Summer 2014

Examining the roles of frame, frequency, and relevance in performance feedback: exploring evaluative and behavioral outcomes of decision making

Elaine Ardis Bossard
University of Iowa

EXAMINING THE ROLES OF FRAME, FREQUENCY, AND RELEVANCE IN PERFORMANCE FEEDBACK: EXPLORING EVALUATIVE AND BEHAVIORAL OUTCOMES OF DECISION MAKING

by

Elaine Ardis Bossard

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

August 2014

Thesis Supervisor: Professor Emeritus Irwin P. Levin

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Elaine Ardis Bossard

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychology at the August 2014 graduation.

Thesis Committee: ___________________________________
Irwin P. Levin, Thesis Supervisor

____
Paul D. Windschitl

____
Jason K. Clark

____
Rebecca Neel

____
Gary J. Gaeth

ACKNOWLEDGMENTS

I would like to thank Dr. Irwin Levin for his guidance and support throughout my graduate school years and during the production of my dissertation. I am also grateful for the helpful suggestions and contributions provided by my committee members Paul Windschitl, Jason Clark, Gary Gaeth, and Rebecca Neel. A special thanks to the many graduate and undergraduate students that have played a role in my development as a researcher in the past several years. In particular, I want to thank all the undergraduate research assistants in the Levin Decision Making Lab that aided in data collection for these dissertation studies and the social psychology area graduate students (plus Matt Calamia) for acting as lab mates and offering a steady stream of support, intellectual stimulation, and fun during my time in graduate school.

ABSTRACT

Feedback is often necessary to provide guidance for future decisions, and factors relating to feedback, including the way feedback information is framed, how frequently it is provided, and the relevance of that feedback in relation to one's decision, have been designated as influential for decision making tendencies. Unfortunately, research on what produces the most effective feedback is mixed, and the relationship between these factors and resulting evaluative and behavioral outcomes is less clear.

Four studies explored the relationship between feedback frame and frequency by addressing whether overall task feedback framed positively and receiving more frequent trial outcome feedback led to more positive performance evaluations and improvements in subsequent task performance (Studies 1A and 1B), how these evaluative and behavioral outcomes varied across different trial feedback frequency intervals (Study 2A), and whether more relevant trial feedback influenced the pattern of these results (Study 2B).

Across the four studies, it was noted that the frequency of trial feedback was more influential for task performance outcomes, while the overall task feedback frame was more influential for performance evaluation outcomes. In addition, more relevant outcome feedback was seen to influence the relationship of feedback variables more for performance evaluations than task performance. Taken together, these studies provide some clarity as to how different types and presentations of feedback produce different evaluative and behavioral outcomes and show initial direction as to when framing task feedback and providing trial feedback more frequently can lead to better, more normatively correct decision making. Theoretical and practical implications, as well as reasons why effects were not consistent across studies, are also discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

People are presented every day with decisions and choices to make, and often, they do not know what the outcome of their decision or choice will be, even if they know the probability associated with each decision outcome. Because there is an element of uncertainty surrounding most choice scenarios, sometimes people choose options that result in more beneficial outcomes and sometimes people choose options that result in more detrimental outcomes; therefore¸ feedback about one's choices can be key to guiding future choices and outcomes.

For example, consider a person during the probation period of a new job. In some cases, a new hire is simply provided with a protocol and guidelines to follow with an assessment at the end of the probation. In other cases, the new hire is provided with training and feedback throughout this period before being assessed. In the former case, the person knows what needs to be accomplished in order to fulfill the position, and based on the protocols received, likely has some idea of the probability of success; however in the latter case, the person may not know the exact contingencies associated with successfully fulfilling the position, and must rely on the provided feedback to discern the probability of this achievement. In either case, the person has to make decisions about how he/she wants to proceed and act on the job (e.g., whether to work overtime or not, whether to deviate from given instructions).

Many kinds of information help humans regulate their thoughts and behavior in order to make better decisions and obtain more optimal outcomes, but information people receive about their past performance has been especially regarded as beneficial for guiding future performance. Most commonly, feedback serves to signal whether people have correctly retrieved or applied information to their thoughts and behaviors (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Indeed, informational feedback can be considered to be a form of reinforcement working to confirm or disconfirm expectancies and provide both motivation and information to correct or improve performance (McClenaghan & Ward, 1987).

What is less well known is why people choose one course of action over another, what influences people to take that course, and how better choices can be encouraged and attained. Though previous literature exhibits some debate and inconsistency on this manner across a variety of domains, three important aspects of feedback in determining more beneficial thoughts and behaviors seem to be the presentation frame, frequency, and relevance of the feedback information one receives.

In the cases above, the person is provided with some overall feedback of his/her previous performance, which could be presented as positive (e.g., completing an assignment or making a sale 75% of the time) or negative (e.g., failure to complete an assignment or losing a sale 25% of the time). The difference in positive versus negative presentations of previous performance suggests that the framing of overall task feedback information (e.g., how much one earned versus failed to earn throughout a task) may be an important variable in investigating different decision choices, evaluations of these choices, and perhaps even eventual changes for better decision making and performance.

In addition, the cases above may represent differences in how knowledge of the probabilities and outcomes associated with various decision choices is obtained and in the proximity with which outcome information is provided following one's choices. On one hand, the new hire may know the contingencies and outcomes for decision choices associated with performing the job based only on the job description and initial instructions; whereas, on the other hand, the new hire may learn the contingencies and outcomes of his/her decision choices associated with performing the job through the direct experience of them and resulting feedback from a supervisor. Differences in how one gains knowledge of the contingencies of various actions, in part, due to the different frequency of receiving feedback about one's decisions and performance on the job may influence how one takes in that feedback and whether the individual feels any control over creating different outcomes in the future. This suggests that the different frequency intervals in which one receives outcome feedback about his/her decisions may also represent an important variable in understanding different decision choices, evaluations of these choices, and changes for better decision making and performance.

Finally, the relevance of the feedback can differ in the cases above. Even when decision feedback is provided, exactly what information is provided in that feedback can determine how it is received and utilized. If the feedback is too generic or only provides information about the outcome, but does not indicate how to change one's actions to improve those outcomes (e.g., "Your work pace is satisfactory"), then the feedback may not provide much guidance for subsequent actions. Making a clear connection between decisions made and decision outcomes allows for potential use of that information in a corrective manner on future decisions, so differences in the relevance of provided feedback may influence how the feedback is interpreted and applied. The level of decision feedback relevance for guiding subsequent thoughts and decisions suggests that taking into consideration how corrective or instructive the decision outcome feedback is in regards to previous choices may be important for investigating different decision choices, evaluations of these choices, and changes for better decision making and performance.

These factors are important to consider when attempting to understand the role of feedback presentation and interpretation during a risky decision making task, as feedback can come in a variety of formats and can contribute to evaluative and behavioral outcomes regarding one's performance. As real-world scenarios often provide frequent and repeated feedback (e.g., weekly quizzes in school or progress reports on a job), using a repeated-choice design for the decision making task of interest in the current studies may better simulate these types of settings and allow for better generalizability of any noted findings. To summarize, the current studies will endeavor to bring clarity to the understanding of specific effects associated with feedback frame, frequency, and relevance and how feedback can affect more general decision making tendencies related to evaluations, decision choices, and task performance. In particular, the current studies will extend the knowledge of what types of outcomes (evaluative and/or behavioral) can be associated with different feedback presentations and what conditions may lead to changes in one's tendency for normatively correct decision making in his/her choices and performance, ultimately leading to better decision making skills.

Influence of Framing Task Performance Feedback

Feedback that focuses on outcomes of performance or on a way in which a task was performed, suggests that there is an evaluative component to that feedback. A negativity bias has been noted in information processing research, such that people pay more attention to and are influenced more by negative information (for reviews, see Fiske & Taylor, 1991; Taylor, 1991); however, some research has found that negative performance feedback does not always follow this pattern. Despite a general tendency for people to believe that evaluations and performance can be improved *only* by information showing how one fell short of a standard (exemplifying negative feedback information), negative feedback can often lead to the opposite effect (Kluger & DeNisi, 1996). Similar effects can be seen across both precise scales (e.g., 98 out of 180 points) and imprecise statements (e.g., "nice job!"; Ilgen & Davis, 2000).

There are many ways in which to present positively and negatively framed feedback, ranging from the number of responses that are correct versus incorrect to achievement versus failure to how many people one was better than versus worse than. As task performance feedback can be framed in a variety of manners, it is important to note that a multitude of variables can play a role in what frame a decision maker falls prey to or adopts, including norms, habits, and personal characteristics (Kahneman & Miller, 1986; Kühberger, 1998; Levin, Schneider, & Gaeth, 1998); however, the dominating factor is how that decision is described.

In spite of the range of possible frames, several different domains of research have found evidence to suggest that positively framing performance feedback can produce positive outcomes. This relationship has been noted separately for both how people evaluate their performance and how they alter their behavior in beneficial ways, though the relationship has more mixed effects for behavioral outcomes. Indeed, a meta-analysis of different feedback manipulations and performance outcomes in the domain of education found that performance feedback is more powerful and produces better outcomes in later task performance if it is provided in a positively framed versus negatively framed manner (Kluger & DeNisi, 1996; see also Hattie & Timperley, 2007 for an overview). It should be noted that a few moderators (e.g., prevention focus of task or individual – Kluger & DeNisi, 1996; Van Dijk & Kluger, 2011) of this general effect have been identified that showcase when negatively framed performance feedback can be beneficial; however, as these moderators are not directly applicable to the current study or task designs (the task is promotion-focused and participants are randomly assigned across conditions), they will not be discussed further.

In related studies, framing health messages in a positive manner was more effective for eliciting preventative health behaviors, and thus obtaining more positive health outcomes, than framing them in a negative manner (Gallagher & Updegraff, 2012; Rothman & Salovey, 1997). In the same studies, however, it was noted that framing health messages in a negative manner could also be effective for obtaining more positive health outcomes, but this effect was weaker and only occurred when eliciting a different type of behavior - illness detection behavior.

While message framing includes a much wider scope of frames, this pattern of effects in conjunction with the previously noted meta-analysis suggests that applying positive or negative framing to overall task performance feedback could also be influential for encouraging decision making to achieve the best outcomes during subsequent task performance. As will be described in more detail later on, the main task and outcomes of interest in the current studies may be more similar to preventative health behaviors, as the behavioral outcome of interest relies on identifying and choosing advantageous options (similar to participating in healthy behaviors), while avoiding disadvantageous options (similar to avoiding unhealthy behaviors). Like preventative health behaviors, promoting making the best choices will help prevent unwanted outcomes, which may be enhanced by positive framing compared to negative framing.

Valence Framing

One of the most common variations on manipulating task feedback is through valence framing. Valence framing is a type of information framing that occurs when the manner in which the information provider's representation of questions, potential answers, actions, or outcomes related to given decision options is presented positively or negatively. The frame valence casts a positive or negative light on that information, which then affects one's evaluations of these things (Kühberger, 1998; Levin, Schneider, & Gaeth, 1998).

As initially described by Kahneman and Tversky's Prospect Theory (1979), when applied to decision making, deviations from a reference point for judgments and decisions framed as either gains or losses can extend to influence the decision maker's ideas regarding given decision options and risk-seeking behavior. As an example, while a majority of people would prefer to face winning $3200 for certain than an 80% chance of winning $4000, a majority of people would also prefer to face an 80% chance of losing $4000 than losing $3200 for certain (Tverksy & Kahneman, 1981). Despite the fact that all these options have the same expected value outcomes (0.8x4000 = 1.0x3200, with a net gain or loss of $3200), one's actions seem to depend on how the scenario is framed.

