for each of the 6 individual difference measure subscales. Participants' who had a higher tendency to rely on gut feelings on the REI-E were more confident in their forecasts $b = 2.03$, $se = 0.97$, $t(225.42) = 2.10$, $p = .0367$. However, reliance on gut feelings on the PID and GDMS did not provide further support for this result. We next looked for an interaction between individual differences and experimental condition. A significant interaction term between preferences for relying on a gut feeling and reasoned analysis by condition would indicate that decisional fit affected participants confidence in their forecasts. However, contrary to our prediction, none of the interaction terms were significant predictors of confidence across the 3 individual difference measures (Table 2.3). Therefore, we did not see evidence that decisional fit improved participants' confidence in their forecasts.

[4] We initially included a random slope for the condition by preference interaction for the games participants made forecasts about. However, as with our models assessing the accuracy of participants' forecasts, we experienced singularity issues. Therefore, we simplified the models and included only participant and game as random intercepts.

Table 2.3 Study 1A Individual Differences and Forecast Confidence

		Forecast Confidence		
	Predictor	*b*	*se*	*t*
PID-I				
	(Intercept)	74.43	4.18	17.82***
	Condition	0.66	0.48	1.37
	Intuition	1.12	0.87	1.28
	Intuition*Condition	1.22	0.87	1.40
PID-D				
	(Intercept)	74.44	4.18	17.81***
	Condition	0.65	0.48	1.34
	Reasoning	0.37	0.87	0.43
	Reasoning*Condition	-1.01	0.87	-1.16
REI-E				
	(Intercept)	74.12	4.09	18.12***
	Condition	0.76	0.52	1.48
	Intuition	2.02	0.97	2.10*
	Intuition*Condition	1.01	0.97	1.04
REI-R				
	(Intercept)	74.25	4.09	18.16***
	Condition	0.82	0.51	1.59
	Reasoning	1.49	0.94	1.59
	Reasoning*Condition	-0.14	0.94	-0.15
GDMS-I				
	(Intercept)	74.06	4.05	18.30***
	Condition	0.87	0.54	1.63
	Intuition	0.19	0.72	0.26
	Intuition*Condition	0.47	0.72	0.65
GDMS-R				
	(Intercept)	74.07	4.05	18.31***
	Condition	0.77	0.53	1.44
	Reasoning	1.51	0.85	1.78
	Reasoning*Condition	-0.17	0.85	-0.20

Note: The results are from separate cross-classified mixed models for each individual difference subscale.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

Exploratory Analyses

Our analyses failed to replicate the results of Halberstadt and Levine (1999). Participants in our study who were asked to go with their gut feeling were not more accurate at forecasting the outcome of NFL games. To further explore our participant's forecasts, we ran the analyses concerning the main effect of experimental condition on participants' accuracy and confidence with all participants (N = 385) who made at least one forecast.

Participants in the gut feeling condition did not select the winning team (M = .75, SD = .16) more often than participants in the reasoned analysis condition (M = .74, SD = .19), $t(383) = 0.50$, p = .62. They were also not more accurate (M = 13.09, SD = 4.05) at forecasting the margin of victory than those in reasoned analysis condition (M = 13.46, SD = 5.35), $t(383) = -0.76$, p = .45. Participants in the gut feelings condition were also not more likely to make forecasts in line with the expert's predictions (M = 6.70, SD = 5.67) compared to participants in the reasoned analysis condition (M = 7.00, SD = 5.64), $t(383) = -0.50$, p = .62. These results further support our conclusion that going with your gut does not improve forecast accuracy. We next re-examined participants' confidence in their forecasts (Hypothesis 7). Participants in the gut feelings condition were not more confident (M = 73.54, SD = 8.10) in their forecasts $t(354) = -0.86$, p = .39 than those in the reasoned analysis condition (M = 74.34, SD = 9.53).

As noted, we again we saw no evidence supporting the conclusion that our manipulation influenced the accuracy of participants' forecasts or their confidence in their forecasts. However, the results reported thus far may have been affected by the

differential attrition we observed. Participants who were asked to think about and provide 3 reasons why they thought one team would win over the other before making their forecasts were less likely to make all 6 forecasts than those asked to rely on their gut feeling about who would win. It is possible that participants who made all 6 forecasts in the reasoned analysis condition were bigger fans than those who made all 6 forecasts in the gut feelings condition.

To address this concern, we compared participants who were asked to think about and provide 3 reasons with those asked to go with their gut on their reported knowledge of the NFL, the time they spend following the NFL, and how closely they follow the NFL. We first looked at everyone who met our inclusion criteria. If differential attrition created unequal groups, we would expect that participants in the reasoned analysis condition would give higher ratings on all 3 measures of how much they follow the NFL than those in the gut feeling condition. Among the participants who provided at least one forecast, those in the reasoned analysis condition did not report spending more time following the NFL, following it more closely, or that they knew more about the NFL (all ts < .43, all ps > .50).

Given that participants in the reasoned analysis condition were not bigger fans than those in the gut feelings condition this reduces the likelihood that differential attrition created systematic differences between the two groups. Just as fanship may have weakened the relationship between experimental condition and accuracy, other variables that we did not measure may have affected our results. For instance, participants' motivation to complete the study might have differed between the two conditions.

Participants in the reasoned analysis condition who made all 6 forecasts may have been more motivated to complete the study than those in the gut feelings condition who made all 6 forecasts.

To further explore the data, we ran correlation analyses to determine if any of other measured variables influenced forecast accuracy and confidence. We first assessed whether participants knowledge about the NFL and how closely they follow the NFL was related to their forecast accuracy and confidence (Table 2.4). We ran the analyses on only those participants who met our inclusion criteria. We included participants' self-reported knowledge of the NFL, how closely they follow the NFL, and how much time they spend watching the NFL. Our primary outcome variables of interest were participants' average misprediction score, their average deviation from experts' predictions and their average confidence in their forecasts.

Somewhat surprisingly participant's reported knowledge of the NFL was only related to participant's confidence and not their misprediction score and how much their forecasts deviated from expert's expectations (Table 2.4). How closely participants follow the NFL and the amount of time they spend following the NFL showed only small correlations with accuracy and confidence (Table 2.4). Finally, we assessed whether the order in which participants made their forecasts and how long they spent making their forecasts affected their accuracy and confidence.

The order in which participants made their forecasts was negatively related to their confidence, $r(1777) = -0.05$, $p = .03$. However, forecast order was unrelated to participants misprediction score, $r(1882) = 0.01$, $p = .64$, or how much their forecasts

Table 2.4 Study 1A Correlations Between Measures of Fanship and Dependent Measures

Variable	*M*	*SD*	NFL Knowledge	Follow NFL (Time)	Follow NFL (Closeness)	Misprediction Score	Deviation from Experts
NFL Knowledge	7.64	1.10					
Follow NFL (Time)	4.85	0.98	.26**				
Follow NFL (Closeness)	3.02	0.81	.51**	.52**			
Misprediction Score	12.77	3.62	.05	.13*	.11*		
Deviation from Experts	5.97	3.78	.09	.13*	.09	.51**	
Confidence	73.91	8.83	.17**	.16**	.12*	.02	-.07

Note: Reported relationships are bivariate correlations.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

deviated from expert's expectations, $r(1882) = 0$, $p = .93$. To test the relationship between forecast order and accuracy and confidence we computed the average amount of time each participant took to make their forecasts looked at correlations separately for participants asked to rely on their gut feelings and reasoned analysis. How long it took participants in the gut feelings condition and reasoned analysis condition to make their forecasts was unrelated to their accuracy and confidence (Table 2.5).

Discussion

The results of Study 1A failed to support our hypotheses and replicate the results of Halberstadt and Levine (1999). In the current study, asking people to go with their gut feelings did not improve forecast accuracy in terms of the teams they picked or their forecast about the team's margin of victory. Similarly, we did not see evidence that participants in the gut feelings condition were more likely to make forecasts closer to those made by expert sports forecasters. In terms of individual difference and accuracy, participants who had a higher tendency for engaging in reasoned analysis as measured by the REI (REI-R) were more accurate in terms of the margin of victory but not who they thought would win. Scores on the REI-R also moderated the relationship between manipulated forecast mode and accuracy. However, given that the relationship was not observed for participants reliance on reasoned analysis as measured by the PID and GDMS we are hesitant to interpret this finding as supporting out hypotheses. In terms of participant's confidence, there was no effect of our manipulation, individual preferences nor an interaction between the two.

While these results cast doubt on our predictions, these results should be taken in

Table 2.5 Study 1A Correlations Between Forecast Time and Dependent Measures

	Forecast Time	
Variable	Gut Feelings	Reasoned Analysis
Misprediction Score	.11	-.01
Deviation from Experts	.04	.15
Confidence	.05	.01

Note: The correlations reported in Table 2.5 are the bivariate correlations. None of the relationships were significant at the .05 level.

light of the differential attrition observed between people asked to think about and provide reasons before making their forecasts and those asked to rely on a gut feeling. Participants asked to provide reasons for their forecasts were less likely to make all 6 forecasts. While we did not see evidence that this created unequal groups in terms of how big of fans participants were, it is possible that other differences exist between the two groups. This differential attrition also affects the individual difference measures which were included at the end of the study. Participants in the reasoned analysis condition who withdrew before the end of the study did not complete the individual difference measures. As an exemplar, 185 people in the gut feeling condition completed all measures of the PID while only 146 in the reasoned analysis condition did so. Therefore, despite these null findings, it is still possible that relying on gut feelings improves forecast accuracy. In subsequent studies we attempt to address this limitation by manipulating how participants make their forecasts as a within-subjects variable (Studies 1B and 2B).

CHAPTER THREE

Study 1B

In Study 1A we did not replicate the results obtained by Halberstadt and Levine (1999). However, in Study 1A many of the participants failed to make all 6 forecasts about the wild card games. Specifically, people in the reasoned analysis condition were less likely to make all 6 forecasts than those in the gut feelings condition. To reduce differential attrition and increase power, in Study 1B we manipulated participants forecast mode (gut feeling, reasoned analysis) as a within-subjects variable. In the current study, participants made forecasts about the same 6 games in the 2022 NFL wild card series as in Study 1A. All hypotheses were the same as in Study 1A.

Method

Participants

Participants, (N = 476) self-selected to take part in a study posted on the same football related subreddits in Study 1A. Participants who completed the study were entered into a drawing for a chance to receive 1 of 4 $25 Amazon gift cards. As in Study 1A. we excluded participants who failed to meet our inclusion criteria (27.3%). We excluded participants who did not answer the two trivia question about the NFL correctly (10.1%) or took longer than 90 seconds to answer them (3.6%). We removed participants who reported taking a previous version of the study (3.4%), failed to make any forecasts (15.7%), or indicated that they consulted outside sources when making their forecasts (4.6%). Therefore, 346 participants who made at least one forecast and met our inclusion criteria were included in the analyses. Overwhelmingly, our sample identified as men (94.5%). The remaining participants identified as women (4.6%), or non-binary or other

(< 1%). Our sample predominantly identified as solely White or European American (77.2%) followed by Asian or Asian American (10.4%), Hispanic or Latinx (4.6%), Black or African American (1.5%), and Native American (< 1%). The remaining participants identified with either multiple ethnic identities or other (7.2%). Further, participants were 19 to 80 years of age (M = 28.3, SD = 7.6).

Materials and Procedures

In Study 1B, participants made forecasts about the same 6-game NFL wild card series as in Study 1A. The materials and procedures for Study 1B are the same as Study 1A with a few important exceptions. Study 1B has a mixed design with condition (gut feeling, reasoned analysis) as a within-subjects variable. Participants completed the same trivia and demographics questions as in Study 1A. Further, participants read the same warning to make every effort to complete the study if they continued to the forecast portion of the study. However, unlike Study 1A, we randomly assigned participants to make 3 of the 6 forecasts by relying on their gut feeling and the remaining forecasts after thinking about and providing 3 reasons why they thought one team would win over the other. Prior to starting their forecasts, participants read,

> "We're interested in how well NFL football fans can predict who will win the upcoming wild card games. In making your predictions, we'd like you to use different strategies for some games than others.
>
> **Intuition-based predictions**
>
> For three of your predictions, we'll ask you to "go with your gut" and choose the team that you feel will win. Rather than coming up with reasons for your

predictions, we just want you to go with your first instinct.

Reasoning-based predictions

For the other three predictions, we'll ask you to think about and analyze the reasons why you think one or the other of the teams will win. We'd also like you to type in the reasons you came up with. Ideally, we'd like you to come up with at least 3 reasons for each of your predictions."

Participants completed the 6 forecasts in random order before reporting their level of confidence about each of their forecasts in the same manner as Study 1A. At the end of the study, we altered the manipulation check questions slightly from Study 1A. In the current study, we asked participants "When we asked you to rely on your intuition, how much did you rely upon your intuition or 'gut feelings'" and "When we asked you to rely on reasoning, how carefully did you think about reasons why one or the other of the two teams would do better?" Participants responded to both questions using a scale from 0 (not at all) to 4 (extremely). Since we did not ask participants to report the degree to which they used both decision modes in each condition we cannot compare their responses across conditions as we did in Study 1A. All other methods and procedures were the same as Study 1A.

Results

Manipulation Checks

To check whether participants followed the instructions we gave them on how to make their forecasts, we first assessed their responses to the two manipulation check questions. Participants on average reported that they followed instructions. That is,

participants reported on average relying on their gut feeling to make forecasts in which they were asked to do so (M = 2.59, SD = 0.95) above the midpoint of the scale, $t(473)$ = 13.48, $p < .0001$. Similarly, participants reported carefully thinking about their reasons why they thought one team would win over the other when asked to do so at (M = 2.91, SD = 0.79) above the midpoint of the scale $t(467)$ = 24.98, $p < .0001$. Therefore, participants at least reported following our instructions about how to make their forecasts.