In essence, Prospect Theory suggests that when people evaluate an uncertain choice of action, they focus on the amount of gain or loss with which they will end up due to their choice/the event and assign values to their choices using that information.

Thus, one of the more important contributions that the theory has provided is the idea that choices are dependent on both how the decision options are presented and how the decision maker interprets that presentation.

A multitude of studies have since helped to demonstrate that information presented in differently valenced frames, or put more generally, a valence framing effect, elicits a situation in which objectively equivalent information in positive versus negative presentations systematically influences what actions or choices a decision maker takes. The standard valence framing effects are rather robust throughout a variety of decision making domains, including gambling decisions, consumer decisions, and health decisions, and can be broken down into three categories: risky choice framing, goal framing, and attribute framing (see Kühberger, 1998 and Levin et al., 1998 for reviews).

Attribute framing. While it is important to note and maintain distinctions between the different types of valence framing, the focus of the current studies is on attribute framing as a mode to present task feedback. As set forth by Levin et al. (1998), attribute framing tends to be the most simple and straightforward of valence framing types, in which a single characteristic of an event or object is the subject of the framing manipulation. This framing manipulation often takes the form of labeling something in a positive light (e.g., 85% lean ground beef) versus a negative light (e.g., 15% fat ground beef) or describing situations in terms of success rates (e.g., 65% of colds cured) versus failure rates (e.g., 35% of colds uncured). Each frame elicits a different evaluation of the event or object, with more favorable evaluations for the positive frame and less favorable evaluations for the negative frame – known as a "valence-consistent shift" (Levin et al., 1998). These differences in evaluations can also lead to different choices and actions. The effect of this frame valence is determined by assessing the difference in evaluations given to the event or object attributes across frame conditions.

Predominantly, attribute framing effects are suggested as originating from how the information received is encoded in relation to its frame valence (Levin & Gaeth, 1988). Positive labeling leads to encoding that evokes favorable associations in one's memory, and negative labeling leads to encoding that evokes more unfavorable associations in one's memory. The shift in response based on frame valence may be caused by these representational differences in associative memory. This idea also suggests that framing a specific characteristic may act similarly to a prime (except that it is not peripheral to the description of the target stimulus), helping to set up the evaluative tone that determines whether the valence of the accessed knowledge is positive or negative (Wyer & Srull, 1989). Evidence for this was found in a series of studies (Levin & Gaeth, 1988; Levin, Johnson, Russo, & Deldin, 1985) demonstrating that positive (75% lean) compared to negative (25% fat) attribute labels of ground beef lead to valence-consistent evaluations of not just the dimension of fat/lean, but also associated dimensions of taste, greasiness, and overall quality.

Employing attribute frames as a vehicle for framing performance feedback could, then, promote valence-consistent performance evaluations of dimensions associated with one's overall score in addition to evaluations of one's actual score. More specifically, positive task feedback frames may lead to a focus on what was done correctly, while negative task feedback frames may lead to a focus on what was done incorrectly.

Other accounts exist regarding the origins of attribute framing effects (for example, see McKenzie & Nelson, 2003), centered on the idea that different attribute frames may directly imply different reference points (through different wordings) for making choices and evaluations rather than the more indirect route of creating different representations in associative memory. These ideas are supported by previous work advising that alternative terminology (e.g., gain versus loss) may introduce unplanned differences in how frames are portrayed and interpreted as compared to simple negation (e.g., gain versus did not gain; Maheswaran & Meyers-Levy, 1990).

Though the difference between these two accounts of attribute framing is noteworthy, regardless of why attribute framing occurs, both accounts inevitably exhibit the same outcomes in terms of attribute framing effects, and the effects are reliable in producing significant results (for reviews see Levin et al., 1998 and Piñon & Gambara, 2005). Thus, the take-home point remains that positive and negative attribute frames can lead to different courses of action and evaluation.

In the following studies, the influence of attribute frames, presented as an outcome (task performance feedback), on guiding people's feelings about and interpretations of their performance is addressed using both simple negation and alternative term manipulations (see *Methods* - Study 1A and Study 1B for exact wording). These experiences may, in turn, influence normatively correct decision tendencies and affect future choices, evaluations, and subsequent performance outcomes. The current studies focus on attribute framing as a component of feedback presentation specifically because it manipulates a single aspect of the scenario in a positive or negative manner to influence how one determines evaluations and takes action. This represents how task performance feedback is commonly given (e.g., how many points earned versus how many points not earned), and it represents the simplest form of framing, allowing for a straightforward, yet relatively unobtrusive, test of positive and negative frame influence in the context of the current studies.

<u>Outcomes of attribute framing.</u> When considering attribute framing as an experimental manipulation of feedback presentation, one may also question what types of outcomes result. Most of the work previously completed on the outcomes of attribute framing has focused on evaluations and judgments, ranging from how much one likes a product (e.g., Levin et al., 1998; Linville, Fischer, & Fischhoff, 1993) to how likely one would be to win a gamble (e.g., Levin et al., 1998; Loke & Tan, 1992), and consistent findings have found that positive frames lead to more positive evaluations and judgments and negative frames lead to more negative evaluations and judgments. Additionally, Kuvaas and Selart (2004) have suggested that positive frames yield a larger change (in a

positive direction) in evaluations of decision outcomes than the change (in a negative direction) resulting from negative frames.

Undeniably, less work has been done investigating behavioral intentions and behavioral outcomes, and most of this work has been in consumer and health decision domains. For instance, the association between positively framed consumer products (e.g., ground beef) and better quality and higher pricing also contributes to a willingness to pay more per pound for that ground beef when its attributes have been framed positively, as compared to negatively (Levin et al., 2002; Levin et al., 1985). In the health decision domain, Howard and Salkeld (2009) found that framing a colorectal cancer screening option as finding one extra cancer (positively framed attribute), compared to one fewer cancers missed (negatively framed attribute) can increase one's willingness to pay for a screening and make one more willing to accept a higher amount of unnecessary colonoscopies (i.e., negative test results). Jasper, Goel, Einarson, Gallo, and Koren (2002) found similar results surrounding behavioral intentions to use allergy drugs during pregnancy. Given the variety of scenarios that can positively or negatively frame different attributes (for instance, the example presented earlier of starting a new job) and the potential outcomes of that framing (e.g., failing to satisfactorily complete an assignment), behavioral intentions and actual behaviors may also be important outcomes to consider in the current studies.

Attribute framing as feedback. One way that attribute framing may contribute to both evaluative and behavioral outcomes in relation to the current studies is by using it to differentially present overall task performance feedback. Of consequence to the current studies, positively framed feedback may be more likely to produce outcomes that suggest change in one's thoughts by increasing the positivity of one's performance evaluations (and relatedly, what one did correctly) and in one's behaviors by improving task performance.

Nygren's (1997) work supports this idea, showing that positively framing attributes of a decision making task can lead to better task performance and more optimal decision strategies than negatively framing attributes of the same task. Vollmeyer and Rheinberg (2005) also found that feedback regarding the number correct on a previous problem-solving task (a positive frame), rather than the number incorrect (a negative frame), led to better performance outcomes on later phases of the same task.

A similar pattern of attribute framing effects emerges when looking at the health behavior change literature. For example, positively framed performance feedback regarding the amount of achievement accomplished towards healthy behavior goals has been found to improve beliefs regarding and increase intentions for improvement on one's choices, as compared to negatively framed performance feedback regarding the amount of work still needed to achieve healthy behavior goals (Choe, Lee, Munson, Pratt, & Kientz, 2013). Though the authors do not address actual behavior change as an outcome measure, the contribution of positively framed performance feedback on intentions to change suggests that behavioral change may follow in a similar fashion.

Initial findings suggest that attribute frames can influence evaluative and behavioral outcomes, like subsequent task performance, in specific settings, but little work has shown how attribute frames can influence these types of outcomes in ways that improve more general decision making skills. Much of the previous research has used tasks where a right and wrong answer exist, but learning a set of rules to achieve the right answer, for example, does not necessarily provide participants with knowledge to induce improvements in decision making skills in other settings (like being able to discriminate between advantageous and disadvantageous options). The current studies expand on the previous research by focusing on how framing overall task feedback can potentially improve evaluations of and performance on a risky decision making task (described in the following section) that has normatively correct response options and has been found to related to real-world decision making tendencies (Brevers, Bechara, Cleeremans, Kornreich, Verbanck, & Noel, in press; Levin, Hart, Weller, & Harshman, 2007a).

The attribute framing applied to format task performance feedback will emulate the above studies, with the positive frame based on attainment of normatively correct responses and the negative frame based on the failure to attain normatively correct responses. More specifically, the current studies employ both a more traditional approach to attribute framing manipulations of feedback, using simple negation (i.e., earned versus failed to earn) and the same objective (false) feedback for each participant, and a more salient and generalizable approach, using opposing terms (i.e., gained versus lost) and actual earned scores for each participant.

The methods with which to showcase the contribution of task feedback frame for different outcomes vary, but by addressing both approaches to framing manipulations within the current studies, a more distinct idea of how attribute frames are affecting performance evaluations and later behavior, especially that which is associated with less biased decision making, can be addressed. However, in everyday decision making, both the manner in which feedback is given and *when* it is given are important factors. The current studies will address not just the effect of attribute framing as task feedback, but how framed task feedback interacts with another variable of interest – the frequency interval with which one receives information regarding task choice probabilities and outcomes (specifically, at the start of the task through explicit statements or through personal experience with them as trial feedback). Combining attribute framing with trial feedback at various intervals (discussed in more detail below) should provide people with a better sense of what their feedback means and how to use it to increase task performance by making more normatively correct decisions.

Current Studies – Manipulating Task Feedback Frame

Each study uses the Cups task as a way to measure the influence of different aspects of feedback presentation (i.e., frame, frequency, or relevance) on evaluative and behavioral outcomes in decision making scenarios. Briefly, the Cups task is used to understand how participants make decisions under differing probabilities of risk and outcome levels, separately for gains and losses, and overall performance measures have been shown to correlate up to three years later, indicating a reliability in measurement of one's decision tendencies (Levin et al., 2007a; Levin, Weller, Pederson, & Harshman, 2007b).

The Cups task also presents the advantage of a repeated gambles design. This helps contribute to the generalizability of a laboratory decision task to real-world decision making, as using aggregated data has shown evidence for stronger links between responses to decision making manipulations and other evaluations and behaviors (Epstein, 1979; Parker & Fischhoff, 2005). Therefore, using a repeated gambles design, such as the Cups task, is an important methodological consideration when relating decision making responses with evaluative and behavioral outcomes.

In the standard format, on each trial, participants are presented with both a set of riskless cups (in which every cup has the same outcome of one quarter) and a set of risky cups (in which one cup has an outcome of a specified number of quarters and the rest have zero quarters) from which to choose and are told to choose the option they think represents the best choice in order to earn more quarters. The number of cups presented for both sets vary in number, and subsequently, the probability of achieving a non-zero outcome when choosing the risky set of cups (see Appendix - Figures A1 and A2). Each combination of probability and potential outcome is presented multiple times, and participants receive immediate results of their choice (i.e., how much money they gained or lost), indicating trial-by-trial decision outcome feedback. The choice between the riskless and risky sets of cups on each trial allows for examination of both overall risk taking and risk sensitivity – which indicates the ability to differentiate between risks that are likely to lead to positive outcomes and risks that are likely to lead to negative outcomes (discussed in further detail in *Materials* – Study 1A).

Risk sensitivity scores serve as the main dependent variable of interest. More specifically, the focus lies on the difference between risk sensitivity scores on two consecutive rounds of the Cups task, with improvement on subsequent task performance noted by positive differences and decline in subsequent task performance noted by negative differences across rounds. By focusing on responses that are tied to normatively correct answers, rather than general risk taking, biased decision making tendencies – representing a type of decision making competence – can be measured; this allows for the important consideration of how factors, such as differences in framing overall task feedback, trial feedback frequency, and feedback relevance, can influence one's ability to become less biased – or more competent – in their decision making tendencies.

Pertinent to the current research design, results have previously shown that framing effects are present in the Cups task in terms of how different decisions are made throughout the task (Levin et al., 2007a; Levin et al., 2007b; Weller, Levin, & Denburg 2011), but no one has yet to investigate how framing overall task performance feedback can influence evaluations of and performance on the Cups task. Overall task performance feedback has thus far been provided as a positive task feedback frame (i.e., the amount of money that one has earned or gained throughout the task). Adding the condition of negatively framed task feedback (i.e., the amount of money that one has failed to earn or that one has lost; see *Materials* – Study 1A and Study 1B for exact wording), then, creates a starting point to examine the potential effects of attribute framing (as a component of task feedback) on both the evaluations of the task and changes in task performance for this paradigm. Attribute framing is implemented as different levels of an experimental factor (i.e., positively versus negatively framed task feedback) rather than as an outcome itself, so the methodology necessitates a between-subjects design and allows for the ability to draw more distinct conclusions from each frame condition. As such, in line with previous attribute framing effects (Kűhberger, 1998; Levin et al., 1998), supplying people with positively framed task feedback should lead to more positive associations with one's performance and the Cups task, perhaps in a way to focus on the positive aspect of one's normatively correct decisions, while supplying people with negatively framed task feedback should lead to more negative associations, perhaps in a way to focus on the negative aspect of one's normatively incorrect decisions. It is an empirical question as to which is more effective in improving performance on the Cups task, but given the current study design (a focus on earning the most quarters), positively framed task feedback should lead to more improved task performance on subsequent rounds, while negatively framed task feedback should lead to less improved task performance on subsequent rounds.

Influence of Trial Feedback Frequency

As feedback serves to signal whether people have correctly retrieved or applied information to their thoughts and behaviors (Bangert-Drowns et al., 1991), the proximity of decision feedback to one's choices and how often that feedback occurs may influence one's ability to recognize that signal. An important line of work from the educational psychology field has developed to address the distinction and effectiveness of immediate versus delayed feedback. Across multiple studies, it has generally been found that when feedback is provided in close proximity to one's performance regarding the correctness of their responses, students are able to assess how well they were performing, alter their approach to future tasks as needed, and show improvement in learning (Kulik & Kulik, 1988; Lhyle, & Kulhavy, 1987; Zahorik, 1987).

Unfortunately, the designation of feedback as immediate across studies is quite chaotic, ranging from feedback after every question or trial to feedback at the end of the test or task. The designation of delayed feedback is similarly chaotic, but generally occurs at some point after the test or task has been finished. Unsurprisingly, there remains some debate over the above general findings, and the literature investigating the effects of immediate versus delayed feedback in educational settings has reported differences in regards to the influence on performance evaluations and behaviors – particularly, that for more difficult tasks, immediate and corrective feedback is most beneficial in leading to improved learning (e.g., Clariana, Wagner, & Roher Murphy, 2000; Epstein & Lazarus, 2002; Epstein, Lazarus, Calvano, Matthews, Hendel, Epstein, & Brosvic, 2002).

When feedback frequency intervals have been considered in health messages, it has been found that having tailored, instructive information provided at numerous time points in the health behavior change process leads to more positive outcomes and longer term impacts (Brug, Glanz, van Assema, Kok, van Breukelen, 1998). At a minimum, it appears that more frequent feedback messages help prevent negative outcomes from occurring by making it possible to use the information to improve one's thoughts and behaviors (Choe et al., 2013).

Overall, the feedback literature across domains suggests that the amount of information provided within the shorter intervals of trial feedback presentation may simply be smaller and easier to synthesize in relation to one's preceding choices compared to larger amounts of feedback information occurring at less frequent intervals. It is also possible that feedback intervals that better allow for synthesizing will be more useful and effective on future tasks. However, it remains in debate at what point the trial feedback interval elicits responses more in line with those considered stemming from immediate feedback compared to delayed feedback. The current studies will implement several intermediary levels of trial feedback presentation (described in detail below) to help clarify this issue.

Description Versus Experience

In addition, the nature of the task itself may vary the frequency with which feedback is provided. Work comparing description-based versus experience-based decision tasks looks at differences in what and how much information is provided when making a decision and receiving its resulting outcome, which can be relevant to the potential influence of trial feedback frequency interval on evaluative and behavioral outcomes. Findings in this area of research suggest that when choosing between options in a decision making task, different choice patterns develop based on whether the scenario is more description-based (the outcome and probability of each choice option is explicitly provided and feedback is rarely received) or more experience-based (the outcome and probability of each choice option is learned through repeated feedback or sampling) in nature (e.g., Barron & Erev, 2003; Fox & Hadar, 2006; Hertwig, Barron, Weber, & Erev, 2004; Weber, Shafir, & Blais, 2004).

When people are only exposed to the descriptive information, they overweight the occurrence of rare events, leading to more risk-taking in the domain of gains and risk aversion in the domain of losses; however, when people are exposed to experiential feedback information, this changes their actions and they exhibit opposite decision making trends (Barron & Erev, 2003; Hertwig et al., 2004; Rakow & Newell, 2010).

As the distinction between description-based scenarios and experience-based scenarios incorporates the component of trial feedback frequency, this difference suggests that one of the reasons the data patterns vary is that the experience-based scenarios simply allow decision makers to better and more easily ascertain which options deliver the highest median payout – or the best outcomes the most often, thus increasing the likelihood to choose options that are advantageous and have potential outcomes in their favor (Yechiam & Busemeyer, 2006; Yechiam, Druyan, & Ert, 2008). Indeed, it has been noted that when payouts and decision feedback are more similar across scenarios, the difference in decision behavior diminishes, but does not disappear (e.g., Hau, Pleskac, Kiefer, & Hertwig, 2008; Ungemach, Chater, & Stewart, 2009).

Relating the standard Cups task used in the current studies to this distinction between types of scenarios, the Cups task presents all necessary description information to make informed choices on any given trial, but repeated trials and receiving immediate feedback for each trial throughout the task allows for the experiential information to influence the decision scenarios. The following sections will describe how the Cups task has been modified to better accommodate the distinction between description-based versus experience-based scenarios in regards to the frequency with which trial feedback is provided.

Varying Frequency Intervals

In addition to the noted choice pattern differences, the way that information is acquired in experience-based scenarios may potentially influence what type and amount of information is used during subsequent decision making due to specific outcomes seeming more certain and possible (Rakow & Newell, 2010), and receiving any amount of experiential feedback information can override previous information about decision choices and outcomes, contributing to these changes in decision making (Jessup, Bishara, & Busemeyer, 2008; Lejarraga & Gonzalez, 2011). Therefore, receiving decision feedback on a frequent basis, compared to not receiving this feedback, is an important factor in driving evaluative and behavioral outcomes in positive or beneficial ways.

Few studies have compared across more than one interval of outcome feedback presentation. Notably, Schmidt, Young, Swinnen, and Shapiro (1989) compared different intervals of trial feedback presentation and found that immediate feedback (received after every trial) elicited better initial performance, compared to a summary of feedback following a specified number of trials, but showed a decrease in performance during follow-up tasks. Countering this finding, Erev, Luria, & Erev (2006) stated that this result did not hold up when the task was more difficult and there was more than one response option to learn throughout in the task (as in the current case, a risky option compared to a riskless option). In these cases, shorter intervals of feedback presentation (which provide better distinction between the outcomes associated with the different responses) lead to better outcomes during initial and subsequent task performance.

Supportive of these ideas, a meta-analysis addressing the instructional effects of feedback on performance by Bangert-Drowns et al. (1991) found that compared to those who received no feedback about their answers, people who received feedback immediately following each of their answers were more likely to correct their answers and improve their performance. When including studies that had alternative feedback conditions (e.g., specific trial feedback presented at the end of a task) in comparison to no trial feedback, similar patterns of results emerged, with any presentation of trial feedback leading to more correct answers and improvement of performance than when this information was not presented. In addition, it was found that across different types of feedback, when feedback signaled not only the correctness or incorrectness of the answers, but also a context for providing a path of correction, the effect on performance improvement and number of correct answers increased.

Moreover, a meta-analytic review by Kulik and Kulik (1988) focusing on the timing of feedback found that receiving immediate feedback about one's choices and outcomes, in comparison to no feedback or delayed feedback conditions, was more effective in promoting improved performance and more correct responses; however, this only seemed to occur when tasks were designed to in a way that did not facilitate an "answer until correct" approach. Combining data across the two meta-analyses, this would suggest that while any amount of trial outcome feedback may be enough to guide thoughts and behaviors in beneficial ways to improve performance, it may also be the case that as the interval between receiving trial feedback increases, the beneficial influence decreases, perhaps due to difficulty in connecting responses to outcomes.

Evaluations of one's performance can also be influenced by how information is learned throughout the task and how frequently trial outcome feedback is provided. In addition to finding that more detailed feedback provided immediately after each choice led to higher achievement on subsequent tasks, Butler (1987) found that this type of feedback was associated with more positive evaluations of interest in the task and desire to improve on the task. Likewise, Vollmeyer & Rheinberg (2000; 2005) found that, in addition to improving task performance, receiving more frequent, instructive feedback led to more positive evaluations of the task and their performance; participants liked the task more, were more confident in task performance, were more confident in improving performance, and felt they knew how to do better on the task.

As will be discussed in more detail in following sections, trial feedback received at varying frequency intervals throughout completion of a task can influence evaluations of one's performance and actual task performance by changing how (or whether) the overall performance feedback is utilized. If participants are more easily able to synthesize the repeated, more immediate feedback across smaller frequency intervals, as is suggested by prior findings (e.g., Yechiam & Busemeyer, 2006; Yechiam, Druyan, & Ert, 2008), then they may be able to advantageously relate that knowledge to the overall task feedback and help streamline what is important to focus on for change and improved performance on subsequent trials and follow-up tasks (i.e., the overall task feedback may strengthen or diminish how interpretable and useful the trial feedback is viewed to direct future action; Hattie & Timperley, 2007). It is possible that any amount of trial feedback should have influence on interpreting and utilizing overall task feedback, but more frequent trial feedback may enhance the ease of integrating this information, so including intermediary frequency intervals will be important for establishing better classifications of immediate versus delayed feedback.

In addition, as more frequent trial feedback throughout a task may provide more instructive and easily integrated feedback information and change how the overall task performance is interpreted, this may make people feel like they can use it to help improve performance. Providing this type of feedback throughout decision tasks may result in more of these positive evaluations, which may potentially mediate the relationships between task feedback frames, trial feedback frequency, and their behavioral outcomes.

When completing tasks with no trial feedback, participants have access only to their overall task feedback. Without receiving the more frequent (and likely more useful and instructive) outcome feedback information throughout the task, they are left to evaluate each trial at a more individual level and may be less able to determine what is important to focus on for effective use of the task feedback in helping improve on subsequent trials and follow-up tasks (in other words, the overall task feedback is less interpretable and useful for directing future action; Hattie & Timperley, 2007).

Delineating feedback presentation intervals as more immediate or delayed throughout the current studies and what effect this has on retention of information for future performance, extends the knowledge of what frequency of trial feedback is necessary and effective to induce the outcomes (both evaluative and behavioral) associated with immediate trial feedback, compared to delayed trial feedback. By conducting studies in a lab-setting (as opposed to field studies and applied settings), some of the potential issues presented in previous research regarding differences in type of feedback across conditions, amount and timing of feedback provided, and relevance of feedback to the task can be controlled more methodically.