Differential Attrition

In Study 1A we observed that participants asked to think about and provide reasons before making their forecasts were less likely to make forecasts about all 6 games. Therefore, before assessing how accurate participants were at making their forecasts, we first looked at the number of forecasts participants made while relying on their gut feeling and engaging in reasoned analysis. Some participants (10%) failed to make at least one forecast in both conditions. The remaining participants made slightly more forecasts relying on their gut feelings (M = 2.92, SD = 0.37) than relying on reasoned analysis (M = 2.87, SD = 0.46), $t(314)$ = 3.02, p = .0027, d = 0.17 (Table 3.1). Therefore, manipulating how participants made their forecasts as a within-subjects variable did not completely eliminate differential attrition. However, it substantially reduced the amount of differential attrition observed compared to Study 1A.

Forecast Accuracy

Overall, participants picked the winning team approximately the same as in Study 1A (M = .79, SD = .41). See Appendix D for a summary of participants responses across games. As with Study 1A we computed their average accuracy scores for the number of

Table 3.1 Study 1B Number of Forecasts by Condition

	Number of Forecasts		
Trial Condition	1	2	3
Gut Feeling	3%	2%	95%
Reasoned Analysis	5%	3%	92%

Note: The frequencies reported here represent the percent of total forecast for each condition.

times they correctly picked the winner, their misprediction score, and the number of points their forecasts deviated from experts' expectations. Contrary to our prediction, participants did not pick the winning team more often when relying on their gut feeling (M = .77, SD = .25), than when asked to provide reasons for their forecasts (M = .76, SD = .25), $t(314) = 0.15$, p = .88. Similarly, participants were not more accurate at forecasting the margin of victory when asked to go with their gut (M = 12.03, SD = 5.29) than when asked to provide reasons (M = 12.26, SD = 5.28), $t(314) = -0.48$, p = .64.We then compared participant's forecasts with expert's predictions. Contrary to our prediction, participants did not make forecasts about the margin of victory closer to experts' predictions when asked to go with their gut (M = 5.65, SD = 3.30) than when asked to provide reasons (M = 5.83, SD = 3.86), $t(314) = -0.82$, p = .41. While our data failed to support our hypotheses based on Halberstadt and Levine's (1999) results, they are in line with Study 1A where we observed no effect of our manipulation on the accuracy of participant's forecasts.

Individual Differences

To determine whether individual difference and decisional fit affected participant's accuracy about the outcome of the NFL wild card games, we submitted the data to cross-classified mixed models. As with Study 1A we ran separate models for each of the 3 individual difference measure's subscales (Table 3.2). This resulted in a total of 6 analyses for each of the three measures of accuracy. We included condition, participants reliance on gut feelings and reasoned analysis, and the condition by reliance interaction

Table 3.2 Study 1B Individual Differences and Forecast Accuracy

		Correct Pick			Misprediction Score			Deviation from Experts		
	Predictor	*b*	*se*	*z*	*b*	*se*	*t*	*b*	*se*	*t*
PID-I										
	(Intercept)	1.34	0.37	3.64***	12.16	3.23	3.76*	5.79	0.40	14.63***
	Condition	-0.03	0.06	-0.52	0.20	0.16	1.22	0.07	0.11	0.68
	Intuition	0.17	0.11	1.52	-0.11	0.30	-0.38	0.46	0.33	1.41
	Intuition*Condition	-0.11	0.11	-0.99	0.38	0.30	1.27	-0.15	0.20	-0.74
PID-D										
	(Intercept)	1.33	0.37	3.65***	12.16	3.23	3.76*	5.79	0.40	14.66***
	Condition	-0.03	0.06	-0.48	0.20	.16	1.22	0.07	0.11	0.68
	Reasoning	0.01	0.12	0.09	-0.15	0.31	-0.49	-0.52	0.34	-1.53
	Reasoning*Condition	0.07	0.12	0.59	-0.31	0.31	-1.00	-0.36	0.21	-1.72
REI-E										
	(Intercept)	1.33	0.37	3.63***	12.18	3.25	3.75*	5.65	0.41	13.82***
	Condition	-0.02	0.07	-0.31	0.17	0.18	0.95	-0.01	0.12	-0.07
	Intuition	0.15	0.12	1.20	-0.14	0.32	-0.44	-0.39	0.32	-1.21
	Intuition*Condition	-0.07	0.12	-0.60	0.04	0.33	0.12	-0.02	0.22	-0.08
REI-R										
	(Intercept)	1.34	0.37	3.63***	12.14	3.22	3.77*	5.70	0.41	14.01***
	Condition	-0.02	0.06	-0.24	0.13	0.17	0.73	0.00	0.12	-0.02
	Reasoning	-0.03	0.12	-0.26	0.13	0.33	0.39	-0.63	0.34	-1.86
	Reasoning*Condition	0.05	0.12	0.44	-0.04	0.33	-0.10	-0.12	0.22	-0.53
GDMS-I										
	(Intercept)	1.32	0.36	3.64***	12.16	3.24	3.75*	5.66	0.43	13.31***
	Condition	-0.01	0.07	-0.14	0.18	0.18	1.00	-0.03	0.12	-0.28
	Intuition	0.07	0.09	0.76	-0.09	0.25	-0.34	-0.02	0.25	-0.09
	Intuition*Condition	0.05	0.09	0.48	-0.31	0.25	-1.23	-0.27	0.17	-1.56
GDMS-R										
	(Intercept)	1.32	0.36	3.64***	12.16	3.24	3.76*	5.66	0.43	13.31***
	Condition	-0.01	0.07	-0.16	0.18	0.18	0.32	-0.03	0.12	-0.28
	Reasoning	-0.08	0.10	-0.78	0.23	0.27	0.38	-0.08	0.26	-0.30
	Reasoning*Condition	0.00	0.10	-0.02	-0.06	0.27	0.83	0.01	0.18	0.03

Note: We conducted separate cross-classified mixed models for each of the individual difference subscales. *$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

as fixed effects and participant and game as random effects[5] (Table 3.2). Across the 6 subscales assessing participants' tendency to rely on gut feelings and reasoned analysis we observed no effect of individual differences on the accuracy of their forecasts (Table 3.2).

If participant's tendency to rely on gut feelings and reasoned analysis moderated the relationship between experimental condition and accuracy, interaction term for subscale by condition would have significantly predict accuracy. However, none of the interaction terms or lower order effects significantly predicted participant's accuracy. Therefore, we do not see evidence that *decisional fit* affected participant's accuracy about who would win each game, the number of point they would win by, or how close their forecasts were to experts' expectations about the outcome of the game (Table 3.2).

Forecast Confidence

We then turned to whether participants were more confident in their forecasts when they relied on their gut feelings as opposed to first when they were asked to engage in reasoned analysis when making their forecasts. Contrary to our predication, participants were not more confident about forecasts in which they were asked to go with their gut (M = 74.68, SD = 9.54), $t(287) = -1.38$, $p = .17$ than when asked to provide reasons (M = 75.68, SD = 9.45).

[5] We initially included a random slope for the condition by preference interaction for the games participant's made forecasts about, however, we experienced singularity issues. Similarly, we experienced singularity issues including participant as a random intercept. The random intercept for participant accounted for no variance and was subsequently removed from the models assessing whether a participant picked the correct team and misprediction scores. Therefore, we simplified the models and only included trial as a random intercept for the models assessing whether a participant correctly picked which team would win and misprediction scores. However, we included both a random intercept for game and participant when assessing deviation from experts' expectations.

To determine whether participant's tendency to rely on gut feelings and reasoned analysis affected their confidence in their forecasts we ran cross-classified mixed models[6]. In line with our prediction, reliance on intuition was a significant predictor of confidence for all three individual difference measures (Table 3.3). Participants who preferred relying on gut feelings more were more confident in their predictions. This is somewhat surprising given the null results in Study 1A. However, this effect was also not replicated in subsequent within-subjects studies (Study 2B and 3). In terms of decisional fit, only the interaction term for reliance on intuition as measured by the GDMS by experimental condition was significant $t(224.31) = 2.21$, $p = .03$, $b = 1.04$, $se = 0.47$. This effect was not replicated in the other 2 individual difference measures (Table 3.3). Therefore, we did not take this as support for our hypothesis that decisional fit would affect participant's confidence.

Discussion

Manipulating forecast strategy as a within-subjects variable produced similar null results as those in Study 1A in terms of accuracy. Specifically, participants were not more accurate in forecasting who would win or the margin of victory when relying on their gut feeling compared to when they were asked to think about and provide 3 reasons for their forecasts. Similarly, when asked to go with their gut feeling, participants did not make forecasts closer to experts' expectations. Therefore, the results did not replicate the results of Halberstadt and Levine (1999). One benefit of manipulating how participants

[6] We experienced singularity issues attempting to include a random slope for the interaction term for condition by preferences. The models for PID and GDMS included only a random slope for condition by trial and condition by participant. We experience convergence issues assessing the same model for the REI. Therefore, we reduced the model to only include a random slope for experimental condition by participant.

Table 3.3 Study 1B Individual Differences and Forecast Confidence

		Forecast Confidence		
	Predictor	*b*	*se*	*t*
PID-I				
	(Intercept)	75.35	3.62	20.80***
	Condition	0.43	0.48	0.88
	Intuition	3.67	0.76	4.85***
	Intuition*Condition	0.31	0.54	0.57
PID-D				
	(Intercept)	75.35	3.62	20.80***
	Condition	0.43	0.47	0.91
	Reasoning	-0.33	0.82	-0.40
	Reasoning*Condition	0.45	0.56	0.80
REI-E				
	(Intercept)	75.07	0.44	168.92***
	Condition	0.71	0.40	1.78
	Intuition	2.76	0.82	3.35***
	Intuition*Condition	0.63	0.74	0.852
REI-R				
	(Intercept)	75.19	0.45	165.72***
	Condition	0.66	0.40	1.67
	Reasoning	-0.51	0.85	-0.60
	Reasoning*Condition	-0.20	0.74	-0.27
GDMS-I				
	(Intercept)	74.94	3.58	20.95***
	Condition	0.37	0.55	0.66
	Intuition	2.97	0.62	4.80***
	Intuition*Condition	1.04	0.47	2.21*
GDMS-R				
	(Intercept)	74.94	3.58	20.95***
	Condition	0.37	0.54	0.68
	Reasoning	0.21	0.68	0.31
	Reasoning*Condition	0.06	0.50	0.13

Note: The results are from separate cross-classified mixed models assessing the effect of individual differences on confidence.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

made their forecasts as a within-subjects variable is that we increased the power of our study. Given the large number of participants (N = 315), if going with a gut feeling increased the accuracy of sports forecasts, we should have observed it in the current study.

However, manipulating how participants made their forecasts in a within-subjects design has several drawbacks. First, it increases the likelihood that participants guessed the true purpose of the experiment and makes the results reported here more susceptible to demand characteristics. Second, it is possible that participants found it difficult to switch between relying on a gut feeling and reasoned analysis when making their forecasts. Specifically, it is possible that after making a forecast for which they were asked to think about and provide reasons, they may have had difficulty relying solely on a gut feeling on subsequent forecasts. However, given that we observed similar null results in Study 1A it seems more likely that engaging in reasoned analysis does not impair forecast accuracy and nor does relying on gut feelings improve forecast accuracy. We include a further discussion of this in the general discussion for Study 1.

As with accuracy, we observed no effect of our manipulation on confidence. However, as predicted, we did observe that participants' tendency to rely on their gut feelings was positively related to their confidence. It is unclear why we only observed this effect in Study 1B. Given that we do not replicate this study in future within-subjects studies we are hesitant to interpret this as supporting our hypothesis.

Despite the positive effect of preferences for gut feelings, we did not see the anticipated effect of decisional fit on confidence. This is not particularly surprising given

that we saw no main effect of our experimental manipulation on confidence. In order for an interaction to occur there would have to have been a crossing interaction between experimental condition and reliance on gut feelings and reasoned analysis. Therefore, in context of the null effect of our experimental manipulation this is less surprising.

In sum, in Study 1B we observed that preferences for relying on gut feelings increases forecast confidence. However, the results did not replicate prior research which has demonstrated that relying on gut feelings improves forecast accuracy. Likewise, individual preferences for relying on gut feelings and reasoned analysis did not moderate the relationship between how we asked participants to make their forecasts and their accuracy or confidence.

Limitations

While the findings reported here cast doubt on the claim that going with your gut improves (and reasoning impairs) forecast accuracy, this claim should be considered in light of a few limitations. While participants in Study 1A asked to go with rely on a gut feeling about who would win reported relying on their gut feeling more than participants in the reasoned analysis condition, the effect was small to moderate. Similarly, in terms of how much participants relied on reasons, in Study 1A participants in the reasoned analysis condition reported relying on their reasons more than those in the gut feeling condition. However, this did not reach significance. In Study 1B we observed that participants reported at least following instructions somewhat, but we would have expected greater reliance on gut feelings and reasoning given our instructions. One interpretation of this is that participants were unaware of how to respond to our

manipulation check questions. This could be because they do not make the same differentiation between gut feelings and reasoning that researchers do. For example, when we asked participants if they relied on their gut the intended alternative (reasoning) may not have come to mind. It is also possible that how they made their forecasts is inaccessible to them. As Halberstadt and Levine (1999) did not ask their participants about the decision mode that they employed when making their forecasts it is impossible to compare across studies. Therefore, while this should be considered when interpreting the results reported here are more likely to generalize given the much larger sample size.

In sum, manipulating decision mode did not affect the accuracy of forecasts about football games. This calls into question prior work showing the benefits of relying on a gut feeling when forecasting the outcome of Sweet 16 games. Likewise, we did not see evidence that people who tend to rely on gut feelings more were more accurate at making forecasts. We saw similar null results in terms of confidence with the exception that individual differences predicted participants' confidence in Study 1B. However, as this was not supported in subsequent studies, we are hesitant to interpret this as supporting the conclusion that people who tend to rely on their gut feelings are more confident in sports forecasts.

CHAPTER FOUR

STUDY 2A

In Study 1, we found no evidence to support the results of Halberstadt and Levine (1999), Similarly, our results did not support our additional hypotheses regarding individual differences. In Study 1, we had participants make forecasts about the team they thought would win and their margin of victory. In Study 2 (and Study 3) we gave participants experts' expectations and asked them to make a forecast about which team would exceed experts' expectations and the number of points they would exceed expectations by. Study 2 also took place in the context of the 2022 wild card games.