Current Studies – Manipulating Trial Feedback Frequency

An additional modification to the original format of the Cups task made in the current studies centers on how much and how often trial outcome feedback is provided to the participants, by creating versions of the task that vary the frequency interval of trial outcome feedback. Opposite of the standard provision of trial-by-trial outcome feedback, the individual trial feedback will be removed (no trial feedback condition); however, participants complete the same combinations of trials, choosing between the risky and riskless options. By removing the immediate, individual trial feedback, the overall task feedback is left as the only information regarding the outcomes of one's decisions (for more detail see *Materials* – Study 1A). The current set of studies also incorporates intermediary trial feedback frequency conditions that present a summary of the feedback resulting from each trial outcome following every three trials (representing the three repetitions of each trial type), every nine trials, and every twenty-seven trials (representing each gain/loss block; see Appendix - Figure A3 for examples). Under these conditions, participants receive all the same individual trial feedback information as those in the trial-by-trial feedback condition, but it is presented in blocks of tallied feedback information in varying frequency intervals rather than following each trial.

Participants in the no trial feedback condition also receive initial task instructions with the information that only one cup in the risky option is nonzero. This provides participants in these conditions a more exact probability of winning (losing) on any given trial. The amount of quarters available to win or lose is presented within the task trials. Participants in the other feedback interval conditions do not receive this outcome contingency information; thus, they are left to discover the exact probability of winning (losing) based on their assumptions and experience with choice outcomes.

As the Cups task has always been constructed using trial-by-trial feedback, creating versions that vary this frequency will allow for an avenue to start examining the potential effects of trial feedback frequency intervals on both performance evaluations and changes in task performance for this paradigm. More specifically, if receiving any amount of experiential information can lead people to discount provided descriptive information and weight the experiential information more heavily in order to guide their decisions (Jessup et al., 2008; Lejarraga & Gonzalez, 2011), supplying people with trial feedback may provide them with a richer set of feedback information useful for future task completion. Therefore, providing some amount of trial feedback, no matter what the frequency, will result in more positive performance evaluations and improved task performance on subsequent rounds of the Cups task. While this general designation may hold, it is also possible that more frequent trial feedback (such as trial-by-trial feedback) may exacerbate these effects, by further facilitating the ease of utilizing this information.

Given the distinctions between the different versions of the Cups task and the prior research on feedback frequency, decision makers may be able to see what options offer the best outcomes the most often when receiving more frequent trial feedback, such as after every trial choice, as they get to directly experience repeated and more immediate results of what choices work and what choices do not. In prior studies, being able to see "what works best" can directly influence decision tendencies by increasing risk taking for those risks that have an actual higher likelihood of earning a payout (Lejarraga & Gonzalez, 2011; Yechiam & Busemeyer, 2006; Yechiam, Druyan, & Ert, 2008). The Cups task used for the current studies will have a similar nature when trial feedback is provided; thus, reliably acting on the choices that seem to elicit the best returns will indicate a better ability to discriminate between risk-advantageous and risk-disadvantageous choices because participants will be choosing options that are in their favor and have a higher expected value outcome (i.e., more normatively correct) while avoiding those that are not and do not.

As preliminary evidence of this, unpublished work from my laboratory suggests that task performance does improve across multiple trials throughout a single round of the Cups task. This dataset was collected using the standard Cups task, and results found that risk sensitivity scores improved across these three trials. In particular, increases in risk sensitivity scores were marked by participants taking less risk from the first trial to the third trial in the risk-disadvantageous conditions (Levin & Bossard, 2011). Risk-disadvantageous trials represent risks that are more likely to lead to negative outcomes; thus, these trials will necessarily be less likely to have positive outcomes. In addition, increases were also seen in risk taking from the first trial to the third trial in the risk-advantageous conditions (representing risks that are more likely to lead to positive outcomes; Levin & Bossard, 2011), but as adult populations tend to exhibit higher levels of risk taking on these trial types to begin with, the amount of increase is much smaller.

Results such as these indicate that as participants are completing the Cups task, they gain more information about the "correctness" of their choices through repeated and immediate feedback. As risk taking decreases for those risks that are less advantageous and increases for those risks that are more advantageous, it appears this additional information allows people to work out that certain conditions, like risk-disadvantageous trials, do not reward positive outcomes very often, but other conditions, like risk-advantageous trials, do. Accordingly, participants modify their choices to focus on taking risks in trials that offer the best outcomes the most often, which indicates a shift towards more normatively correct decision making. While this preliminary dataset does not provide a full picture of what is expected to be shown in the current set of studies, it does provide a starting point for generating predictions.

These findings suggest that if improvement in task performance and risk sensitivity can be seen and that people seem to be applying their experience across a single round of the Cups task, then similar patterns may exist extending people's ability to directly apply their experience in round one to show improvement and application of experience in consecutive rounds of the Cups task. The preliminary dataset does not include responses from versions of the Cups task that involve alternative trial feedback frequency or no trial feedback at all, so it is harder to say whether any improvement could potentially be found in these conditions; however, prior research on how feedback is interpreted and utilized (as mentioned in more detail below; Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2000; 2005) suggest that conditions that do not receive trial feedback will lead to less improvement in task performance than those that do.

Influence of Trial Feedback Relevance

The influence feedback holds on evaluative and behavioral outcomes should be regarded under the context of the relevance of that feedback. In other words, the influence of feedback may be determined by how much detail is provided to make connections between one's response and its outcome, the relative correctness of one's response, and whether one should relate feedback regarding correct retrieval or application of information to future decisions. Differences in relevance of the decision feedback to one's choices may influence how one applies feedback to and improvement in subsequent decision making through more interpretable and directive information about one's choices and performance. When investigating the effectiveness of feedback on evaluations and behaviors, feedback relevance has often coincided with feedback frequency. Indeed, the general consensus in the research on the effects of feedback is that the more often feedback is received, the more instructive it is for guiding thoughts and decisions; thus, when examining decision choices that can elicit feedback based on decision outcomes and performance, this feedback can certainly be corrective and guide one to better decisions (e.g., knowing what to choose and what to avoid on future decisions).

Following from this, much of the arguments made above to support investigating differences in trial feedback frequency are applicable to support investigation of differences in feedback relevance, as well. For example, knowing the outcome of one's decision quickly after making it can make the best options become clearer to the decision maker; thus this information becomes relevant feedback for the decision maker to use going forward for future decisions (e.g., Yechiam & Busemeyer, 2006; Yechiam, Druyan, & Ert, 2008). Nevertheless, not all trial feedback is directly relevant in providing guidance for how to proceed on subsequent tasks. Sometimes the trial feedback is too generic or only provides information about the outcome, but does not indicate the course of action needed to change one's decisions to improve those outcomes. This can lead to viewing the feedback as less useful and uncertainty about how to respond (Fedor, 1991; Kluger & DeNisi, 1998). Therefore, knowing the frequency intervals with which to provide trial outcome feedback is important, but the relevance of that feedback, provided at any interval, is also something to consider when determining the most effective forms of feedback for task performance outcomes.

As with the effects of trial feedback frequency, previous research has suggested that more detailed and individualized feedback (indicating more relevant feedback) can lead to more positive performance evaluations and increased task performance. Butler (1987) found that feedback on a creative use task containing more detailed information (comments on what was done correctly and incorrectly by that individual) compared to an overall performance score (just the amount correct) led to more positive evaluations of interest and desire to improve on the task and a higher amount of achievement on subsequent tasks. Likewise, Vollmeyer and Rheinberg (2005) found that receiving more detailed and instructive feedback information led to more positive evaluations of the task (i.e. participants were more confident in task performance, felt they knew how to do better on the task), and an improvement in performance on a problem solving task.

While trial feedback has been shown to have positive benefits for evaluative and behavioral outcomes, it appears that it is not sufficient to simply tell a person whether he/she made correct or incorrect responses. The feedback also needs to convey the actual correct answer in some way (Bangert-Drowns et al., 1991; Pashler, Cepeda, Wixted, & Rohrer, 2005). When the correct answer is made available after an error, people are able to better integrate that information into memory and answer more questions correctly on a follow-up test (Anderson, Kulhavy, & Andre, 1971).

Indeed, other work addressing the importance of feedback relevance suggests that simple correct/incorrect feedback transmits some information for directing improved task performance, but is not as useful as being provided with information regarding the right answer or the context in which the right answer exists (Fazio, Huelser, Johnson, &

Marsh, 2010). Providing feedback throughout a task, however, can reduce redundant cognitive processes and supply people with schemas to help correct their errors, resulting in improved performance compared to receiving no feedback or generic correct/incorrect feedback (e.g., Leutner, 1993; Moreno & Mayer, 2004). As will be addressed in the later studies, trial outcome feedback in the Cups task does not explicitly provide the correct answer, so measures were taken to make the correct answer more clear and to make the trial outcome feedback more relevant by providing information about one's choice and the decision scenario in which it was deemed correct or incorrect in addition to the trial outcome. Doing so should provide more guidance for which options to approach (because they are more advantageous) and which to avoid (because they are more disadvantageous) in order to achieve better, more normatively correct outcomes, feel better about one's performance, and improve task performance.

Looking to the health psychology literature, relevant feedback is a key component of constructing tailored health messages. For health messages to act as an intervention and influence one's subsequent attitudes and behaviors, they need to be relevant, personalized, and able to be acted upon (Contento, Black, Bronner, Lytle, Malony, Olson, & Swadener, 1995). Consequently, a multitude of research has investigated the differential influence of more tailored versus more generic health messages on a variety of outcomes (e.g., seen as more relevant and engaging – DiClemente, Marinilli, Singh, & Bellino, 2001; seen as more effective – Oenema & Brug, 2003; improve different health behaviors – Campbell, DeVellis, Strecher, Ammerman, Devellis, & Sandler, 1994).

While many health messages aim to provide feedback to increase engagement in the message and inspire change, tailored feedback offers a better connection between one's choices or behaviors and specific outcomes to be more instructive, corrective, and relevant for improving one's outcomes. Because tailored feedback promotes feelings of relevance, this may make the information more salient and instructive, particularly when in relation to one's previous performance or a normative standard, leading to more positive outcomes (Brug et al., 1998).

Although health messages often relate to risk-taking behaviors, as represented in the current studies, there appears to be little evidence to suggest how this more detailed feedback relates to development of skills regarding these risk-taking behaviors, especially for those with higher or lower skills to begin with (Brug et al., 1998). Therefore, the opportunity exists to clarify exactly how more direct, corrective information (like tailored messages) may affect subsequent risky decision making and what conditions may be necessary to produce the most normatively correct outcomes.

Additionally, more elaborate feedback is not always seen as more effective compared to more simple, less detailed feedback. In some cases, the additional information provided merely adds distraction from what one needs to know for improving task performance and does not enhance the effects of feedback on performance (Butler, Godbole, & Marsh, 2013), but this is not necessarily the case for the current studies. In the Cups task, the choice outcomes are random and one can win or lose on any given trial, so simply receiving feedback about those outcomes may not be enough to increase comprehension of the information. Elaborating on the standard trial outcome feedback to include the context in which it occurs (i.e., making a direct connection between one's choice and whether the probability of winning/losing a trial is favorable, indicating the normatively correct choices) can provide decision makers with more helpful, rather than superfluous, information regarding which choices were better or worse and why that may be in order to guide their future actions.

Despite some lack of consensus regarding how more relevant feedback can affect evaluative and behavioral outcomes, similar to trial feedback frequency intervals, the amount of detail provided by more relevant feedback may change how (or whether) the overall performance feedback is utilized. The ability to synthesize relevant trial feedback across smaller frequency intervals may lead to better relations between trial outcome feedback knowledge and the overall task feedback and better focus on what is needed for improving performance (Hattie & Timperley, 2007; Yechiam & Busemeyer, 2006; Yechiam, Druyan, & Ert, 2008).

As with the arguments for the effects of trial feedback frequency intervals, this suggests that increasing the relevance of the trial feedback should serve to provide more detailed and instructive information and change how the overall task performance is incorporated, strengthening or diminishing one's view of how usable the trial feedback is to direct future action. Therefore, the effects of task feedback frames and trial feedback frequency on more positive performance evaluations and improvement in subsequent task performance (and the likelihood of finding these effects), including the potential mediational relationship between task feedback frames, trial feedback frequency, and their behavioral outcomes by performance evaluations, should only be strengthened by increasing the relevance of trial feedback in the Cups task in comparison to the standard trial outcome feedback.

Current Studies – Manipulating Trial Feedback Relevance

The final modification to the Cups task in the current studies involves what trial outcome feedback information is provided to participants. Trial outcome feedback was altered to include the context under which one's choices were made and whether the trial was associated with advantageous or disadvantageous risk taking. Across all task feedback frame and trial feedback frequency conditions, participants received more detailed trial feedback set to enhance one's ability to connect the decision scenario, his/her choice, and the associated decision outcome (e.g., "On Trial 1, there were 3 cups and the chance to win 5 quarters, and this represented a risk advantageous trial. You chose the risky side and won 5 quarters"; "On Trial 3, there were 3 cups and the chance to lose 5 quarters, and this represented a risk disadvantageous trial. You chose the risky side and lost 0 quarters") to indicate whether the choice was normatively correct or not and provide direction for future choices. Regardless of any differences in the trial feedback provided, the participants completed the same combinations of trials, choosing between the risky and riskless options.

As the standard trial feedback is presented as outcome information (i.e., “You won XX quarters”; “You lost XX quarters”), adding relevant context information to the trial feedback allows for examination of the potential effects of trial feedback relevance on both performance evaluations and changes in task performance for this paradigm and whether the trial feedback normally received in the Cups task is relevant enough to guide future decisions. Given the preliminary findings regarding task performance improvement in the Cups task (see *Current Studies – Manipulating Trial Feedback Frequency*), the standard trial feedback in the Cups task may be sufficient for providing this guidance; however, this trial feedback is somewhat generic and not explicitly tied to the choices one previously made (particularly when the trial feedback is less frequent).

If, as stated previously, supplying trial feedback can provide a richer set of feedback information useful for future task completion, decision makers may be better able to see what options offer the best outcomes the most often, choosing options that are in their favor and have a higher expected value outcome while avoiding those that are not and do not (i.e., making more normatively correct decisions). Thus, trial feedback that is more relevant to one’s choices by making it more detailed, specific to the decision context, and connected to one’s responses may augment any noted effects that trial feedback produces (and the likelihood of finding those effects) on more positive performance evaluations and improved task performance in subsequent decision making tasks.

Relationship Between Feedback Variables

Despite a multitude of research done on feedback effects across different domains, as noted earlier, little work has addressed the lack of consistency in designating different types and frequencies of feedback and how these different aspects of feedback may combine to create the most effective scenarios. Different aspects of feedback representing the context of framing, frequency, and relevance exist widely in the real-world (e.g., consumer, education, and health decisions), so understanding how these factors relate and their potential impact on evaluations, behavioral intentions, and behaviors can have large and important implications for how people can be influenced to make different, and potentially better, decisions. In particular, the distinction between the different trial feedback frequencies is important because it may help determine the influence of positively and negatively framed task feedback.

When considering both the variables task feedback framing and trial feedback frequency in one decision making task, it is possible that combinations of the two may lead to different degrees of performance evaluations and subsequent task improvement across multiple rounds of the task. Recall that choices are often dependent on both how decision options are presented and how the decision maker interprets that presentation (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981), so evaluative (performance evaluations) and behavioral outcomes (task performance) may vary based both on the positive or negative frame of the task feedback and how the presentation of the trial feedback allows for interpreting task performance and ability to improve. Converging evidence suggests that a basic distinction between responses to tasks and scenarios presented in a more positive (gain) frame and a more negative (loss) frame exists for most people (e.g., risk aversion - Kahneman & Tversky, 1979; Levin et al., 2007b; ambiguity - Lauriola, Levin, & Hart, 2007; brain activation - Xue, Lu, Levin, & Bechara, 2011); however, the additional distinction between different trial feedback frequencies may act as an influencing factor as to the interpretation and lasting impact of positively and negatively framed task feedback.

Specifically, an interaction may occur between task feedback frame and trial feedback frequency; the more positive performance evaluations and improved task performance on consecutive rounds of the task that result both from receiving trial feedback and from the positive task feedback frame should combine to produce the most positive performance evaluations and the largest improvement in overall task performance. As negative task feedback frames and conditions that receive no trial feedback will likely produce less positive performance evaluations and less improvement in task performance in subsequent rounds of the Cups task, less (or the least) positive performance evaluations and smaller (or no) improvements in task performance would be expected for other combinations. In essence, measures of performance evaluations and improved task performance will show larger differences between those conditions that receive any frequency interval of trial feedback and those conditions that do not receive trial feedback frequency when the task feedback frame is positive than when it is negative (i.e., a Trial Feedback Frequency X Task Feedback Frame interaction).

Evaluations as Mediator

As alluded to previously, whether or not one receives trial feedback throughout completion of a task can influence performance evaluations and task performance by changing how (and whether) the overall task feedback is utilized. In a review of how researchers have previously categorized feedback, Hattie and Timperley (2007) summarized that feedback comes in four main varieties that differ in level of information provided and interpretability gained. The two most common forms of feedback are task feedback (feedback that exhibits how well a task was completed) and self-regulative feedback (feedback that allows for monitoring, directing, and correcting actions). In the context of the current studies, the overall task performance feedback (framed positively or negatively) would be categorized as task feedback, since it focuses on a straightforward and overall score, but adding concurrent trial feedback at various frequency intervals can change one's perception of the task performance feedback.

When trial feedback is provided more frequently throughout a task and in closer proximity to one's response, it is perceived as more instructive and allows for the feedback to be interpreted as more self-regulative. As self-regulative feedback is considered the most effective in producing positive outcomes, the ability to recognize what past decisions were good or bad and how to adjust accordingly leads to feeling confident with one's performance, liking the task more, and feeling like improvement is possible in the future (Butler, 1987; Hattie & Timperley, 2007), in addition to improving actual task performance. Therefore, receiving trial feedback, in general, may provide the information needed to feel positive about one's performance and improve task performance, but also act as a lens through which to view overall task feedback; positive versus negative frames may strengthen or diminish these positive feelings and intentions to improve.

Putting feedback, evaluations, and task performance together, Vollmeyer & Rheinberg (2000; 2005) found a mediational relationship among these variables such that receiving more detailed, easily integrated, and instructive feedback led to better evaluations of the task and task performance – including feeling able to use the feedback to improve on the task – and these evaluations led to improvement in actual performance on subsequent trials of the same task.

Applying this relationship to the current studies, a mediated moderation pattern may occur. In this case, the Trial Feedback Frequency X Task Feedback Frame interaction effect on improved task performance will occur; however, the Trial Feedback Frequency X Task Feedback Frame interaction effect on subsequent task performance may be mediated by performance evaluations. Presuming initial effects of both task feedback frame and trial feedback frequency to produce more positive or negative performance evaluations following initial completion of the Cups task (the most positive evaluations being shown in the positive frame/trial-by-trial feedback condition), participants who receive positively framed task feedback and more frequent trial outcome feedback will incorporate these performance evaluations into their understanding and ability to improve task performance on a subsequent round. Participants who do not receive trial feedback and thus, a lack of instructive and easily integrated feedback, will not utilize their performance evaluations in this way.

In sum, a mediated moderation pattern will exist if the anticipated Trial Feedback Frequency X Task Feedback Frame interaction on subsequent task performance is produced by one's performance evaluations of round one, and when these evaluations are controlled, the direct effect of the Trial Feedback Frequency X Task Feedback Frame interaction is reduced. This will indicate that differences in subsequent task performance are the result of differences in evaluations of round one performance, such that when people rate their performance more positively as a result of both receiving positively framed task feedback and more frequent, individual trial feedback, people use these evaluations to direct their approach to subsequent decision choices.

It is also important to keep in mind that while feedback frame and frequency together can influence better decision making, the relevance of the feedback matters as well. If the task feedback is framed positively and trial feedback follows in close proximity to one's choices, but one cannot extract the information necessary to apply to future decisions, then the feedback may not make much impact. Providing relevant feedback in a positive manner and in close proximity to one's choices will serve to augment the predicted relationship of task feedback frame and trial feedback frequency regarding what is important for improving one's decision making, including the mediated moderation pattern – due in part to more associated positive evaluations of the task and one's task performance.

Overview of Current Studies

Four studies are designed to examine the relationship between the roles of task feedback presentation (attribute framing), frequency of trial feedback, relevance of trial feedback, and their outcomes. While all three aspects of feedback have received attention in previous research, the exact influence of both feedback information gained throughout a decision making task and feedback about one's overall performance on evaluative and behavioral outcomes – in particular, taking action to change one's performance and make better decisions – is less clear. Particular questions of interest are when completing multiple rounds of the same task, what conditions provide for more positive evaluations of one's performance and lead to more normatively correct decision making (by way of improved task performance).

The same general procedure will be used in all four studies (see Appendix - Table A1 for a breakdown of how the different feedback variables are incorporated across studies). Upon arrival, participants were told that they would be participating in a study to understand how performance feedback influences later performance outcomes. Participants completed a version of the Cups task, with the frequency interval of trial feedback varied between participants, and received either positively or negatively framed feedback regarding their overall task performance. After receiving their task feedback, participants answered a series of questions about their performance and the task itself. Upon completing these questions, participants again completed the Cups task (under the same conditions as before) and answered a second set of questions about their performance and the task.

Studies 1A and 1B establish a paradigm to examine the influence of task feedback frame and trial feedback frequency on decision choices, evaluations, and performance of a risky decision making task. Positively or negatively framed overall task feedback will follow completion of a risky decision making task in which participants receive or do not receive trial outcome feedback after each individual choice. Across the two studies, different feedback framing manipulations were used, but the general predictions are the same.

Some research has suggested that positively framed task feedback should result in more positive performance evaluations and improved task performance on subsequent rounds of the Cups task, compared to negatively framed task feedback (although see Kluger & DeNisis, 1996 and Van Dijk & Kluger, 2011 for times when this is not the case). Given the design of the Cups task, this pattern of results is expected to hold; therefore, it is predicted that framing the presentation of overall task feedback in the current studies positively (indicating achieving normatively correct responses), compared to negatively (indicating failing to achieve normatively correct responses), will lead to more positive performance evaluations and improved performance (i.e., increased risk sensitivity) in consecutive rounds of a risky decision making task.

In addition, repeatedly experiencing successful versus unsuccessful outcomes across different Cups task trials should give participants an idea of which choices are most advantageous. Acting in a way to optimize decision making by maximizing one's advantage and favorable outcomes would be more easily accomplished when trial feedback is provided; thus, it is predicted that participants will show more positive performance evaluations and improved performance (i.e., increased risk sensitivity) in consecutive rounds of a risky decision making task when trial outcome feedback is provided (i.e., trial-by-trial condition) than when it is not (i.e., no trial feedback condition).