As with Study 1, we predicted that participants asked to rely on their gut feeling would be more accurate at forecasting which team would exceed experts' expectations and the number of points they would exceed them by (Hypothesis 1 & 2). However, we simplified our predictions about the effect of individual differences and decisional fit. As the misprediction score considers both the team a participant picks and the number of points they expect them to exceed expectations by, we only assessed individual differences and decisional fit on participant's accuracy about the amount a team would exceed expectations by (misprediction score). Therefore, we predicted that participants who tend to rely on gut feelings would be more accurate, and that this effect would be greater for those participants who tend to rely on gut feelings and were asked to do so (Hypothesis 3 & 4).

Since we gave participants experts' expectations, we could not assess their accuracy in terms of how much their forecasts deviated from experts' expectations.

However, people often substitute a more complex question (e.g., which team will exceed experts' expectations) with a simpler question (e.g., who will win the game?). This type of substitution is often reflected in people's gut feeling response (Kahneman & Frederick, 2002). Since the *favorite* team is the one that is expected to win the game, we would expect people relying on their gut feeling to pick the favorite team to exceed expectations more often than those asked to think about and provide reasons for their forecasts. In this case engaging in reasoned analysis should allow people to overcome this initial gut feeling and answer the more complex question they are asked (i.e., which team will exceed experts' expectations?)

In a related series of studies Simons and Nelson (2006) asked participants to forecast the outcome of NFL football games and found that people were more likely to pick experts' favorite team to exceed expectations compared to the underdog. Therefore, we predicted that participants asked to rely on their gut feeling would be more likely to pick the favorite team to exceed expectations (Hypothesis 5). We further predicted that people who prefer to rely on their gut feelings would pick the favorite to exceed experts' expectations more often and this would be greater for those who tend to rely on their gut feelings and were asked to do so when making their forecasts (Hypothesis 6 & 7). Our hypotheses regarding participants' confidence remain the same as in Study 1.

Method

Participants

Participants were Reddit users (N = 441) who affirmed they were over the age of 18, had not taken part in a previous study, and provided informed consent. Participants

were recruited on the same Subreddits as Study 1, saw the same study advertisement, and participated for an opportunity to win 1 of 4 $25 Amazon gift cards. We removed participants who failed to make any forecasts (22%). We further removed those who did not meet the same inclusion criteria (7%) as in Study 1. We excluded participants who did not correctly answer our two trivia questions (1.4%) and participants who did not answer the questions in under 90 seconds (< 1%). We further excluded people who received outside help (5.3%) when making their predictions. We included the remaining participants (N = 335) of which a greater number (n = 208) were asked to rely on their gut feelings than were asked to think about and provide 3 reasons for their forecasts (n = 153). Our sample predominantly identified as men (95.5%) followed by women (3.6%), non-binary (< 1%) and those who did not provide a response (< 1%). Participants were 19 to 56 years old (M = 29.49, SD = 6.89). Participants also predominantly identified as White or European American (77.3%) followed by Asian or Asian American (8.4%) Hispanic or Latinx (4.5%), Black or African American (1.2%), Native American (< 1%), the remaining participants identified with a non-listed identity or multiple ethnic identities (6%).

Measures and Procedures

The general procedure of study 2A was the same as that in Study 1A with the exception that participants made forecasts about which team would exceed bookmakers' expectations. After being randomly assigned to either make their forecasts based on a gut feeling or think about and provide 3 reasons for their forecasts, participants read a description of how we wanted them to make their forecasts. The instructions for both

conditions were the same in Study 1A with the exception that we told them they would make predictions about which team would exceed bookmakers' expectations and that this would be described in more detail on the next page. To familiarize participants with bookmakers' expectations, they read a brief description before making their forecasts:

> "In the week leading up to each game, bookmakers in Las Vegas determine which team is the favorite and which is the underdog.
>
> Bookmakers also set something known as "the spread"--the number of points that the favorite is expected to beat the underdog by.
>
> In the Rose Bowl, for instance, Ohio State was favored to beat Utah by 4 points.
>
> As it turns out, Ohio State won by a score of 48-45. They won, but only by 3 points. That is:
>
> > Ohio State won the game but fell short of bookmakers' expectations by 1 point, and
> >
> > Utah lost the game but exceeded bookmakers' expectations by 1 point.
>
> For each NFL wild card game, we're going to tell you:
>
> which team is expected to win, and the number of points that the favorite is expected to win by (that is, 'the spread')."

Participants then went through an example in which they were asked to pick which team *exceeded bookmakers' expectations* in a college football bowl game. Participants read:

> "In the Texas Bowl, Kansas State was favored by 10 points against LSU.
>
> As it turns out, Kansas State won by a score of 42-20. That is, Kansas State won by 22 points."

There were 4 choices below the example, “Kansas State exceeded expectations by 12 points”, “Kansas State exceeded expectations by 3 points”, LSU exceeding expectations by 3 points”, and LSU exceeded expectations by 12 points”. Participants who got the question correct (Kansas State exceeded expectations by 12 points) were told that they were correct and asked how well they understood what in meant for the favored team to exceed expectations and the underdog to exceed expectations. Participants who got the question incorrect were given the correct answer and a brief explanation of how to calculate it before indicating how well they understood.

Forecasts

Participants made forecasts about the 6 NFL wild card games as in Study 1. On a given trial, participants were reminded how we would like them to make their forecasts. Participants in the gut feeling condition read:

> “As we mentioned earlier, we're interested in the benefits of intuition (‘gut feelings’) in these sorts of judgments. Try to avoid any drawn-out analysis of your decisions, and make your predictions based on your first instinct.”

Participants in the reasoned analysis condition read,

> “We're especially interested in the benefits of reasoning in these sorts of predictions. Therefore, in order to prepare yourself for your predictions, we would like you to think about, analyze, and type out the reasons for those predictions.”

For each forecast, we provided participants with the teams that would be playing, the team bookmakers expected to win, and the number of points they were expected to win by. Participants also read how each team could exceed *bookmakers’ expectations*. For

example, when making forecasts about the game between the Buccaneers and Raiders they read,

> "The Buccaneers are favored to beat the Raiders by 9 points.
>
> **Buccaneers**
>
> In order to exceed expectations, the Buccaneers need to win by more than 9 points.
>
> **Raiders**
>
> The Raiders can exceed expectations either by
>
> - losing by fewer than 9 points, or
> - winning the game outright."

To make their forecasts, participants selected the team they thought would exceed expectations with the favored team always listed first followed by a slider scale. Participants in the reasoned analysis condition were first asked to think about and provide 3 reasons why they thought one team would exceed expectations before making their forecast.

For 3 of the games in the wild card series bookmakers' expectations were set at a whole point value (e.g., the Buccaneers and Raiders game above). In such cases we asked participants to imagine the favorite would not win by the number of points bookmakers expected (e.g., 9 points in the example above). Participants made their forecasts on a slider scale from 1 to 30 points. Bookmakers' expectations for the remaining 3 games ended in half-point increments (e.g., 3.5) a score that is impossible to achieve. For these

games participants made their point forecasts on a slider scale from 0.5 to 30.5. The remaining methods and procedures were the same as Study 1A.

Results

Manipulation checks

To test whether our manipulation affected the degree to which people relied on their gut feeling or reasoned analysis we compared participant's responses to the 2 manipulation check questions. As expected, participants in the gut feelings condition reported relying on their gut feelings (M = 2.63, SD = .94) when making their forecasts more than participants in the reasoned analysis condition (M = 2.26, SD = .91), $t(236)$ = 3.04, p = .0027, d = 0.39. Likewise, participants in the reasoned analysis condition reported relying more on their reasons (M = 2.83, SD = .78) than those in the gut feeling condition (M = 2.32, SD = .98), $t(236)$ = -4.31, $p < .0001$, d = 0.57. Combined these analyses indicate that participants reported at least somewhat following our instructions about how we wanted them to make their forecasts.

Differential Attrition

In study 1A, which had a between-subjects design, we observed that participants in the reasoned analysis condition made fewer forecasts than those asked to rely on their gut feeling. Therefore, we assessed whether there was evidence of differential attrition in the current study. Overall participants asked to think about and provide 3 reasons for their forecasts (M = 5.18, SD = 1.73) made fewer forecasts than those asked to rely on a gut feeling (M = 5.75, SD = 0.95), $t(333)$ = 3.85, p = .0001. Fewer participants asked to think about and provide 3 reasons for their forecasts made all 6 forecasts (79%) as compared to

those asked to rely on a gut feeling (92%). Therefore, the results reported here should be taken in context of the differential attrition we observed. However, we did not see a difference between groups in terms of how big of fans participants were in Study 1 which is one potential confounding variable that might have been introduced by attrition.

Forecast Accuracy

As in Study 1, we first computed average accuracy scores before running the analyses. In order to assess the accuracy of participant's forecasts about who would exceed expectations, we took the average number of games in which they correctly forecasted the team that would exceed expectations. Likewise, we averaged participants misprediction scores about the number of points they forecasted a team would exceed expectations by. See Appendix E for a summary of participants' responses across games. Contrary to your prediction, participants asked to rely on a gut feeling (M = .52, SD = .21) did not select the correct team to exceed expectations more frequently, $t(333) = 0.10$, p = .9187. than those asked to think about and provide 3 reasons for their forecasts (M = .51, SD = .25). Participants in the gut feeling condition (M = 12.15, SD = 2.63) were also not more accurate than those in the reasoned analysis condition (M = 12.40, SD =3.59), in forecasting the number of points a team would exceed expectation by, $t(333) = 0.72$, p = .47.

We next assessed whether participants who prefer to rely on gut feelings were more accurate in terms of their misprediction scores. We ran separate cross-classified mixed models for the 6 subscales including experimental condition, subscale, and the condition by subscale interaction as fixed effects and trial and participant ID as random

intercepts[7]. We initially included random slopes for participant ID and game, but the models had singularity issues. Therefore, we included game and participant ID as only random intercepts. The analyses reported in Table 4.1 indicated that none of the individual differences measures significantly predicted participants accuracy. Similarly, we did not see evidence that individual differences moderated the relationship between experimental condition and accuracy. Therefore, we failed to find evidence in support of our hypothesis that participants who had tend to rely on gut feelings would be more accurate in their forecasts and that this would be greater for people who made their forecasts based on a gut feeling.

Forecasts about the Favorite

We then tested whether participants in the gut feeling condition were more likely to pick the *favorite* team to exceed expectations. However, contrary to our prediction, participants in the gut feeling condition (M = .52, SD = .20) were not more likely to pick the favored team to exceed expectations as compared to those in the reasoned analysis condition (M = .50, SD = .24), $t(333) = .80$, $p = .42$. We then ran models to assess whether individual differences and *decisional* fit predicted which team (*favorite, underdog)* participants picked would exceed expectations (Table 4.1). For each subscale, we included experimental condition, preference, and the condition by preference interaction with only game as a random intercept. We tried to include random slopes for participant ID and game but experienced singularity issues. Likewise, there were singularity issues for models in which we include Participant ID as a random intercept.

[7] The model assessing participants preferences for the REI-E had singularity issues. Including participant ID as a random intercept accounted for little variance (se < .0001). We therefore removed the random intercept for participant ID.

Table 4.1 Study 2A Individual Differences and Forecasts

		Misprediction Score			*Favorite Picks*		
	Predictor	*b*	*se*	*t*	*b*	*se*	*z*
PID-I							
	(Intercept)	12.46	3.27	3.81*	0.05	0.07	0.74
	Condition	0.00	0.14	0.01	0.00	0.05	0.02
	Intuition	0.24	0.24	0.98	-0.06	0.09	-0.67
	Intuition*Condition	0.01	0.24	0.03	-0.09	0.09	-1.03
PID-D							
	(Intercept)	12.47	3.27	3.82*	0.07	0.07	1.00
	Condition	0.00	0.14	-0.03	0.01	0.05	0.29
	Reasoning	0.04	0.25	0.18	0.10	0.09	1.11
	Reasoning*Condition	0.06	0.25	0.28	0.18	0.09	1.90
REI-E							
	(Intercept)	12.34	3.31	3.73*	0.05	0.08	0.57
	Condition	0.06	0.14	0.42	0.03	0.05	0.48
	Intuition	-0.11	0.26	-0.42	-0.06	0.10	-0.59
	Intuition*Condition	0.01	0.26	0.02	-0.02	0.10	-0.18
REI-R							
	(Intercept)	12.35	3.31	3.73*	0.05	0.08	0.63
	Condition	0.03	0.14	0.22	0.03	0.05	0.58
	Reasoning	0.05	0.27	0.19	0.01	0.10	0.13
	Reasoning*Condition	-0.04	0.27	-0.15	0.13	0.10	1.28
GDMS-I							
	(Intercept)	12.35	3.30	3.74*	0.05	0.08	0.55
	Condition	0.06	0.14	0.45	0.03	0.05	0.51
	Intuition	0.05	0.22	0.21	0.03	0.08	0.36
	Intuition* Condition	0.14	0.22	0.66	-0.09	0.08	-1.14
GDMS-R							
	(Intercept)	12.36	3.30	3.74*	0.06	0.08	0.69
	Condition	0.05	0.14	0.37	0.04	0.05	0.79
	Reasoning	-0.13	0.21	-0.59	0.15	0.08	1.78
	Reasoning* Condition	0.16	0.21	0.76	0.08	0.08	0.95

Note: The results are from separate cross-classified mixed models assessing the effect of individual differences on accuracy and the number of forecasts in which a participant picked the *favorite* team.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

Participants who had a higher preference for relying on gut feelings were not more likely to pick the *favorite* (Table 4.1). Individual differences similarly did not moderate the relationship between experimental condition and which team participants picked would exceed experts' expectations (Table 4.1). Therefore, the results do not support our 3 hypotheses regarding decision mode and which team participants would forecast would exceed experts' expectations.

Forecast Confidence

To examine participant's confidence in their forecasts about which team would exceed experts' expectations we first assessed whether participants asked to rely on their gut feeling were more confident about their forecast. However, contrary to our prediction, participants in the gut feeling condition (M = 73.12, SD = 8.95) were not more confident about who they forecasted would exceed experts' expectations, $t(289) = .42$, $p = .68$, than those in the reasoned analysis condition (M = 72.68, SD = 8.57). We then assessed whether individual differences predicted participant's confidence or moderated the relationship between the experimental manipulation and their confidence.

As in the earlier analyses assessing individual differences and decisional fit, we ran separate models for each of the 6 subscales (Table 4.2). In the model we included experimental condition, participant's preference, and the condition by preference interaction as fixed effects and participant ID and game as random intercepts.[8] As predicted participants who had a higher tendency to rely on gut feelings on the PID (PID-I) were more confident in their forecasts, $t(276.03) = 2.39$, $p = .0098$, $b = 2.39$, $se = .92$.

[8] We initially included random slopes for participant ID and game, but the models experienced singularity issues. Therefore, we simplified the models to include participant ID and game as random intercepts.

Table 4.2 Study 2A Individual Differences and Forecast Confidence

		Forecast Confidence		
	Predictor	*b*	*se*	*t*
PID-I				
	(Intercept)	72.80	1.17	62.02***
	Condition	-0.26	0.52	-0.50
	Intuition	2.39	0.92	2.60**
	Intuition*Condition	-3.89	0.92	-4.23***
PID-D				
	(Intercept)	72.86	1.19	61.47***
	Condition	-0.45	0.54	-0.83
	Reasoning	-1.53	0.98	-1.55
	Reasoning*Condition	-0.64	0.98	-0.65
REI-E				
	(Intercept)	72.69	1.20	60.41***
	Condition	-0.08	0.53	-0.15
	Intuition	1.37	1.01	1.36
	Intuition*Condition	-1.46	1.01	-1.45
REI-R				
	(Intercept)	72.77	1.21	60.29***
	Condition	-0.16	0.54	-0.29
	Reasoning	-0.49	1.03	-0.47
	Reasoning*Condition	0.41	1.03	0.40
GDMS-I				
	(Intercept)	72.95	1.22	59.74***
	Condition	-0.07	0.53	-0.13
	Intuition	1.59	0.81	1.96.
	Intuition*Condition	-0.69	0.81	-0.85
GDMS-R				
	(Intercept)	73.03	1.23	59.57***
	Condition	-0.05	0.54	-0.09
	Reasoning	0.32	0.82	0.39
	Reasoning* Condition	0.67	0.82	0.82

Note: The results are from separate cross-classified mixed models assessing the effect of individual differences on confidence.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

However, this was not replicated in the other two measures assessing participants' tendency to rely on gut feelings (Table 4.2). Therefore, we are hesitant to take this as evidence supporting our hypothesis that people who prefer to rely on their gut feelings are more confident in their forecasts.

Similarly, as measured by the PID-I participant's preference for relying on gut feelings moderated the relationship between experimental condition and confidence (Table 4.2). We observed a significant interaction between scores on the PID-I and experimental condition, $t(276.03) = -4.23$, $p < .0001$, $b = -3.89$, $se = 0.92$. Participants who preferred to rely on their gut feelings were more accurate if they were asked to do so when making their forecasts as opposed to those who preferred to rely on gut feelings and were asked to think about and provide 3 reasons for their forecasts. Despite this significant result, the remaining analyses regarding decisional were not significant (Table 4.2).

Discussion

In Study 2A, we asked participants to make forecasts about who they thought would exceed experts' expectations in upcoming NFL football games. As in Study 1 we did not find evidence supporting our hypothesis that participants who made their forecasts based on a gut feeling would be more accurate. This further casts doubt on the results of Halberstadt and Levine (1999) who had previously demonstrated this effect although in a new sports forecasting paradigm. Participants asked to go with their gut feeling were not more accurate nor did they demonstrate previously observed biases (i.e., picking the winner more often). They were not more likely to pick the *favorite* to exceed

expectations as in previous research (Simmons & Nelson, 2006). Similarly, despite some significant results in terms of individual differences they were not consistent across individual difference measures and therefore we are hesitant to draw any conclusion from them.

The current study, while well powered should be considered in light of the differential attrition we observed. Participants asked to think about and provide 3 reasons dropped out of the study more frequently and were less likely to make all 6 forecasts than those asked to rely on their gut feelings. Therefore, it is possible that there were systematic differences between the two groups created by people in the reasoning condition withdrawing from the study.

CHAPTER FIVE

STUDY 2B

In study 2A, we again so no evidence supporting our hypotheses that going with a gut feeling improves the accuracy of forecasts about which team will exceed expectations, or the number of points the team will exceed expectations by. We saw similar results in terms of confidence. People who relied on their gut feeling and those who provided reasons for their forecasts were equally as confident about their forecasts. However, we observed that participants in the reasoned analysis condition were less likely to make all 6 forecasts. To address the concerns raised by differential attrition in Study 2B we manipulated whether people relied on a gut feeling or thought about and provided reasons as a within-subjects variable. We made the same predictions as in Study 2A. Study 2B took place during the same 2022 NFL wild card series. All predictions remained the same from Study 2A.

Method

Participants

Participants were recruited using the same Reddit posts as in the previous studies. Participants were Reddit users (N = 328) who affirmed they were over the age of 18 and provided informed consent. We excluded participants (27%) who did not meet the same inclusion criteria as in the previous studies. We removed participants who had taken part in a previous version of the study (1%). We removed participants who did not answer the two trivia questions correctly (3%) or did not answer them within 90 seconds (2%). Further, we exclude participants who reported receiving help from outside resources (4%). Finally, we removed participants who failed to make any forecasts (26%) All

remaining participants (n = 240) made at least one forecast and were included in the analyses. Participants predominantly identified as men (97.5%) with the remaining participants (2.5%) identifying as women. Participants also predominantly identified as White or European American (74.6%) followed by Hispanic or Latinx (7.9%), Asian or Asian American (7.1%), Black or African American (2.5%), Native American (< 1%), the remaining participants identified with a non-listed identity or multiple ethnic identities (10%). Participants were 19 to 65 years old (M = 29.12, SD = 7.20). Participants completed the study for a chance to win 1 of 4 $25 Amazon gift cards.

Measures and Procedures

As in Study 2A, participants made forecasts about which team would exceed experts' expectations in the 6 NFL wild card games. However, in Study 2B we manipulated whether people were asked to go with their gut feeling or reasons as a within-subjects variable. We randomly selected 3 games for which we asked them to go with their gut feeling when making their forecasts. For the remaining 3 we asked them to think about and provide 3 reasons why they thought one team would exceed experts' expectations when making their forecasts. We further randomized the order in which participants made their forecasts. All other methods and procedures were the same as Study 2A.

Results

Manipulation checks

We first assessed whether participants made their forecasts in line with how we asked them to. After completing their forecasts, participants reported relying on their gut

feeling slightly less than we expected. Participants reported on average relying on their gut on average just above "slightly" (M = 2.21, SD = .86) for forecasts which they were asked to go with their gut. For forecasts in which they were asked to think about their reasons they also reported "slightly" (M = 2.74, SD = .92) relying on their reasons.

Differential Attrition

To assess for differential attrition, we first excluded participants (14%) who did make at least one forecast in both conditions. We then looked at the number of forecasts participants made relying on their gut feeling and those they made using reasoned analysis. Participants made approximately the same number of forecasts for which they relied on their gut feelings (M = 2.9, SD = 0.38) and those for which we thought about and provided reasons (M = 2.86, SD = 0.49). Therefore, we concluded that we did have differential attrition in the current study.

Accuracy of forecasts

As in the prior studies we first computed participants' average accuracy scores for both the number of correct picks and their misprediction score. We then assessed whether participants asked to rely on their gut feeling were more accurate at forecasting which team would exceed experts' expectations. Contrary to our prediction, participants were not more accurate when relying on a gut feeling (M = .54, SD = .28) compared to when they were asked to think about and provide reasons for their forecasts (M = .56, SD = .32), $t(206) = 0.54$, p = .59. Similarly, participants' forecasts in the gut feeling condition (M = 10.32, SD = 3.73) were not more accurate about the number of points a team would exceed expectations by compared to those for which they were asked to think about and

provide 3 reason for (M = 10.83, SD = 4.35), $t(206) = -1.11$, $p = .27$. Together we observed that participants were not more accurate in their forecasts when they relied on a gut feeling than when they engaged in reasoned analysis. See Appendix E for a summary of participants' responses across games.

We next assessed whether participants who prefer to rely on gut feelings were more accurate in terms of their misprediction scores. We ran separate cross-classified mixed models for the 6 subscales including experimental condition, subscale, and the condition by subscale interaction as fixed effects and trial and participant ID as random intercepts. We initially included random slopes for participant ID and game, but the models had singularity issues. Therefore, we included game and participant ID as only random intercepts. The analyses reported in Table 5.1 indicated that none of the individual differences measures significantly predicted participants accuracy. However, there was a significant interaction between preferences for reasoned analysis as measured by the PID-D and experimental condition Table 5.1. Participants who had a higher preference for reasoned analysis were more accurate when they were asked to think about and provide 3 reasons for their forecasts compare to when they were asked to rely on a gut feeling $t(180) = 2.18$, $p = .03$, $b = 0.98$, $se = 0.45$. However, this was not supported by the analyses for the other subscales measuring participants preferences for reasoned analysis Table 5.1.

Forecasts about the Favorite

We then tested whether participants in the gut feeling condition were more likely to pick the *favorite* team to exceed expectations. Contrary to our prediction, participants

Table 5.1 Study 2B Individual Differences and Forecasts

		Misprediction Score			Favorite Picks		
	Predictor	*b*	*se*	*t*	*b*	*se*	*z*
PID-I							
	(Intercept)	8.48	2.60	3.26*	-0.02	0.21	0.91
	Condition	0.20	0.23	0.88	0.00	0.11	0.98
	Intuition	0.04	0.42	0.10	0.11	0.18	0.53
	Intuition*Condition	-0.59	0.39	-1.52	0.35	0.18	0.05
PID-D							
	(Intercept)	8.44	2.61	3.23*	-0.01	0.20	0.96
	Condition	0.20	0.23	0.89	0.00	0.11	0.99
	Reasoning	0.00	0.49	0.01	0.10	0.21	0.63
	Reasoning*Condition	0.98	0.45	2.18*	-0.29	0.21	0.16
REI-E							
	(Intercept)	8.58	2.62	3.28*	0.02	0.18	0.91
	Condition	0.27	0.24	1.12	-0.01	0.11	0.91
	Intuition	-0.44	0.44	-1.01	0.30	0.18	0.11
	Intuition* Condition	-0.17	0.40	-0.43	0.06	0.19	0.74
REI-R							
	(Intercept)	8.58	2.62	3.27*	0.01	0.18	0.96
	Condition	0.27	0.24	1.14	-0.01	0.11	0.92
	Reasoning	0.32	0.51	0.64	-0.21	0.21	0.32
	Reasoning*Condition	0.42	0.46	0.90	0.13	0.21	0.55
GDMS-I							
	(Intercept)	8.62	2.64	3.26*	0.01	0.18	0.97
	Condition	0.35	0.24	1.45	-0.01	0.11	0.90
	Intuition	0.06	0.36	0.16	0.25	0.15	0.11
	Intuition*Condition	-0.33	0.33	-0.99	0.20	0.15	0.20
GDMS-R							
	(Intercept)	8.56	2.64	3.24*	0.02	0.16	0.89
	Condition	0.34	0.24	1.45	-0.01	0.11	0.90
	Reasoning	-0.23	0.47	-0.49	0.17	0.20	0.40
	Reasoning*Condition	0.81	0.43	1.89	-0.22	0.20	0.27

Note: The analyses reported here were separate cross-classified mixed models.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

were not more likely to pick the favorite team to exceed expectations when relying on their gut (M = .53, SD = .29), $t(206) = -0.16$ p = .88, than when thinking about reasons (M = .54, SD = .31). We then ran cross-classified mixed models to assess whether individual differences and *decisional* fit predicted which team (*favorite, underdog)* participants picked would exceed expectations (Table 5.1). For each subscale, we included experimental condition, preference, and the condition by preference interaction with only game as a random intercept. We tried to include random slopes for participant ID and condition but experienced singularity issues. Likewise, there were singularity issues for models in which we include Participant ID as a random intercept. The random intercept for Participant ID accounted for very little variance and was subsequently dropped from the models. Participants who had a higher preference for relying on gut feelings were not more likely to pick the *favorite* for any of the individual difference measures (Table 5.1). Likewise, we so no effect of *decisional fit* on the likelihood participants would pick the *favorite* to exceed expectations (Table 5.1).

Forecast Confidence

Participants made their confidence judgments after completing all 6 forecasts. Therefore, some participants (4%) did not make confidence judgments for forecasts made using both decision modes. Before running a *t*-test to compare participants confidence we removed these participants. Contrary to our prediction, participants were not more confident about forecasts made using their gut feeling (M = 73.70, SD = 9.85) than those made in the reasoned analysis condition (M = 74.98, SD = 10.14), $t(188) = -1.58$, p = .12.

We next turned to the individual difference measures. We ran the same models as discussed above with participant's confidence as the outcome measure. Participants who prefer to rely on their gut feelings were not more confident in their forecasts (Table 5.2). Similarly, we did not see evidence of an interaction between preferences and experimental condition (Table 5.2). Therefore, in the current study we individual differences nor *decisional fit* affected participant's confidence in their forecasts. Note that while condition was a significant predictor of confidence when controlling for individual differences on the PID, it was not significant in the *t*-tests reported earlier nor replicated in models assessing the other individual difference measures.

Discussion

In study 2A we observed differential attrition. People asked to rely on their gut feeling were more likely to make all 6 forecasts about the 2022 NFL wild card series than those asked think about and provide 3 reasons for their forecasts. Based on this concern, in Study 2B we manipulated how people made their forecasts as a within-subjects variable. In the current study we saw that participants were approximately equally as likely to make forecasts in both conditions. However, despite effectively eliminating differential attrition we, as in the previous studies, so no evidence for our hypothesis that relying on gut feelings improves forecast accuracy. Similarly, we did not see evidence that relying on gut feelings increased the confidence participants had in their forecasts. Likewise, individual differences and *decisional fit* did not reliably affect participants' accuracy or confidence.