Beyond these main effects, it is predicted that the effect of trial feedback frequency on performance evaluations (following round one of the Cups task) and improved task performance (across rounds of the Cups task) will be moderated by whether overall task feedback is framed positively or negatively. The difference in performance evaluations and in task performance improvement (indicating an increase in normatively correct decision making) between conditions that receive trial feedback and those that do not will be larger when the overall task feedback is framed positively versus negatively, with the most positive evaluations and the largest improvement in performance in consecutive rounds of the task seen under the condition of receiving trial feedback combined with positive task feedback frames.

In addition, a mediated moderation pattern may occur in both Studies 1A and 1B. The anticipated influence of trial feedback frequency combined with task feedback frames on improved task performance (i.e., the Trial Feedback Frequency X Task Feedback Frame interaction) may be produced by one's performance evaluations following the initial round of the Cups task, and when these evaluations are controlled, the direct effect of the interaction will be reduced. While both task feedback frames and trial feedback frequency may influence performance evaluations to be more positive or negative, when people rate their performance more positively as a result of receiving both trial feedback (providing more instructive feedback and ease of integrating feedback information) and positively framed task feedback, people use these evaluations to guide their subsequent task performance.

Studies 2A and 2B extend understanding of the role trial feedback frequency (and the related concept of trial feedback relevance) can play in connection with evaluative and behavioral outcomes in a risky decision making task. Using a similar design as Study 1B, Study 2A also includes the addition of multiple intermediary trial feedback conditions (presentation of trial feedback following every three, nine, or twenty-seven trials).

Adding these conditions will allow for a better test of what trial feedback frequency intervals provide enough corrective, instructive information to elicit more positive evaluations and improved task performance on subsequent rounds of the Cups task. With the additional trial feedback frequency intervals, the effect that more immediate, compared to delayed, trial feedback should have on evaluative and behavioral performance outcomes can be clarified, particularly in how different types of feedback that relate to what choices had favorable outcomes versus unfavorable outcomes can heighten one's normatively correct decision tendencies to improve task performance.

Following from Study 1B, it is expected that the relative data patterns of the main effects of task feedback frame and trial feedback frequency will remain intact, such that both positively framed task feedback and the trial-by-trial feedback condition will show more positive performance evaluations and improved task performance across rounds of the Cups task. In addition, in light of evidence from the health and educational psychology literature that positive outcomes resulting from immediate feedback stem from the instructive and integrative nature of the feedback (Brug et al., 1998; Dihoff et al., 2003; Epstein et al. 2002) and from work on the description versus experience decision scenarios suggesting that any amount of experiential outcome information will lead participants to discount known decision contingency information (Jessup, Bishara, & Busemeyer, 2008; Lejarraga & Gonzalez, 2011), it is expected that the intermediary trial feedback conditions will not be significantly different from the trial-by-trial feedback condition, but will be significantly different from the no trial feedback condition. Thus, the interaction of trial feedback frequency and task feedback frame across task rounds (described above) will show similar patterns, but the difference between conditions receiving trial feedback at any interval than the no trial feedback condition will be significantly larger when overall task feedback is framed positively versus negatively.

While any amount of trial feedback in combination with positive task feedback frame may result in more positive evaluations and improved performance on subsequent rounds of the task, it is possible that the most positive evaluations and the largest improvement in performance will be seen under the positive frame/trial-by-trial feedback condition, as this represents the condition in which it is potentially most easy to recognize favorable and unfavorable outcomes and adjust responses accordingly. It is also predicted that the mediated moderation pattern associated with the Trial Feedback Frequency X Task Feedback Frame interaction will hold.

Study 2B addresses how trial feedback relevance influences the relationship between task feedback frame, trial feedback frequency, and their evaluative and behavioral outcomes. While maintaining the same study design as Study 2A, the trial feedback relevance was modified to incorporate not just outcome information but also the context within which the decisions were made (i.e., designation of trial as favorable/unfavorable; whether the risky or riskless option was chosen). This information serves to provide participants with a better understanding of how their choice related to the outcome they received and whether that is a course they should continue to take on subsequent decisions.

While it is predicted that the relative data patterns, including the predicted mediated moderation pattern associated with the Trial Feedback Frequency X Task Feedback Frame interaction, stated in Study 2A will be more likely to be found, it is also possible that any noted effects of task feedback frame and trial feedback frequency will be augmented due to the feedback information seeming more salient and instructive to use in improving performance (Brug et al., 1998). Additionally, as more relevant information may help facilitate synthesizing trial feedback, smaller frequency intervals may lead to better incorporation of feedback knowledge with the overall task feedback and better focus on what is needed for improving performance (Hattie & Timperley, 2007; Yechiam & Busemeyer, 2006; Yechiam, Druyan, & Ert, 2008). Therefore, it is a possibility that the positive task feedback frame/trial-by-trial feedback condition will be augmented the most by increasing the relevance of trial feedback.

To summarize, the proposed studies will address the following hypotheses and research questions:

1. Framing overall task feedback may lead to a valence-consistent shift.
    a. When task feedback is framed positively, task performance will be more positively evaluated and subsequent risk sensitivity scores will improve; however, when task feedback is framed negatively, task performance will be more negatively evaluated and subsequent risk sensitivity scores will show less (or no) improvement.

2. Trial feedback frequency will lead to differences in performance evaluations and subsequent task performance.
    a. When trial feedback is provided, task performance will be more positively evaluated and subsequent risk sensitivity scores will show more improvement than when trial feedback is not provided.
    b. Alternative: When trial feedback is provided more frequently (e.g., trial-by-trial feedback), task performance may be more positively evaluated and subsequent risk sensitivity scores may show more improvement than when trial feedback is provided less frequently or not at all.
3. Task feedback frame and trial feedback frequency manipulations will interact to show different levels of positive performance evaluations and change (improvement) in subsequent task performance.
    a. The difference in performance evaluations and risk sensitivity improvement between trial feedback frequency conditions that receive trial feedback versus conditions that receive no trial feedback will be larger when task feedback frames are positive versus negative.
        i. Because both positively framed task feedback and receiving trial feedback can lead to more positive performance evaluations and improvement in subsequent risk sensitivity scores, the most positive evaluations and the largest improvement seen under this condition.
        ii. Because both negatively framed task feedback and not receiving trial feedback can lead to less positive performance evaluations and less improvement in subsequent risk sensitivity scores, the least positive evaluations and the least amount of improvement will be seen under this condition.

   b. The influence of trial feedback frequency and task feedback frame on subsequent task performance will be mediated by evaluations of one's initial performance.
      i. Participants whose more positive performance evaluations result from both receiving trial feedback (providing more instructive feedback and ease of integrating feedback information) and positively framed task feedback will use these evaluations to guide and improve subsequent risk sensitivity scores.
4. How does increasing the relevance of trial feedback information (by establishing a more direct connection between one's choices and their outcomes and providing information regarding the relative advantage of one's choices) affect performance evaluations and subsequent task performance?
   a. Relative data patterns and relationships between task feedback frame and trial feedback frequency will remain, but more relevant trial feedback will result in augmented effects.

CHAPTER I

STUDY 1A

Over the past few decades, research on the effects of feedback on evaluations and performance outcomes has shown that feedback frame and frequency are both important for understanding how people feel about and adjust their actions. Pertinent to the current studies, the research has expressed interest in what choices are made under various circumstances. Under circumstances in which trial-by-trial feedback is combined with positive feedback framing of task performance, the perception of the trial feedback and how useful it is deemed for future tasks may change; therefore, it is particularly important to look at how task feedback frames and trial feedback frequency together can help elucidate potential outcomes on decision choices, evaluations, and task performance. Critically, this can further the understanding of what factors can encourage people to make better, more normatively correct decisions in a variety of scenarios where people perform repeated behaviors.

The primary goal of Study 1A was to develop and test an experimental paradigm that could be used to understand the relationship between task feedback frame and trial feedback frequency and their influence on evaluative and behavioral outcomes. Any time changes are made to a task (in this case, the framing of the task feedback and the interval of providing trial feedback), not all elements of that task may be maintained. Therefore, analyses using only round one responses were run to address the ability of the new Cups task versions to replicate prior Cups task data patterns.

Following these initial analyses, Study 1A addresses three main research questions. First, Study 1A addresses whether positively versus negatively framed overall task feedback will result in valence-consistent response shifts. It is expected that the positive task feedback frame condition will exhibit more positive evaluations of task performance (e.g., interest in continuing, felt control over task) and more improved task performance (as measured by larger, positive differences in risk sensitivity scores across consecutive rounds of the Cups task), while the negative task feedback frame condition will exhibit more negative evaluations of task performance (e.g., less interest in continuing, less control over task) and less improved task performance (as measured by negative or no differences in risk sensitivity scores across consecutive rounds of the Cups task).

Second, Study 1A addresses whether trial feedback frequency (receiving or not receiving immediate trial feedback following each decision) contributes to more positive performance evaluations and more improvement in subsequent task performance. It is expected that receiving outcome feedback following each trial in the Cups task lead to more positive evaluations of task performance and more improved task performance, while receiving no trial feedback in the Cups task will show less positive performance evaluations and less improved task performance.

Finally, Study 1A explores whether framing overall task feedback as positive or negative will lead to differences in performance evaluations and improved task performance across varying frequency of trial feedback. Specifically, the more positive performance evaluations and a larger improvement in subsequent task performance predicted when receiving trial feedback compared to receiving no trial feedback is expected to be moderated by the task feedback frame; the difference in performance evaluations and improved task performance between trial feedback frequency conditions will be larger when task feedback is framed positively versus negatively. The most positive performance evaluations and the most improved task performance is expected to be seen under the condition of positive task feedback frames combined with receiving feedback after each trial.

Further, a mediated moderation pattern is predicted whereby the anticipated Trial Feedback Frequency X Task Feedback Frame interaction effect on subsequent task performance may be mediated by performance evaluations. While both task feedback frames and frequency of trial feedback may influence performance evaluations to be more positive or negative following initial completion of the Cups task, when people rate their performance more positively as a result of both receiving positively framed task feedback and feedback after every trial (providing more instructive feedback and ease of integrating feedback information), people use these evaluations to guide their understanding and improve task performance on subsequent trials.

Method

Participants and Design

Two hundred and thirty-six undergraduate students (*M* = 18.99 years, *SD* = 1.88) were recruited from the University of Iowa Psychology Department Subject Pool as partial fulfillment of a research requirement. Of the recruited participants, 68.2% were males and 66.5% were Caucasian. Six participants were dropped from analysis due to failure to recall task performance or failure to pass suspicion checks (note: no changes in data patterns or significance levels occurred whether these participants were included or excluded from the dataset). The study used a 2 (task feedback frame: positive vs. negative) X 2 (trial feedback frequency: no trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) design with all factors manipulated between-subjects.

Materials and Procedure

Procedure overview. All instructions, tasks, and questions were delivered on a personal computer, and participation lasted approximately 30 minutes. Upon arrival, participants consented to participate in the study, during which they completed two rounds of the Cups task (Levin et al., 2007b), which involves winning and losing hypothetical money, and a series of questions regarding task performance and evaluations. Participants were told that the purpose of the study was to understand how performance feedback influences later performance outcomes.

During the experimental session, participants first completed the initial round of the Cups task, being told to make the best decisions they can. Counterbalancing was employed so participants received trial-by-trial feedback regarding their choices or no trial feedback throughout the task. Immediately after finishing all trials of the first round, the total amount earned or not earned, compared to the maximum possible, appeared on the screen, counterbalanced within each trial feedback level so that participants received overall task feedback framed in either a positive or a negative manner.

After receiving their overall performance feedback on round one detailing the amount of money earned or not throughout the entire task, participants answered a series of dependent measures (detailed below) before completing the Cups task for a second time. The participant's second round of the Cups task presented the same trial feedback frequency and feedback framing conditions as for the first round.