Asking people to rely on their gut feeling when making forecasts about which team will

Table 5.2 Study 2B Individual Differences and Forecast Confidence

		Forecast Confidence		
	Predictor	*b*	*se*	*t*
PID-I				
	(Intercept)	73.60	1.90	38.74***
	Condition	1.44	0.63	2.29*
	Preference for Intuition	0.53	1.28	0.42
	Preference for Intuition by Condition	1.06	1.06	0.99
PID-D				
	(Intercept)	73.59	1.88	39.20***
	Condition	1.44	0.63	2.28*
	Preference for Reasoning	0.00	1.50	0.00
	Preference for Reasoning by Condition	0.47	1.25	0.38
REI-E				
	(Intercept)	73.75	2.17	33.99***
	Condition	1.06	0.64	1.66
	Preference for Intuition	1.02	1.29	0.79
	Preference for Intuition by Condition	-0.05	1.08	-0.05
REI-R				
	(Intercept)	73.82	2.10	35.15***
	Condition	1.06	0.63	1.67
	Preference for Reasoning	0.53	1.53	0.35
	Preference for Reasoning by Condition	0.28	1.25	0.22
GDMS-I				
	(Intercept)	73.79	2.26	32.65***
	Condition	1.09	0.65	1.67
	Preference for Intuition	0.57	1.07	0.54
	Preference for Intuition by Condition	0.37	0.90	0.41
GDMS-R				
	(Intercept)	73.78	2.24	33.01***
	Condition	1.09	0.65	1.67
	Preference for Reasoning	0.97	1.40	0.70
	Preference for Reasoning by Condition	0.05	1.18	0.05

Note: The analyses reported here are from separate cross-classified mixed models.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

exceed experts' expectations did not increase the accuracy of their forecasts. Study 2 provided a conceptual replication of Halberstadt and Levine (1999). However, unlike the previous authors, we observed no effect of relying on gut feelings. As noted in the introduction to Study 2A we had anticipated that asking participants to make forecasts about which team would exceed experts' expectations was a task in which people were likely to have affective reactions. It is similar to how Halberstadt and Levine (1999) asked their participants to make their forecasts, but it more closely mimics sports betting in which bets are placed on who will exceed expectations. However, we observed similar null results as in Study 1. Beyond not replicating the results of Halberstadt and Levine, (1999) we also saw no evidence supporting the findings by Simons and Nelson (2006) which had previously demonstrated that people more often pick the *favorite* to exceed expectations. We expected that participants relying on gut feelings would be more likely to substitute the question we had asked them ("which team will exceed experts' expectations?") with the simpler question, "which team will win the game?" However, both groups of participants picked the *favorite* team equally and picked the *favorite* and *underdog* in approximately equal proportion. Participants in Study 2 were also not more confident in their forecast when they made them based on their gut feeling. Likewise, across study 2, we did not support our additional hypothesis regarding individual differences and *decisional fit*.

While the results of Study 2 cast further doubt that relying on gut feelings improves the accuracy of sports forecasts and how confident people are in their forecasts, there is one important limitation to note. Participants in both Studies 1 and 2, on average,

only reported somewhat following our instructions about how we wanted them to make their forecasts (Table 5.3). Therefore, it is possible that participants were not relying on a gut feeling when making forecasts in which we asked them to. In the reasoned analysis condition, asking participants to write in 3 reasons for their forecasts likely engaged them in reasoned analysis. However, it is also possible that they made a pick before reasoning about which team would win. In such a case participants may have engaged in post-choice rationalization of their pick based on a gut feeling, and recorded reasons that were in line with the forecasts that already had in mind. With this limitation in mind, it is still possible that relying on gut feelings does affect sports forecasts, we just did not manipulate decision mode. We attempt to address this concern in Study 3.

Table 5.3 Effect of Experimental Manipulation on Reported Reliance on Decision Mode

	Reliance on Gut feeling		Reliance on Reasons	
Condition	Gut Feeling	Reasons	Gut Feeling	Reasons
Study 1A	*M* = 2.42 (1.05)	*M* = 2.06 (0.96)	M = 2.64 (0.94)	*M* = 2.82 (0.86)
Study 1B	*M* = 2.59 (0.95)			*M* = 2.91 (0.79)
Study 2A	*M* = 2.63 (0.94)	*M* = 2.32 (0.98)	*M* = 2.26 (0.91)	M = 2.83 (0.78)
Study 2B	*M* = 2.21 (0.86)			*M* = 2.74 (0.92)

Note: Participants reported the degree to which they relied on their gut feelings and reasoned analysis on scale of 0(not at all) to 4(extremely). Table 5.3 reports the means and standard deviations in parentheses for participants in each study separated by the condition they were assigned to.

CHAPTER SIX
STUDY 3

In Studies 1A through 2B we observed no effect of our experimental manipulation on the accuracy of (or confidence in) forecasts about the outcome of NFL football games. However, as noted in Table 5.3 we observed a modest effect of our manipulation on our whether participants reported relying on their gut feeling or reasons. While Halberstadt and Levine (1999) did not ask participants to report the degree to which participants relied on their gut feeling and reasons when making their forecasts it is possible that they were better able to manipulate how participants made their forecasts and therefore observed an effect of manipulated decision mode on accuracy. Halberstadt and Levine (1999) had their participants come for an in-person laboratory study in which they made their forecasts while we asked people to complete a survey online. Asking people to make their forecasts in-person may increase their adherence to the instructions about how to make their forecasts. We were unable to conduct Study 3 in person, and therefore we used the same Subreddits as in Study 1 and 2. However, we tried to strengthen our manipulation to encourage participants to make their forecasts based on either a gut feeling or reasoned analysis. Along with telling participants an abridged version of the instructions used in studies 2A and 2B, we had participants in the gut feeling condition make their forecasts under *high time pressure* and those asked to first think about the reasons why they thought one team would win over the other under *low time pressure*. Study 3 took place during the 4 game 2022 NFL Divisional Playoffs (Appendix F). Predictions for Study 3 were the same as in Study 2.

Method

Participants

Interested redditors read a similar advertisement as earlier studies on the same NFL subreddits. Participants (N = 862) self-selected to take part in the study for a chance to receive 1 of 4 $25 Amazon gift cards. Participants were approximately equally assigned to the gut feeling condition (n = 420) and reasoned analysis condition (n = 442).

We excluded participants who did not meet the same exclusion criteria used in the previous studies (23%). We removed participants who had taken part in a previous study (5%). We excluded those that did not answer the two trivia questions correctly (4%) and those that failed to answer the questions in under 90 seconds (< 1%). Further participants who received help from outside sources were excluded from the analyses (6%). Participants who failed to make at least one forecast were also removed (11%). The remaining participants (N = 656) were split between the *high time pressure* condition (n = 327) and *low time pressure* condition (n =329).

Participants were of a similar demographic makeup as the other Reddit based samples in our previous experiments. Participants overwhelming identified as men (93.3%) with the remaining participants identifying as woman (5%), non-binary or other (2%). The majority of participants also identified as White or European American (76%), followed by other or multiple ethnic groups (9%), Hispanic or Latinx (7%), Asian or Asian American (6%), Black or African American (1%) and Native American (< 1%). Participants included in the analyses were equally assigned to the *high time pressure* and *low time pressure* conditions. Participants were 19 to 77 years old (M = 29.9, SD = 8.23).

Survey

The general survey design was the same as study 2A with a few important exceptions. As in Study 2A participants first completed the two trivia questions, demographics questionnaire and were told to make every effort to complete the study if they continued. Participants were then randomly assigned to one of two conditions. Participants were either asked to make their forecasts based on a gut a feeling and under high time pressure, or they were asked to think about reasons why they thought one team or another would exceed experts' expectations under low time pressure. Both sets of participants made forecasts about which teams would exceed expectations in the 2022 NFL Divisional Playoff series. Participants in the *high time pressure* condition were required to make their forecasts in under 10 seconds and asked to rely on their gut feeling. Therefore, before making their forecasts, they read:

> "We're interested in how well NFL football fans can predict which team will exceed bookmakers' expectations in the upcoming divisional playoff games. (We'll talk more about bookmakers' expectations on the next page.) We'd like you to go with your intuition or "gut feeling" when making your predictions. In other words, we would like you to make your predictions based on your first instinct. To help you go with your gut, you will only have 10 seconds to make each prediction."

Participants in the *low time pressure* condition were required to wait 15 seconds after reading about a given matchup. Participants in the *low time pressure* condition were told:

> "We're interested in how well NFL football fans can predict which team will exceed bookmakers' expectations in the upcoming divisional playoff games. (We'll talk more about bookmakers' expectations on the next page.) We'd like you to think about and analyze the reasons why you think one team will win over the other. Therefore, we will ask you to think about each game for at least 15 seconds before entering your prediction."

As in Study 2, on the following page, participants read a description of experts' expectations and were asked a trivia question about which of two teams in a previous game exceeded expectations to help them get familiar with experts' expectations.

Before each forecast, participants were reminded to either rely on their "gut feeling" (*high time pressure* condition) or to think carefully about why one team or the other would exceed expectations (*low time pressure* condition). In *the high time pressure* condition participants saw a countdown timer at the top left of the screen that indicated how much time they had left to make their forecast. Below the timer they read information about the game they would be making a forecast about. For example, when making a forecast about the game between the Tennessee Titans and Cincinnati Bengals participants read that the "Titans are favored to beat the Bengals by 3.5 points." Below they were asked to pick which team they felt would exceed expectations with the favored team listed first, and the number of points they felt they would exceed expectations by on a slider scale from 0.5 to 30.5 points. Because experts' expectations for all 4 games in the divisional playoffs ended in a half point (e.g., 3.5 points in the game above), participants made all 4 forecasts on the same scale. If a participant did not make a forecast within 10

seconds, the survey automatically advanced. Participants who missed a forecast were reminded that they needed to make their forecasts within 10 seconds and asked if they would like to make a forecast for the previous game. If they opted to return to the previous forecast, they were not under time constraints to make their forecast. Therefore, we removed forecasts about games in which the participant failed to make the forecast in time before running the analyses.

Participants in the *low time pressure* condition made forecasts in a similar way, however they saw a timer counting down from 15 seconds. Below the timer was the same information about the teams that would be playing and a reminder that "for at least the next 15 seconds, please think carefully about which team will exceed expectations." After the timer expired, Qualtrics displayed the two questions with which participants could make their forecasts.

To reduce the number of missing forecasts due to user error in *the high time pressure* condition, participants completed one practice trial in which they made a forecast about two teams who were not playing in the divisional playoffs. To keep the survey consistent across conditions, participants in both conditions completed a practice trial that matched how they would make their forecasts. For each forecast a participant made, they reported how confident they were that their chosen team would exceed expectations on a scale from 1 (I am not at all confident) to 7 (I am extremely confident. Participants then either completed the REI or PID before answering the same post-experiment questions from the previous studies.

Results

Differential Attrition

Based on the results of the previous 4 experiments in which we manipulated decision mode as a between-subjects variable, we examined whether there was differential attrition between the 2 conditions. The *low time pressure* condition took longer to complete than the *high time pressure* condition. However, in the *high time pressure* condition we removed forecasts in which participants did not make their forecast within the 10 second time limit. Therefore, of the forecasts included in the analyses, participants in the low time pressure condition were more likely to have made all 4 forecasts (95%) compared to those in the high time pressure condition (80%). However, of participants in the high time pressure condition who did not make all 4 forecasts in time, the majority made 3 forecasts (79%). Therefore, it is possible that the forecasts that were included from participants in the high time pressure condition represent only those that they had a strong gut feeling about. However, since this is in line with the experimental manipulation, we think it is unlikely that this introduces a confound. While we included fewer forecasts made in the *high time pressure* condition, there were an equal number of participants in each condition. Therefore, at the participant level we did not see evidence of differential attrition.

Manipulation checks

The primary aim of study 3 was to strengthen the effect of our manipulation. Therefore, we first assessed whether having participants make forecasts under high and low time pressure increased their reliance on the decision mode we asked them to use.

Participants in the *high time pressure* condition (M = 2.41, SD = 0.98) reported relying on their gut feelings more than participants in the *low time pressure* condition (M = 2.13, SD = .94), $t(602) = 3.61$, $p = .0003$, $d = .29$. Likewise, participants in the *high time pressur*e condition (M = 2.52, SD = .97) reported relying on their reasons less, $t(600) = -5.10$, $p < .0001$, $d = -.42$ than the those in the *low time pressure* condition (M = 2.89, SD = .82). Therefore, participants at least reported somewhat following our instructions making their forecasts based on either their gut feeling or reasoned analysis.

Forecast Accuracy

We first looked at how accurate participants were at forecasting which team would exceed expectations. Contrary to our expectation, participants in the *high time pressure condition* (M = .48, SD = .24) did not pick the correct team to exceed expectations more often than those in the *low time pressure* condition (M = .50, SD = .26), $t(654) = -.08$, $p = .40$. Likewise, participants in the *high time pressure* condition (M = 7.51, SD = 2.56) were not more accurate at forecasting the number of points a team would exceed expectations by than those in the *low time pressure* condition (M = 7.52, SD = 2.76), $t(654) = -0.01$, $p = .99$. See Appendix F for a summary of participants' responses across games.

We then assessed whether individual differences and *decisional fit* affected participants accuracy (Table 6.1). We ran cross-classified mixed models for each of the 4 subscales measuring participants preferences for relying on gut feelings and reasoned analysis. We included preference, experimental condition, and the preference by condition interaction as fixed effects and a random intercept for the game participants

Table 6.1 Study 3 Individual Differences and Decisional Fit

		Misprediction Score			*Favorite Picks*			Forecast Confidence		
	Predictor	*b*	*se*	*t*	*b*	*se*	z	*b*	*se*	*t*
PID-I										
	(Intercept)	7.46	0.94	7.91**	-0.30	0.19	-1.63	4.44	0.12	37.55***
	Condition	0.20	0.24	0.82	0.00	0.09	0.05	0.11	0.08	1.40
	Intuition	0.25	0.32	0.79	0.18	0.13	1.44	0.15	0.10	1.46
	Intuition*Condition	0.04	0.45	0.08	-0.07	0.18	-0.42	0.09	0.15	0.58
PID-D										
	(Intercept)	7.48	0.94	7.92**	-0.30	0.19	-1.61	4.45	0.12	37.58***
	Condition	0.17	0.24	0.73	0.00	0.09	-0.03	0.10	0.08	1.28
	Reasoning	0.90	0.32	2.82**	-0.02	0.13	-0.14	0.15	0.10	1.45
	Reasoning*Condition	-0.70	0.46	-1.52	0.19	0.18	1.07	0.05	0.15	0.32
REI-E										
	(Intercept)	7.66	0.99	7.70**	-0.30	0.19	-1.61	4.59	0.12	39.01***
	Condition	-0.17	0.48	-0.35	-0.14	0.19	-0.74	-0.36	0.15	-2.34*
	Intuition	-0.41	0.57	-0.72	-0.37	0.23	-1.61	0.24	0.18	1.35
	Intuition*Condition	1.23	0.83	1.47	0.16	0.34	0.48	0.11	0.27	0.42
REI-R										
	(Intercept)	7.68	0.96	8.04**	-0.31	0.18	-1.72	4.55	0.12	37.91***
	Condition	-0.18	0.47	-0.37	-0.11	0.19	-0.58	-0.32	0.16	-2.02*
	Reasoning	0.53	0.62	0.86	0.27	0.25	1.12	0.03	0.20	0.14
	Reasoning*Condition	-0.59	1.11	-0.53	-0.23	0.44	-0.51	-0.17	0.37	-0.46

Note: The results reported here are from separate cross-classified mixed models.

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

made forecasts about. Contrary to our prediction preference for relying in reason measured with the PID-D significantly predicted accuracy, $t(1877.01) = 2.84$, $p = .0048$, $b = 0.90$, $se = 0.32$, with higher preference related to greater accuracy. However, this was not replicated with scores on the REI-R. None of the interaction terms between condition and preferences were significant indicating that *decisional fit* did not increase accuracy (Table 6.1).

Forecasts about the Favorite

Contrary to our expectation participants in the high time pressure condition ($M = .43$, $SD = .25$) were not more likely to select the *favorite* to exceed expectations than those in the low time pressure condition ($M = .42$, $SD = .24$), $t(654) = 0.54$, $p = .59$. Likewise, individual differences nor *decision fit* significantly predicted whether someone picked the *favorite* to win the games (Table 6.1).

Confidence

Contrary to our expectation participants in the high time pressure condition were not more confident in their forecasts ($M = 4.49$, $SD = 0.92$) than those in the low time pressure condition ($M = 4.47$, $SD = 0.89$), $t(646) = 0.24$, $p = .81$. Similarly individual differences nor *decision fit* significantly predicted whether someone participants' confidence in their forecasts. (Table 6.1).

Discussion

In study 3, we attempted to bolster our manipulation to encourage participants to rely on their gut feelings and reasoned analysis. However, participants' responses to our manipulation check questions calls into question whether adding time pressure affected

which decision mode they used when making their forecasts. More specifically, we required that participants asked to rely on their gut feeling make their forecasts within 10-seconds of reading the two teams that would be playing. Gut feelings come to mind quickly and allow people to make responses faster than engaging in reasoned analysis (Kahneman & Frederick, 2002). Therefore, we reasoned that time pressure would increase reliance on gut feelings because participants would not have time to engage in reasoned analysis. Prior research has linked latency speed to intuitive processing (Evans & Curtis-Holmes, 2005). Evans and Curtis-Holmes (2005) demonstrated that responses made in under 10 seconds reflect heuristic, gut feeling answers when solving syllogisms. In the low time pressure condition, we told participants to first think about reasons why one team or another would exceed experts' expectations for 15 seconds. We chose 15-seconds to encourage reasoned analysis without implementing a delay that might frustrate participants. As mentioned, we expected this to increase the amount that participants reported relying on their assigned decision mode. However, we saw similar results on our manipulation check questions as in other studies. One explanation for this, expressed previously, is that the manipulation checks were poorly calibrated. It is possible that participants were either unable to report how they made their forecasts (based on gut feelings or reasoned analysis) or that the choices (e.g., I relied on my gut feelings an extreme amount) discouraged participants from selecting them. However, despite these limitations we replicated the results of our previous 4 experiments which together provide evidence contradicting Halberstadt and Levine's (1999) findings.

In terms of individual differences, we again found that preferences for relying on gut feelings and reasoned analysis did not predict the accuracy of participants' forecasts. Likewise, individual differences did not predict how confident people were about their forecasts. While these results are surprising given previous research, they support and replicate the results reported here in earlier experiments.

CHAPTER SEVEN
GENERAL DISCUSSION

The series of studies we reported here were in service of two aims. First, we sought to replicate previous research which demonstrated the benefits of relying on gut feelings when making forecasts about sporting events. Second, we attempted to extend previous work to examine individual differences and forecast confidence. However, across 5 studies, we were unable to replicate the results of Halberstadt and Levine (1999) using a robust online, Reddit sample. Asking people to rely on their gut feeling or reasoned analysis when making forecasts about NFL games did not affect the accuracy of their forecasts. Our results provide little evidence that decision mode affects the accuracy of sports forecasts or how confident people are about their forecasts. As discussed previously, sports forecasting represents a special case of decision making in which there is an objective and verifiable criterion for which to compare forecasts (i.e., the outcome of the game). The results reported here indicate that these types of forecasts may be beyond the types of decisions in which gut feelings improve accuracy.

Decision Mode and Forecast Accuracy

Our results regarding the effect of manipulated decision mode and individual differences on accuracy are surprising considering previous work. However, there are several possible reasons for why our results did not replicate those of Halberstadt and Levine (1999). We will first review some of the limitations of the current work which may explain the difference between our results and those of Halberstadt and Levine. We will next discuss the strengths of the current work and why the current studies represent a

more robust test of the hypothesis that relying on gut feelings improves the accuracy of forecasts about sporting events.

Population and Sample

It is first important to note that the samples in the present work were not representative. In all 5 studies our participants predominately identified as white and as men. Therefore, the conclusions drawn from this research may not generalize to other identities. Women and men differ in which decision mode they tend to use. Women tend to rely more on gut feelings when making decisions. For example, women tend to report being more experiential than men on the REI (Epstein, 2003; Sladek et al., 2010). Likewise, women tend to have a higher faith in their intuition than men (Epstein et al., 1996; Aarnio & Lindeman, 2005; Ward & King, 2020). Men also tend to have a higher tendency to use reasoned analysis than women (Sladek et al., 2010). Based on this it is possible that our sample of mostly men had lower trust in their intuition and found it difficult to rely on their gut feelings when making their forecasts. This may have (a) led to lower overall accuracy in the gut feelings condition or (b) led participants in the gut feelings condition to rely more on reasoned analysis. Both of these could, in part, explain our null results.

However, it is unlikely that this explains why our results differed from Halberstadt and Levine (1999). They do not report the demographics of their participants, so we are unable to compare the representativeness of our samples with theirs (e.g., the gender makeup of the two samples). However, it is likely that their sample also predominantly identified as white and as men. Therefore, future research should attempt

to collect a more diverse and representative sample to further examine how decision mode affects the accuracy of sports forecasts.

There are some reasons to believe that our sample did differ from Halberstadt and Levine (1999). Our participants completed the study online as opposed to in-person as done in Halberstadt and Levine's study. Participants recruited online have been found to give similar results as those recruited for in-person studies (Birnbaum, 2004). Likewise, recruiting participants from social media is becoming more common (Amaya et al., 2021) including from Reddit. However, it is possible that our participants differed from those in Halberstadt and Levine's study. In terms of demographics, our sample had a more diverse range of ages and consisted of an overall older group of participants. Previous research has shown that age and reliance on intuition are negatively correlated (Sladek et al., 2010). This provides further evidence that our participants may have had a more difficult time relying on their gut feelings when asked to make their forecasts based on gut feeling than those in previous research.

Manipulation of Decision Mode

We may not have manipulated decision mode as effectively as Halberstadt and Levine (1999) despite using a similar instructions manipulation. Our participants may have paid less attention to our instructions or did not follow them as rigorously. As noted, participants completed our study online. This may have reduced how likely they were to follow our instructions. While it is unknown how many participants asked to provide 3 reasons in Halberstadt and Levine (1999) did so before making their forecasts, we observed that many of our participants did not. Therefore, it seems plausible that our

participants did not follow our instructions closely, which may have reduced the effectiveness of our manipulation. Likewise, some participants made negative comments about the number of questions on the individual difference measures which reduces the likelihood that they completed them accurately or may have completed them using a response set (e.g., acquiescence).

There are additional reasons to suspect that we did not manipulate decision mode. In the current studies, we offered a small potential incentive for completing the current studies (chance at winning a $25 gift card). Therefore, this increases the likelihood that participants paid less attention to our instruction or were less motivated to follow them. It is also logical that this would be higher for our participants who were asked to think about and provide reasons for their forecasts. Coming up with reasons for each forecast is a more time consuming and difficult process than making forecasts based on gut feelings. This may have led to some of the differential attrition we observed. It is also possible participants in both conditions may not have followed our instructions on how to make their forecasts, in part, due to the low incentive for completing the study.

Another reason we may not have manipulated decision mode effectively is that we told participants that we would post their results anonymously online after the end of the study. We did this to increase recruitment and to encourage participants to make all the forecasts we asked them to. However, this may have encouraged participants not to follow our instructions. Participants may have used an alternative decision mode in order to beat other fans' forecasts. They may have used a decision mode that they thought would allow them to make the most accurate forecasts or consulted outside resources

when making their forecasts. We removed participants who told us that they had consulted outside resources when making their forecasts which reduced the possibility that this affected our results. However, some participants may not have responded honestly when we asked them whether they had consulted outside resources. Likewise, participants who withdrew from the study early did not answer this question.

As discussed, it is possible that we did not manipulate the decision mode participants used in the studies reported here. If participants did not follow our instructions, it likely would have increased their reliance on the decision mode they tend to use. If this was the case, and gut feelings do improve the accuracy of sports forecasts, we would have observed significant effects of our individual difference measures on accuracy. This assumption is supported by previous research (e.g., Raab and Laborder, 2011; Betsch & Kunz, 2008; Bastick, 1982) in which individual differences measures are correlated with decisions. However, we did not observe a consistent relationship between participants' tendency to rely on each decision mode and accuracy. Therefore, despite the likelihood that we failed to manipulate participants' decision modes, our results still cast doubt on previous research which has observed that decision mode affects the accuracy of sports forecasts.

Responses to our manipulation check questions also raise the concern that we did not manipulate how people made their forecasts. Participants in all 5 studies did report at least somewhat following our instructions (e.g., using reasoned analysis to make their forecasts when asked to do so). However, the effect size of the manipulation varied across studies and was not as strong as we had expected. While it is possible that

participants were unsure how they made their forecasts when asked later, it is also likely that they either did not follow our instructions or were unable to conform their decision making process to our instructions.

In studies 1 through 4, we used a similar instructions manipulation as in Halberstadt and Levine (1999). However, participants reported only somewhat following how we asked them to make their forecasts. Likewise, in Study 3 adding a time pressure manipulation had little effect on how participants reported making their forecasts. It is worth noting that Halberstadt and Levine did not ask participants how they made their forecasts. Consequently, we are unable to compare the degree to which participants relied on gut feelings and reasoned analysis in our study and theirs. We also failed to appropriately assess how closely participants followed our instructions in Studies 2 and 4. For these studies, we only asked participants how much they used the decision mode we asked them to for a given trial and not how much they used the other. Specifically, we asked them how much they relied on their intuition or gut feelings when making forecasts on which we asked them to and how much they used reasoning on forecasts where we asked them to. Therefore, we were only able to compare their responses to these questions against the midpoint of the scale.

Differential Attrition

We observed significant differential attrition in our studies. Participants asked to engage in reasoned analysis about each game made fewer forecasts than those asked to rely on their gut feeling. While we did not see that participants in the two conditions varied in fanship, there are other potential confounding variables that we cannot account

for. For example, participants in the reasoned analysis condition who made all 6 forecasts may have been more motivated to complete the study and took their forecasts more seriously than those in the gut feelings condition. This may have affected our results by increasing the accuracy of participants in the reasoned analysis condition. While we are unable to assess this in the current research, prior research has demonstrated that differential attrition in online studies can affect the strength and direction of results (Zhou & Fishbach, 2016). Therefore, as mentioned previously the results should be considered in context of the differential attrition. Future research on the effect of decision mode should be careful to try to reduce participants withdrawal in the reasoned analysis condition. One likely first step is to use a within-subjects design as done here which reduced differential attrition substantially.

Individual Differences

Despite the limitations discussed, the current studies represent a more robust test of the hypothesis that relying on gut feelings improves sports forecasts. There are two remaining explanations for why our studies do not support the results of Halberstadt and Levine (1999). First, it is possible that Halberstadt and Levine observed significant results due to Type-I error. Their small sample size, and the fact that they only conducted one study encourages such an interpretation. The more roust samples utilized across the 5 studies in the current work further support this assertion. However, the results of Plessner et al., 2006 as reported in Plessner and Czenna (2007), support Halberstadt and Levine's (1999) results. Plessner and colleagues' results were, however, involved in higher level

interactions including expertise, available information, and time until the game would be played.

The second interpretation is that other changes between the two experimental designs impacted the results. Specifically, it is possible, although unlikely, that NFL sports forecasting is somehow different than forecasts about college basketball and FIFA soccer. For one, there are far fewer teams in the NFL than those that play in NCAA division I college basketball. Therefore, it is likely that all of our participants had some familiarity with the teams that would be playing. In contrast, participants who made forecasts about the Sweet 16 may have been largely unfamiliar with some of the teams that would be playing. In such a case, participants asked to rely on their gut feelings likely picked the team they were most familiar with, a stand-in for how well a team usually plays. Relying on recognition as information can be an effective decision-making strategy (i.e., the *recognition heuristic*; Goldstein & Gigerenzer, 2002; Pachur et al., 2011). In contrast, those in the reasoned analysis condition who were asked to provide their reasons for why a team would win may have changed their attitudes about the two teams in line with inconsequential or irrelevant information that came to mind.