Following the second round of the Cups task, participants again filled out a series of dependent measures and additionally, the individual difference measure of Numeracy (Weller, Dieckmann, Tusler, Mertz, Burns, & Peters, 2013) and demographics. Finally, participants were thoroughly debriefed and thanked.

Cups task. Following the modified, computerized version as set forth by Levin and colleagues (2007b), the Cups task is used to examine individuals' decision trends under uncertainty. In this task, participants make choices regarding winning and losing hypothetical monetary outcomes (0 to 5 quarters). Each within-subject trial consists of making a choice between a riskless and a risky set of cups and are constructed by manipulating whether quarters will be won or lost, the probability of winning or losing if the risky option is chosen (.20, .33, or .50, representing respectively, choosing from an array of 5, 3, or 2 cups), and the amount one can potentially win or lose by choosing the risky option (2, 3, or 5 quarters). Thus, gain (loss) trials involve the choice between an option that offers a sure gain (loss) of one quarter and an option that offers a probability of winning (losing) multiple quarters (see Appendix - Figure A1 for a depiction of the different trials).

In every condition, participants were presented with a set of two, three, or five cups shown on each side of the screen for each trial. One side of the screen was the riskless side; it was stated that if a cup was selected on this side, one quarter was gained or lost depending on the domain. The other side of the screen was the risky side; it was stated that if a cup was selected on this side, participants either gained or lost the designated number of quarters (2, 3, or 5) or no quarters. The outcomes occurring when the risky side was chosen were structured such that the distribution of the three successive trial outcomes matched what should occur probabilistically given the number of cups in those trials. For each trial type, a random list of outcomes was created equal to the probability of interest (.50, .33, .20). When there were two cups, there were four outcomes with two nonzero outcomes. When there were three cups, there were three outcomes with one nonzero outcome. When there were five cups, there were five outcomes with one nonzero outcome. The outcomes for the three successive trials-occurring when the risky side was chosen were drawn from the first three numbers in these lists. This means that trials having two cups on each side had at least one nonzero outcome, trials having three cups on each side had one nonzero outcome, and trials having five cups on each side may or may not have had a nonzero outcome. Additionally, the side of the screen that the riskless and risky options appear was randomized.

To accommodate the manipulation of trial feedback frequency within this task design, participants in the no trial feedback conditions received task instructions prior to starting the task with the information that only one cup in the risky option was nonzero. For example, for gain trials, the no trial feedback condition task instructions stated that "Choosing from one side will result in a 100% chance of winning money ($.25). Under one cup on the other side, you have a chance to win a larger amount of money. All other cups on this side result in winning $0." This provided participants in the no trial feedback conditions the exact probability of winning (losing) on any given trial prior to completing the task. Participants in the trial-by-trial feedback conditions did not receive this exact probability information. The trial-by-trial feedback condition task instructions for gain trials stated that "Choosing from one side will result in a 100% chance of winning money ($.25). Choosing from the other side, you have a chance to win a larger amount of money." Thus, these participants were left to discover the probability of winning (losing) based on their assumptions and experience with choice outcomes. Under both conditions, the amount of quarters available to win or lose was presented within the task trials. Comparable instructions were provided for loss trials.

In addition, the awarding of quarters gained or lost varied between the two conditions (see Appendix - Figure A3 - parts a and b). For the trial-by-trial feedback condition of the Cups task, the bottom of the computer screen depicted a bank where coins were gained or lost from the total after every trial choice. This provided participants with immediate feedback regarding specific trial gains and losses. In contrast, for the no trial feedback condition, the bottom of the computer screen again depicted a bank of coins, but no coins were exhibited as being gained or lost on-screen for each trial choice. Therefore, in the no trial feedback condition, successive choices were made with no immediate feedback regarding previous trial gains or losses.

To manipulate the other main factor of task feedback framing, at the end of the task, all participants were shown a monetary total reflecting their overall performance on the task. To make the task feedback framing more objectively identical in its provided information, all participants, regardless of experimental condition, received the same false feedback for their total score after the first round. This total was presented in one of two ways: positively framed or negatively framed. For the positive frame, monetary totals represented how much money one (falsely) earned throughout the task, with higher totals equaling better performance ("You have earned 98 quarters out of a possible 180 for a total of $24.50."). For the negative frame, monetary totals represented how much money one (falsely) failed to earn throughout the task, with higher totals equaling lower performance ("You have failed to earn 82 quarters out of a possible 180 for a total of $24.50."). Task feedback on round two of the Cups task was also presented to match the participant's framing condition, but reflected their actual performance scores.

As noted above, participants in both versions of the Cups task received a set of each combination of probability levels (2/3/5 cups) and outcome levels (2/3/5 quarters) presented in randomized order under both gains and losses, totaling 54 trials. Gain and loss trials were constructed into separate trial blocks, with order of trial blocks counterbalanced across participants.

By manipulating gain/loss domain, probability of winning/losing, and magnitude of winning/losing for the risky option, three decision contexts are created under both gains and losses: risk advantageous (RA), equal expected value (EQEV), and risk disadvantageous (RD; see Figure A1 for a depiction of the different trials). RA trials exist when choosing the risky option is more beneficial (has a higher expected value outcome) than choosing the riskless option. RD trials exist when choosing the risky option is less beneficial (has a lower expected value outcome) than choosing the riskless option. EQEV trials exist when the benefit for choosing either option is the same.

Taking the different types of trials into consideration, risk-taking behavior for individuals is measured by overall risk-taking and sensitivity to risk. Overall risk-taking can be measured by the number of times the risky option was chosen for a score ranging from 0 (took no risks) to 54 (only took risks). Risk sensitivity for individuals can be measured by subtracting the number of risky choices made for RD trials from the number of risky choices made for RA trials; this creates scores ranging from -18 (took risks only in RD trials) to +18 (took risks only in RA trials). Higher scores on risk sensitivity indicate that more risky decisions were made when the expected value was in one's favor than not.

As noted previously, the measurement of task performance in the current studies was taken from risk sensitivity scores in each round of the Cups task, with the first round scores acting as a baseline of comparison for improvement. Risk sensitivity scores, as compared to overall risk taking, allows for assessment of decision choices in comparison to normatively correct responses. In turn, this allows for a better and more clear assessment of potential improvement in subsequent performance outcomes and overall decision making skills.

Numeracy. Numeracy was measured using the recently developed Weller et al. (2013) 8-item abbreviated numeracy scale. The scale incorporates questions from previous variations of numeracy scales including Lipkus, Samsa, and Rimer, (2001; e.g., "In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?"), Peters, Hibbard, Slovic, and Dieckmann, (2007; e.g., "Suppose you have a close friend who has a lump in her breast and must have a mammography . . . The table below summarizes all of this information. Imagine that your friends tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?"), and Frederick (2005; e.g., "A bat and ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?"). Correct answers are scored as one, and incorrect answers are scored as zero; lower scores on this measure indicate a lower ability to comprehend and manipulate numeric information, while higher scores indicate a higher ability to comprehend and manipulate numeric information.

Dependent measures. After completion of round one of the Cups task, the dependent measures were asked regarding one's level of agreement on a 7-point scale (1 = Strongly Disagree, 7 = Strongly Agree). Questions included the topics of satisfaction ("I feel satisfied with my performance"), willingness to complete the task again ("I would participate in the Cups task again"), how much the participants liked the task ("I enjoyed the Cups task"), feelings of control over the task ("I feel that I can control my performance in the Cups task"), and confidence in improvement ("I am confident I would show improvement on the Cups task if I were to do it again"). Following completion of round two of the Cups task, participants used the same response scale to answer these same questions. At each time point, an overall performance evaluation index was created from compiling all the dependent measures for use in further analyses. Cronbach's alpha for each index was $\alpha = .66$ and $\alpha = .73$ for round one and round two, respectively. Since these indexes maintain acceptable reliability for exploratory studies (Nunnally, 1978; Robinson, Shaver, & Wrightsman, 1991), further analyses using these dependent measures were conducted with the index. Deletion of any item did not appreciably improve reliability, and the item-total correlations are .3 or higher, indicating that each item is meaningful to the scale (Nunnally & Bernstein, 1994).

As a manipulation check, participants were asked to recall their own scores, with the question framed according to their designated feedback framing condition, and participants described their ideas about the hypotheses of the study as a suspicion check. Finally, participants were asked standard demographic information.

Results

Descriptives and Preliminary Analyses

To assure that any changes made to the Cups task did not alter how the task is perceived and completed, one-sample t-tests were run to compare the current sample's overall risk-taking and risk sensitivity scores from round one to the average response from previously collected college student data (Weller et al., 2011). It was expected that round one responses (before any framing manipulations were introduced) would exhibit similar risk-taking and risk sensitivity tendencies as previous samples. A significant difference was seen for overall risk-taking with the current sample taking less risks (M = 30.72, SD = 10.12) than the previous sample (M = 35.25), $t(229) = -6.79$, $p < .001$, d = .89. A significant difference was also seen for risk sensitivity with the current sample showing a lower risk sensitivity (*M* = 0.24, *SD* = 0.29) than the previous sample (*M* = 0.48), $t(229) = -12.37$, $p < .001$, $d = 1.64$. No changes in the pattern of significant differences occurred when separately assessing gain and loss trials. While there were significant differences in both dependent variables, the qualitative differences between the samples were less noteworthy. When assessing performance levels for monetary payouts in previous studies, the differences seen between the current sample and previous samples would not necessitate different payout categories. Establishing similarities in task performance allows for using round one performance as a basis of comparison for improvement in round two and understanding how the different feedback given in round one will affect performance evaluations and influence potential improvement.

In regards to the dependent variables of interest, descriptive statistics were assessed for each variable. Based on differential risk choices in consecutive rounds of the Cups task, across all conditions, the average risk sensitivity difference score was 1.19 (*SD* = 6.78), indicating a general improvement in risk sensitivity. Responses to the five dependent measure questions were combined to form one index of performance evaluation (*M* = 22.33, *SD* = 4.92).

Risk Sensitivity Scores

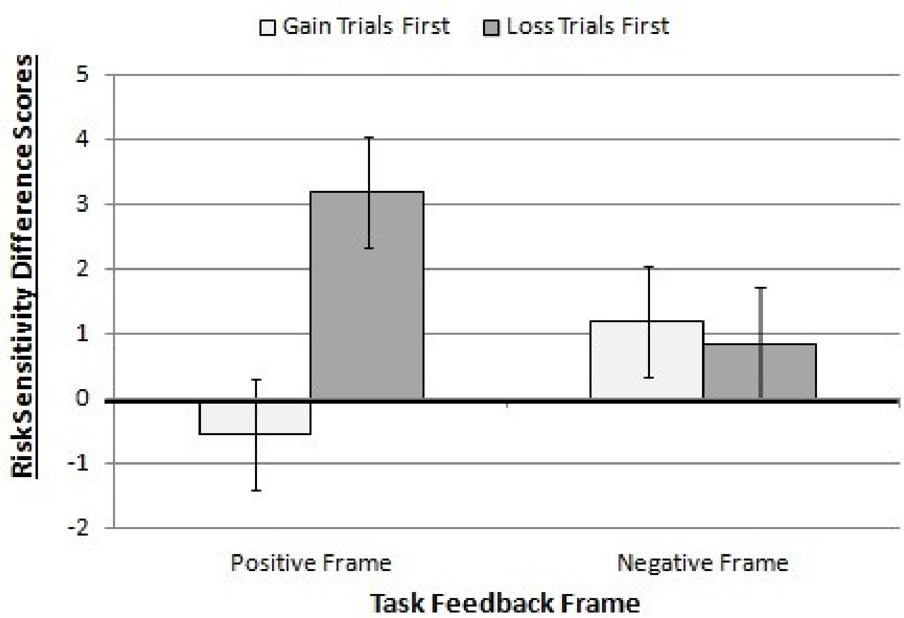
Participants' choices across the different trial types were aggregated to form risk sensitivity scores following the calculations detailed in the *Materials* section above. Risk sensitivity scores for participants acted as the primary dependent measures throughout the current studies and were used to assess the degree of change in task performance, as calculated by finding the difference of participants' risk sensitivity scores between round two and round one of the Cups task. The use of difference scores for analysis in the context of the current studies was supported by guidelines presented in Williams and Zimmerman (1996) and Culpepper (in press).