Another difference between our study and Halberstadt and Levine (1999) is that they collected data in 1995 and 1996. It is possible that our participants had more information about the teams that would be playing due to information being more widely available via the internet. While this is likely, it is doubtful that this would have affected our results. Plessner et al., 2006 as reported in Plessner and Czenna (2007) found that people asked to rely on a gut feeling were more accurate when they had greater expertise

and when they had more information about the teams that would be playing. Therefore, it is unlikely that this is the reason why our results do not support those of Halberstadt and Levine (1999).

In the current work we see little evidence that manipulating decision mode affects the accuracy of sports forecasts. While future research should continue to examine decision mode and sports forecasts, our results support the conclusion that sports forecasting may fall outside of the domain of decisions that are affected by decision mode.

As mentioned briefly above, despite the limitations regarding the efficacy of our manipulation, measuring individual differences in reliance on decision modes provides a strong test of the hypothesis that gut feelings improve sports forecasting accuracy. Therefore, the conclusion that relying on gut feelings does not improve the accuracy of sports forecasts is further supported by our results that participants' tendency to rely on gut feelings and reasoned analysis were unrelated to the accuracy of their forecasts. Given that our manipulation had a modest effect on how participants reported that they made their forecasts, we get an especially clear look at individual differences and forecast accuracy.

An intuitive decision mode has previously been demonstrated to improve decision making in complex environments like playing handball (Raab and Laborder, 2011). Similarly, many studies have found a positive relationship between reliance on an intuitive decision mode and business outcomes (e.g., Sadler-Smith, 2004). However, we see no evidence that relying on gut feelings improves decision making in sports

forecasting. Therefore, it is likely that decision mode does not affect the accuracy of sports forecasts specifically. However, decision mode may still affect other decisions including those made by athletes.

In the current studies, we also saw no evidence that decisional fit increases the accuracy of sports forecasts. Given that we did not observe main effects of individual differences or manipulated decision mode it is not surprising that we did not observe the predicted interaction between decision mode and individual differences. However, this is in contrast to previous research which has observed that decisional fit increases, for example, the positive attributes attributed to a chosen object (Betsch & Kunz, 2008). Likewise, Betsch and Kunz observed that decisional fit decreases post decision regret. However, our results further support the assumption that sports forecasts are outside of the types of decisions that decision mode influences.

Forecast Confidence

Similar to our results regarding forecast accuracy, people asked to rely on their gut feelings were not more confident about their forecasts than those asked to think about and provide reasons for their forecasts. We had predicted that basing forecasts on gut feelings which come to mind more easily would increase participants' confidence in their forecasts based on Simons and Nelson (2006). They demonstrated that people frequently choose intuitive as opposed to non-intuitive choices. This is ostensibly because people hold their intuitions with high confidence. In line with this hypothesis, people also tend to be more confident in decisions made quickly (Zakay & Tuvia, 1998; Ackerman & Zalmanov, 2012). We did observe that preferences for relying on gut feelings in Study 1B

was positively related to forecast confidence. However, since this was not replicated in the remaining experiments, we are hesitant to interpret it as support for our hypothesis. One reason we did not observe a relationship between decision mode and confidence, is that our participants may not have had strong gut feelings about which team would win a game. If a gut feeling did not come to mind quickly and easily then it is unlikely that relying on a gut feeling would increase accuracy compared to using reasoned analysis. We conclude that there is little evidence that relying on gut feelings improves peoples' confidence in sports forecasts. This further supports our conclusion that sports forecasting may be a type of decision that is unaffected by decision mode.

Conclusions

Previous research had suggested that sports forecasting is one area of decision making in which relying on gut feelings would improve decision making. People are likely to have affective reactions toward which team they think will win the game and who will win a game is influenced by innumerable variables (i.e., a complex decision). Given prior research that has shown that in these types of decisions, gut feelings improve decision making (e.g., McMackin and Slovic, 2000) we had predicted that people who rely on gut feelings would make better forecasts. Based on our results, however, it is likely that sports forecasts are a boundary condition. Sports forecasting may fall outside the domain for which decision mode influences the outcome of decision making.

Why sports forecasting may be unaffected by decision mode is unclear from the present research. One potential reason is that it is difficult to forecast who will win a game with substantial accuracy. Sports betting depends on the inability of people to

create a reliable strategy for picking who will win a game. This is likely to be especially difficult in playoff games as the two teams that are playing are likely to be at least relatively evenly matched. As a reference, prediction markets, bookmakers, tipsters (professional betting services), and statistical models all perform similarly poorly at forecasting the outcome of sporting events (Spann & Skiera, 2009). Therefore, future studies should examine the effects of decision mode on decision making using other objective and verifiable domains (e.g., stock markets and political elections).

However, other decision making domains fitting these criteria are likely to be equally difficult to forecast with any substantial degree of accuracy. Despite this, people often make forecasts about the future, and it is important to understand the effect of decision mode on the quality of these types of decisions. This is especially true since many of the forecasts people routinely make have potentially life altering consequences (e.g., who will win an election or which stock to invest in). It should also be noted that other tasks in the judgment and decision-making literature may be similarly difficult. For example, estimating the amount of mail handled by the U.S. Postal Service and other unknown quantities estimated in MacGregor et al. (1988) and McMakin and Slovic (2000). Therefore, beyond the difficulty of the task, there may be other variables that influence when a decision mode is likely to improve decision making.

Across 5 experiments we observed that people who rely on gut feelings and reasoned analysis are equally as accurate at forecasting the outcome of sporting events. This includes manipulating decision mode as a within- and between-subjects variable, using an instructions and time-pressure manipulation, and two different types of sports

forecasts. Likewise, we see no evidence that individual differences in reliance on gut feelings and reasoned analysis affects the accuracy of sports forecasts. Across these studies we also observed the effect of decision mode on how confident people are in their forecasts. Therefore, based on the results reported here, sports forecasting is likely a type of decision that falls outside of the influence of decision mode.

REFERENCES

Aarnio, K., & Lindeman, M. (2005). Paranormal beliefs, education, and thinking styles. *Personality and Individual Differences, 39*(7), 1227–1236. https://doi.org/10.1016/j.paid.2005.04.009

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychon Bull Rev 19***,** 1187–1192. https://doi.org/10.3758/s13423-012-0305-z

Amaya, A., Bach, R., Keusch, F., & Kreuter, F. (2021). New data sources in social science research: Things to know before working with reddit data. *Social Science Computer Review*, *39*(5), 943–960. https://doi.org/10.1177/0894439319893305

Bastick. T. (1982). *Intuition: How we think and act*. J. Wiley.

Betsch, C. (2004). Präferenz für Intuition und Deliberation (PID) : Inventar zur erfassung von affekt- und kognitionsbasiertem entscheiden [Preferences for Intuition and Deliberation (PID): An inventory for assessing affect- and cognition based decision-making]. Zeitschrift für Differentielle und Diagnostische Psychologie, *25*, 179-197.

Betsch, C. (2007) Chronic preferences for intuition and deliberation in decision making: lessons learned about intuition from an individual differences approach. In Plessner, H., Betsch, C., & Betsch, T. (2007) *Intuition in Judgment and Decision Making* (pp. 231–248). Psychology Press. https://doi.org/10.4324/9780203838099-24

Betsch, C., & Iannello, P. (2010). Measuring individual differences in intuitive and deliberate decision-making styles: A comparison of different measures. In Glockner, A. & Witteman, C. *Foundations for tracing intuition* (pp. 259–279). Psychology Press. https://doi.org/10.4324/9780203861936-18

Betsch, C., & Kunz (2008). Individual strategy preferences and decisional fit. *Journal of Behavioral Decision Making*, *21*(5), 532-555. https://doi.org/10.1002/bdm.600

Bruin, W. B., Parker, A. M., & Fischhoff, B. (2007). individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*(5), 938–956. https://doi.org/10.1037/0022-3514.92.5.938

Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology*, *55*(1), 803-832. https://doi.org/10.1146/annurev.psych.55.090902.141601

Caplan, R. D. (1987). Person-environment fit theory and organizations: Commensurate dimensions, time perspectives, and mechanisms. *Journal of Vocational Behavior*, *31*(3), 248-267. https://doi.org/10.1016/0001-8791(87)90042-X

Cook, C., & Gonzales, H. (2016). Australian individual decision styles, intuitive and rational decision making in business. *International Proceedings of Economic Development and Research*, *86*, 93-100.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*(5), 819–829. https://doi.org/10.1037/0022-3514.66.5.819

Dijksterhuis, A. J. (2004). Think Different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87*(5), 586–598. https://doi.org/10.1037/0022-3514.87.5.586

Dijksterhuis, A. J., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1*(2), 95–109. https://doi.org/10.1111/j.1745-6916.2006.00007.x

Dijksterhuis, A. J., & van Olden, Z. (2006). On the benefits of thinking unconsciously: Unconscious thought can increase post-choice satisfaction. *Journal of Experimental Social Psychology*, *42*(5), 627–631. https://doi.org/10.1016/j.jesp.2005.10.008

Epstein, S. (1990). Cognitive-experiential self-theory. In Pervin, L., H*andbook of personality theory and research* (pp. 165-192). Guilford Press.

Epstein, S. (2003). Cognitive-experiential self-theory of personality. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology*, Vol. 5, pp. 159–184). John Wiley & Sons, Inc. https://doi.org/10.1002/0471264385.wei0507

Epstein, S. (2010). Demystifying intuition: What it is, what it does, and how it does it. *Psychological Inquiry*, *21*(4), 295–312. https://doi.org/10.1080/1047840X.2010.523875

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of*

Personality and Social Psychology, *71*(2), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Evans, B. T., & Curtis-Holmes, J. (2007). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382-389. https://doi.org/10.1080/13546780542000005

Frederick. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gärtner, M., Andersson, D., Västfjäll, D., & Tinghög, G. (2022). Affect and prosocial behavior: The role of decision mode and individual processing style. *Judgment and Decision Making*, *17*(1), 1-13. https://doi.org/10.1017/S1930297500008998

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

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

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Halberstadt, J. B. & Hooton, K. (2008). The affect disruption hypothesis: The effect of analytic thought on the fluency and appeal of art. *Cognition and Emotion*, *22*(5), 964–976. https://doi.org/10.1080/02699930701597668

Halberstadt, J. B., & Levine, G. M. (1999). Effects of reasons analysis on the accuracy of predicting basketball games. *Journal of Applied Social Psychology*, *29*(3), 517–530. https://doi.org/10.1111/j.1559-1816.1999.tb01399.x

Halberstadt, J. B., & Wilson, T.D. (2008). Reflections on conscious reflection: Mechanisms of impairment by reasons analysis. In Adler, J. E. & Rips, L. J. *Reasoning: Studies of human inference and its foundations*. (pp. 548-565). Cambridge University Press.

Higgins, T. E. (2000). Making a good decision: Value from fit. *The American Psychologist*, *55*(11), 1217–1230. https://doi.org/10.1037/0003-066X.55.11.1217

Higgins, T. E. (2005). Value from Regulatory Fit. *Current Directions in Psychological Science*, *14*(4), 209–213. https://doi.org/10.1111/j.0963-7214.2005.00366.x

Higgins, T. E., Idson, L. C., Freitas, A. L., Spiegel, S., & Molden, D. C. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology*, *84*(6), 1140–1153. https://doi.org/10.1037/0022-3514.84.6.1140

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* Free Press.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of cognitive neuroscience*, *9*(4), 462–475. https://doi.org/10.1162/jocn.1997.9.4.46

Kahneman. (2011). *Thinking, fast and slow* (1st ed.). Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*(1), 1–24. https://doi.org/10.1006/jmla.1993.1001

Kirkpatrick L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *63*(4), 534–544. https://doi.org/10.1037/0022-3514.63.4.534

Levine, G. M., Halberstadt, J. B., & Goldstone, R. L. (1996). Reasoning and the weighting of attributes in attitude judgments. *Journal of Personality and Social Psychology, 70*(2), 230-240. http://doi.org/10.1037/0022-3514.70.2.230

MacGregor, D., Lichtenstein, S., & Slovic, P. (1988). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes*, *42*(3), 303–323. https://doi.org/10.1016/0749-5978(88)90003-9

McMackin, J., & Slovic, P. (2000). When does explicit justification impair decision making? *Applied Cognitive Psychology*, *14*(6), 527–541. https://doi.org/10.1002/1099-0720(200011/12)14:6<527::aid-acp671>3.0.co;2-j

Mikels, J. A., Maglio, S. J., Reed, A. E., & Kaplowitz, L. J. (2011). Should I go with my gut? Investigating the benefits of emotion-focused decision making. *Emotion*, *11*(4), 743–753. https://doi.org/10.1037/a0023986

Nisbett, R., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/10.1037//0022-3514.35.4.250

Nordgren, L.F., & Dijksterhuis, A. J. (2008). The Devil Is in the deliberation: Thinking too much reduces preference consistency. *The Journal of Consumer Research*, *36*(1), 39–46. https://doi.org/10.1086/596306

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00147

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*(6), 972–987. https://doi.org/10.1037/0022-3514.76.6.972

Plessner, H. & Czenna, S. (2007) The Benefits of Intuition. In Plessner, H., Betsch, C., & Betsch, T. (2007) *Intuition in Judgment and Decision Making* (pp. 251–265). Psychology Press. https://doi.org/10.4324/9780203838099-24

Raab, M., & Laborde, S. (2011). When to blink and when to think: preference for intuitive decisions results in faster and better tactical choices. *Research Quarterly*

for Exercise and Sport, *82*(1), 89–98. https://doi.org/10.1080/02701367.2011.10599725

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richetin, J., Perugini, M., Adjali, I., & Hurling, R. (2007). The moderator role of intuitive versus deliberative decision making for the predictive validity of implicit and explicit measures. *European Journal of Personality*, *21*(4), 529–546. https://doi.org/10.1002/per.625

Risen, J. L. (2016). Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions. *Psychological Review, 123*(2), 182-207. http://doi.org/10.1037/rev0000017

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82*(3), 416-425. http://doi.org/10.1037/0021-9010.82.3.416

Sadler-Smith, E. (2004). Cognitive style and the management of small and medium-sized enterprises. *Organization Studies*, *25*(2), 155–181. https://doi.org/10.1177/0170840604036914

Sanchez, C., & Dunning, D. (2020). Decision fluency and overconfidence among beginners. *Decision, 7*(3), 225-237. https://doi.org/10.1037/dec0000122

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variable components. *International Biometric Society*, *2*(6), 110-114. https://doi.org/10.2307/3002019

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71. https://doi.org/10.1016/0010-0285(90)90003-M

Schunk, D., & Betsch, C. (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology*, *27*(3), 386–401. https://doi.org/10.1016/j.joep.2005.08.003

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818-831. https://doi.org/10.1177/0013164495055005017

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General, 135*(3), 409-428. http://doi.org/10.1037/0096-3445.135.3.409

Sladeck R. M., Bond, M. j., & Phillips, P. A. (2010) Age and gender difference in preferences for rational and experiential thinking. *Personality and Individual Differences, 49*(8), 907-911. https://doi.org/10.1016/j.paid.2010.07.028

Spann, M., & Skiera, B. (2009). Sports Forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, *28,* 55-72. https://doi.org/10.1002/for.1091

Walco, D. K., & Risen, J. L. (2017). The empirical case for acquiescing to intuition. *Psychological Science, 28*(12), 1807–1820. https://doi.org/10.1177/0956797617723377

Ward, S. J., & King L. A. (2020). Examining the roles of intuition and gender in magical beliefs. *Journal of Research in Personality, 86*, 103956. https://doi.org/10.1016/j.jrp.2020.103956.

Welsh, M. B., Delfabbro, P. H., Burns, N. R., & Begg, S. H. (2014). Individual differences in anchoring: Traits and experience. *Learning and Individual Differences, 19*, 131-140. https://doi.org/10.1016/j.lindif.2013.01.002

Wilson, T. D., & LaFleur, S. J. (1995). Knowing what you'll do: Effects of analyzing reasons on self-prediction. *Journal of Personality and Social Psychology, 68*(1), 21-35. http://doi.org /10.1037/0022-3514.68.1.21

Wilson, T. D., Hodges, S. D., & LaFleur, S. J. (1995). effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology*, *69*(1), 16–28. https://doi.org/10.1037/0022-3514.69.1.16

Wilson, T.D. & Dunn, D. S. (1986). Effects of introspection on attitude-behavior consistency: Analyzing reasons versus focusing on feelings. *Journal of Experimental Social Psychology*, *22*(3), 249–263. https://doi.org/10.1016/0022-1031(86)90028-4

Wilson, T.D. & Hodges, S. D. (1992). Attitudes as temporary constructions. In *The construction of social judgments* (pp. 37–66).

Wilson, T.D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*(2), 181–192. https://doi.org/10.1037/0022-3514.60.2.181

Wilson, T.D., Kraft, D., & Dunn, D. S. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology*, *25*(5), 379–400. https://doi.org/10.1016/0022-1031(89)90029-2

Wilson, T.D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality & Social Psychology Bulletin*, *19*(3), 331–339. https://doi.org/10.1177/0146167293193010

Zakay, D. (1985). Post-decisional confidence and conflict experienced in a choice process. *Acta Psychologica*, *58*(1), 75–80. https://doi.org/10.1016/0001-6918(85)90035-6

Zakay, D., & Tsal, Y. (1993). The impact of using forced decision-making strategies on post-decisional confidence. *Journal of Behavioral Decision Making*, *6*(1), 53–68. https://doi.org/10.1002/bdm.3960060104

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica*, *98*(1), 103-115. https://doi.org/10.1016/S0001-6918(97)00037-1

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research

conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493-504. https://doi.org/10.1037/pspa0000056

APPENDICES

Appendix A Rational and Experiential Self Inventory (REI)

Rationality scale (REI-R)

1. I try to avoid situations that require thinking in depth about something (re-)
2. I'm not that good at figuring out complicated problems (ra—)
3. I enjoy intellectual challenges (re)
4. I am not very good at solving problems that require careful logical analysis (ra—)
5. I don't like to have to do a lot of thinking (re-)
6. I enjoy solving problems that require hard thinking (re)
7. Thinking is not my idea of an enjoyable activity (re—)
8. I am not a very analytical thinker (ra—)
9. Reasoning things out carefully is not one of my strong points (ra—)
10. I prefer complex problems to simple problems (re)
11. Thinking hard and for a long time about something gives me little satisfaction (re)
12. I don't reason well under pressure (ra—)
13. I am much better at figuring things out logically than most people (ra)
14. I have a logical mind (ra)
15. I enjoy thinking in abstract terms (re)
16. I have no problem thinking things through carefully (ra)
17. Using logic usually works well for me in figuring out problems in my life (ra)
18. Knowing the answer without having to understand the reasoning behind it is good enough for me (re-)

19. I usually have clear, explainable reasons for my decisions (ra)
20. Learning new ways to think would be very appealing to me (re)

Experientiality scale (REI-E)

1. I like to rely on my intuitive impressions (ee)
2. I don't have a very good sense of intuition (ea—)
3. Using my gut feelings usually works well for me in figuring out problems in my life (ea)
4. I believe in trusting my hunches (ea)
5. Intuition can be a very useful way to solve problems (ee)
6. I often go by my instincts when deciding on a course of action (ee)
7. I trust my initial feelings about people (ea)
8. When it comes to trusting people, I can usually rely on my gut feelings (ea)
9. If I were to rely on my gut feelings, I would often make mistakes (ea—)
10. I don't like situations in which I have to rely on intuition (ee-)
11. I think there are times when one should rely on one's intuition (ee)
12. I think it is foolish to make important decisions based on feelings (ee—)
13. I don't think it is a good idea to rely on one's intuition for important decisions (ee)
14. I generally don't depend on my feelings to help me make decisions (ee-)
15. I hardly ever go wrong when I listen to my deepest gut feelings to find an answer (ea)
16. I would not want to depend on anyone who described himself or herself as intuitive (ee-)

17. My snap judgments are probably not as good as most people's (ea-)

18. I tend to use my heart as a guide for my actions (ee)

19. I can usually feel when a person is right or wrong, even if I can't explain how I know (ea)

20. I suspect my hunches are inaccurate as often as they are accurate (ea—)

Note: The REI-40 was developed by Pacini and Epstein (1999). "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" (Pacini & Epstein, 1999, p. 976). Items beginning with r are part of the rational subscale while those beginning with e are part of the experiential subscale. Items are assessed on a scale from 1 (definitely not true of myself) to 5 (definitely true of myself).

Appendix B General Decision-Making Scale (GDMS)

Rational Subscale (GDMS-R)

1. I double-check my information sources to be sure I have the right facts before making a decision
2. I make decisions in a logical and systematic way
3. My decision making requires careful thought
4. When making a decision. I consider various options in terms of a specific goal
5. I explore all of my options before making a decision

Intuitive Subscale (GDMS-I)

1. When making a decision. I rely upon my instincts
2. When I make decisions. I tend to rely on my intuition
3. I generally make decisions that feel right to me
4. When I make a decision. it is more important for me to feel the decision is right than to have a rational reason for it
5. When I make a decision. I trust my inner feelings and reactions

Note: The GDMS was developed by Scott and Bruce (1995). The original GDMS includes additional subscales for dependent, avoidant, and spontaneous decision making styles not included here. Items are assessed on a 5 point scale from 0 (strongly disagree) to 5 (strongly agree).

Appendix C Preference for Intuition and Deliberation Scale (PID)

Preference for Deliberation (PID-D)

1. Before making decisions I first think them through
2. Before making decisions I usually think about the goals I want achieve
3. I consider myself
4. I prefer making detailed plans rather than leaving things to chance
5. I am a perfectionist
6. I think about a decision particularly carefully if I have to justify it
7. When I have a problem I first analyze the facts and details before I decide
8. I think before I act
9. I think more about my plans and goals that other people do

Preference for Intuition (PID-I)

1. I listen carefully to my deepest feelings
2. With most decisions it makes sense to complexly rely on your feelings
3. I don’t like situations that require me to rely on my intuition (-)
4. I prefer drawing conclusions based on my feelings, my knowledge of human nature, and my experience of life
5. My feelings play an important role in my decisions
6. When it comes to trusting people, I can usually rely on my gut feelings
7. I prefer emotional people
8. I am a very intuitive person
9. I like emotional situations, discussions, and movies

Note: The PID was developed by Betsch (2004). Items marked with (-) are reverse coded. Items are assessed on a scale from 1 (very much disagree) to 5 (very much agree).

Appendix D Study 1 Forecasts and Game Outcome

	Final score	Margin of victory	Experts' expectations	Favorite's over-performance	Percent of correct picks		Misprediction		Deviation from experts' expectations		Confidence	
					Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis
Study 1A												
Bengals vs. Raiders	26-19	7	6.5	0.5	84	81	5.6	5.9	5.7	6.0	65.7	65.9
Bills vs. Patriots	47-17	20	4.5	15.5	73	64	26.0	27.3	5.8	6.9	69.2	70.5
Buccaneers vs. Eagles	31-15	16	9.0	7.0	94	95	6.2	6.0	5.2	5.5	85.5	87.4
Chiefs vs. Steelers	42-21	21	12.5	8.5	94	93	8.4	9.2	5.3	5.6	88.7	87.2
Cowboys vs. 49ers	17-23	-6	3.0	-3.0	35	33	10.2	11.3	6.4	7.3	68.4	70.7
Rams vs. Cardinals	34-11	23	4.0	19.0	73	75	19.2	18.4	6.0	5.9	65.8	67.7
Study 1B												
Bengals vs. Raiders	26-19	7	6.5	0.5	89	90	4.4	4.8	4.5	5.0	73.5	74.9
Bills vs. Patriots	47-17	20	4.5	15.5	71	65	26.4	27.0	5.1	5.7	67.1	67.6
Buccaneers vs. Eagles	31-15	16	9	7.0	81	80	8.5	8.5	5.9	6.8	81.8	84.9
Chiefs vs. Steelers	42-21	21	12.5	8.5	94	91	7.9	9.4	5.3	5.2	90.8	86.3
Cowboys vs. 49ers	17-23	-6	3	-3.0	51	45	8.6	9.8	7.4	7.2	68.0	71.5
Rams vs. Cardinals	34-11	23	4	19.0	73	80	15.6	15.0	5.4	5.4	67.7	68.3

Appendix E Study 2 Forecasts and Game Outcome

	Final score	Margin of victory	Experts' expectations	Favorite's over-performance	Percent of correct picks		Misprediction		Deviation from experts' expectations		Confidence	
					Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis
Study 1A												
Bengals vs. Raiders	26-19	7	6.5	0.5	54	57	4.58	4.56	4.62	4.62	70.6	69.8
Bills vs. Patriots	47-17	20	4.5	15.5	48	44	25.2	25.6	4.71	4.39	72.8	71.3
Buccaneers vs. Eagles	31-15	16	9.0	7.0	57	50	7.55	8.37	6.13	6.12	76.6	75.1
Chiefs vs. Steelers	42-21	21	12.5	8.5	59	57	8.35	8.94	5.72	6.72	77.3	75.9
Cowboys vs. 49ers	17-23	-6	3.0	-3.0	49	51	9.36	9.29	6.08	6.06	72.1	72.1
Rams vs. Cardinals	34-11	23	4.0	19.0	45	54	19.4	18.5	5.1	5.02	70.2	72
Study 1B												
Bengals vs. Raiders	26-19	7	6.5	0.5	57	60	4.85	4.72	4.91	4.8	70.7	72.2
Bills vs. Patriots	47-17	20	4.5	15.5	44	40	16	16.8	5.03	4.96	69.5	72.9
Buccaneers vs. Eagles	31-15	16	9	7.0	55	55	6.94	7.14	5.47	6.04	76.6	78.3
Chiefs vs. Steelers	42-21	21	12.5	8.5	60	61	7.95	8.99	5.44	7.3	80.2	80.3
Cowboys vs. 49ers	17-23	-6	3	-3.0	55	53	8.74	9.13	6.74	6.25	74.6	73.3
Rams vs. Cardinals	34-11	23	4	19.0	59	58	18.4	18.1	5.04	5.73	70.2	72.7

Appendix F Study 3 Forecasts and Game Outcome

	Final score	Margin of victory	Experts' expectations	Favorite's over-performance	Percent of correct picks		Misprediction		Deviation from experts' expectations		Confidence	
					Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis	Gut feeling	Reasoned analysis
Buccaneers vs. Rams	27-30	-3	2.5	-5.5	59	61	5.62	5.54	5.14	5.49	4.36	4.33
Chiefs vs. Bills	42-36	-6	1.5	-7.5	31	33	8.18	8.79	5.38	6.05	4.34	4.39
Packers vs. 49ers	10-13	-3	5.5	-8.5	45	45	9.79	9.52	5.23	5.38	4.71	4.76
Titans vs. Raiders	16-19	-3	3.5	-6.5	56	56	6.49	6.48	5.35	5.67	4.52	4.41

VITA

Andrew Thomas Langbehn grew up in Portage, Michigan. After graduating from high school, he attended Kalamazoo Valley Community College where he completed an Emergence Medical Technology Basic Certificate attending the University of Wisconsin – Madison where he graduated with a Bachelor of Arts degree in Psychology and a minor in Criminal Justice in 2020. He chose to attend the University of Tennessee, Knoxville and began pursuing a Doctor of Philosophy degree in Experimental Psychology the following semester. Broadly, his research seeks to uncover the complexity of human emotions including how people experience and communicate emotions and how emotions and feelings shape behavior. He is thankful for all the support from his family and from members of the University of Tennessee, Knoxville community.