Risk sensitivity differences scores potentially range from -36 (indicating that the participant took risks only on RA trials for round one and only on RD trials for round two) to +36 (indicating that the participant took risks only on RD trials for round one and only on RA trials for round two). Positive difference scores of task performance for risk sensitivity indicates that during round two, participants were more likely to choose the option with the more favorable expected value outcome and, subsequently, made more normatively correct decisions during round two. Negative difference scores of task performance for risk sensitivity indicates that during round two, participants were less likely to choose the option with the more favorable expected value outcome and, subsequently, made less normatively correct decisions during round two.

The overall risk sensitivity difference scores were submitted to a 2 (task feedback frame: positive vs. negative) X 2 (trial feedback frequency: no trial feedback vs. trial-by-trial feedback) X 2 (order: gains first or losses first) three-way independent ANOVA, with all variables as between-subjects factors. From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. No significant main effect of task feedback frame was found (p = .78), so the hypothesized effect of task feedback frames on improving risk sensitivity and one's tendency to make normatively correct decisions was not supported.

For the effect of trial feedback frequency, it was expected that the trial-by-trial feedback version of the Cups task would promote improvement in risk sensitivity scores across consecutive rounds, while the no trial feedback condition of the task would show less improvement in risk sensitivity scores (or no change at all). As predicted, a significant main effect of trial feedback frequency was found such that the trial-by-trial feedback condition (M = 2.39, SD = 6.11) had larger improvements in risk sensitivity scores across rounds than the no trial feedback condition (M = -0.04, SD = 7.21), $F(1, 222) = 9.24$, p = .003, η^2 = .037.

Figure 1.1 Order X Task Feedback Frame Effect on Risk Sensitivity Difference Scores



Note. The bolded line at the zero mark indicates no change in risk sensitivity scores. Positive scores indicate improvement in risk sensitivity; negative scores indicate decline in risk sensitivity. Error bars indicate ±1 SE.

No main effect was predicted for the variable of order (gain trials first vs. loss trials first); however, a significant main effect was found such that those participants who experienced loss trials first (*M* = 2.07, *SD* = 6.60) showed larger improvements in risk sensitivity scores across rounds than those participants who experienced gain trials first (*M* = 0.33, *SD* = 6.67), $F(1, 222) = 4.62$, $p = .033$, $\eta^2 = .019$.

Although main effects on subsequent risk sensitivity are important to note, the more interesting pattern is potential interaction effects. A significant two-way interaction of Order X Task Feedback Frame was found, $F(1, 222) = 6.51$, $p = .011$, $\eta^2 = .026$ (see Figure 1.1 above).

Specifically, the amount of risk sensitivity improvement associated with negatively framed task feedback did not differ depending on whether gains trials or loss trials came first ($M = 1.20$, $SD = 6.28$ and $M = 0.83$, $SD = 6.76$, respectively), $p = .76$. Improvements in risk sensitivity associated with positively framed task feedback did differ by order; receiving loss trials first lead to larger improvements in risk sensitivity on consecutive rounds of the Cups task ($M = 3.20$, $SD = 6.29$) than receiving gain trials first ($M = -0.55$, $SD = 7.38$), $t(115) = -2.97$, $p = .004$, $d = .55$. No other interactions were found to be significant, *p*s > .13 thus the predicted Trial Feedback Frequency X Task Feedback Frame interaction effect on improved subsequent task performance was not supported.

Performance Evaluations

A between-subjects independent ANOVA with the three main factors (task feedback frame, trial feedback frequency, and order) was used to test differences on the performance evaluation index (including evaluations of satisfaction, willingness to complete the task again, control over the task, confidence in improvement, and enjoyment). It was hypothesized that task feedback frame and trial feedback frequency effects on judgments and intentions may exist either in addition to behavioral effects; thus, the composite scale for round one was expected to show higher (more positive) ratings when task feedback was framed positively than negatively and also when outcome feedback was provided after every trial versus no trial feedback, such that the positive task feedback frame/trial-by-trial feedback condition will show the most positive performance evaluations.

The performance evaluation index was submitted to a 2 (task feedback frame: positive or negative) X 2 (trial feedback frequency: trial-by-trial feedback vs. no trial feedback) X 2 (order: gains first or losses first) three-way independent ANOVA, with all variables as between-subjects factors. From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. For the main effect of task feedback frame, it was expected that positively framed overall task feedback would promote more positive performance evaluations, while negatively framed overall task feedback would show less positive (or even negative) performance evaluations. In support of this hypothesis, a significant main effect of task feedback frame was found such that the positive task feedback frame condition ($M = 23.31$, $SD = 4.79$) had more positive performance evaluations after round one than the negative task feedback frame condition ($M = 21.31$, $SD = 4.86$), $F(1, 222) = 9.81$, $p = .002$, $\eta^2 = .042$. No other main effects or interactions were found to be significant, *p*s > .22; thus the predicted patterns of higher (more positive) ratings when trial feedback was provided and the Trial Feedback Frequency X Task Feedback Frame interaction were not supported.

As no effect of interaction between task feedback frame and trial feedback frequency was seen for risk sensitivity difference scores or round one performance evaluations, and the results from this study suggest that trial feedback frequency has influence only on subsequent risk sensitivity while task feedback frame has influence only on performance evaluations, no mediation analyses were deemed necessary. Therefore, the predicted mediated moderation pattern showing that differences in improved risk sensitivity are a result of more positive or negative performance evaluations, such that participants who receive trial feedback (and thus more instructive and easily integrated feedback information) would use these evaluations to guide their understanding of the task and improve task performance on subsequent trials was not supported.

Additional Analyses

A 2 (task feedback frame: positive vs. negative) X 2 (trial feedback frequency: trial-by-trial feedback vs. no trial feedback) X 2 (order: gains first vs. losses first) three-way ANCOVA was run to determine whether risk sensitivity differences scores significantly differed among participants who had different feedback conditions when adjusted for numeracy scores as a covariate. No significant main effects or interactions were seen in relation to numeracy scores, *p*s > .12, and the noted significant main effects of trial feedback frequency ($F(1, 221) = 9.05$, $p = .003$, $\eta^2 = .037$) and order ($F(1, 221) = 4.51$, $p = .035$, $\eta^2 = .018$) and the Order X Task Feedback Frame interaction ($F(1, 221) = 6.33$, $p = .013$, $\eta^2 = .026$) remained significant. Therefore, it does not appear that numeracy scores account for the relationship between task feedback framing, trial feedback frequency, and subsequent task performance.

Analyses were run to control for potential demographic moderators, such as gender, but no differences were found other than those already noted, *p*s > .20. When assessing gender as a separate factor in relation to the outcomes of interest, no theoretically interesting results or trends were found in the data; thus, these effects were not pursued further.

Relevant ANOVAs were also assessed for overall risk-taking difference scores and gain/loss domain differences in risk sensitivity and found no theoretically interesting results; thus, these effects were not pursued further. While a few significant differences in overall risk taking were seen across consecutive rounds of the Cups task, these differences occurred in conjunction with corresponding differences in risk sensitivity scores; therefore, they do not provide additional information about the influence of feedback on decision making beyond what is noted in relation to risk sensitivity difference scores.

Discussion

Overall, mixed results emerged, with some support for predicted hypotheses. The significant main effect of trial feedback frequency on subsequent task performance indicates that, as predicted, receiving trial-by-trial feedback does elicit more improvement on risk sensitivity scores. This provides initial evidence that the frequency with which one receives trial feedback can influence behavioral outcomes and one's tendency to make more normatively correct decisions in a risky decision making task.

These findings suggest a novel way to relate the effects of feedback frequency to choices between relatively advantageous versus disadvantageous options - the Cups task - and that individual trial feedback is important for adapting behavior to maximize outcomes in a normatively correct fashion throughout the Cups task. This suggests that in line with previous literature, it may not be enough to simply have contingency information about one's choice options (present across all conditions) or know how to complete a task (as in the no trial feedback condition); rather, immediate and corrective feedback along the way is needed to guide one's actions to be more normatively correct, particularly if the task is somewhat difficult (e.g., Clariana et al., 2000; Epstein & Lazarus, 2002; Epstein et al., 2002). This pattern of results also suggests that when given the opportunity to learn from various pieces of feedback, people appear to take advantage of that opportunity and make an attempt to apply the feedback knowledge to their subsequent actions. This provides some support for previous results that show learning can occur across trials of the Cups task (Levin & Bossard, 2011) and places emphasis on distinguishing the role of individual trial feedback that is somewhat unregulated in important domains, such as educational attainment and health prevention outcomes.

Additionally, the significant main effect of task feedback frame on performance evaluations indicates that, as predicted, the positive task feedback frame does elicit more positive performance evaluations, and that manipulating overall task feedback as positive versus negative can influence how one views their overall task performance, confidence in improving performance, and how much one likes the task. These findings are important and add to the literature regarding the influence of attribute framing on evaluations of and associations with various objects. Because the current study emphasized the simplicity and objective similarity between task feedback frames, the difference between the positive and negative frame manipulations may have been strong enough for participants to understand and interpret at a broad, evaluative level but not strong enough to evoke changes in specific behavioral outcomes. Study 1B (described below) will address this issue by utilizing a different, more salient task feedback frame manipulation.

Though not predicted, a significant main effect of order (gain trials first vs. loss trials first) and a significant Order X Task Feedback Frame interaction were found. The main effect of order indicates that overall, experiencing loss trials first elicits more improvement in risk sensitivity scores and may influence one's behavioral outcome tendencies to become more advantageous and normatively correct, but this pattern is qualified by whether overall task performance is framed positively or negatively. If one's overall task feedback is negatively framed, subsequent task performance does not vary (or improve) whether gain or loss trials are presented first; however, if one's overall task feedback is positively framed, subsequent task performance improves if loss trials are presented first.

The effects found in relation to the gain/loss trial order are harder to interpret, as these effects are not generally seen in Cups task data (Levin et al., 2007a, 2007b; Weller et al., 2011) and may not be maintained across subsequent studies. Nevertheless, it is possible that loss trials and the information they provide are viewed more negatively, are harder to understand, and heighten one's tendency towards loss aversion (which then needs to be overcome to avoid taking unfavorable risks; Armstrong, Schwartz, Fitzgerald, Putt, & Ubel, 2002; Ledgerwood & Boydstun, 2014). The noted Order X Task Feedback Frame interaction suggests that at the same time, positively framed task feedback may be spurring more positive performance evaluations and creating some confidence that improved performance is possible in the future. In doing so, there may be a stronger focus on how to achieve better outcomes and avoid losses when given the opportunity on the following round of the Cups task, which leads to improved task performance.

Regardless, the primary goal of Study 1A was accomplished. The primary goal of Study 1A was to develop and test an experimental paradigm that could be used to understand the relationship between two important aspects of performance feedback – task feedback frames and trial feedback frequency – and their evaluative and behavioral outcomes. While no behavioral outcomes were seen in relation to task feedback frames, the trial feedback frequency manipulation did provide evidence to help distinguish what behavioral outcomes may be associated with more immediate versus delayed feedback and the potential importance of providing it in a variety of decision making scenarios. Furthermore, the paradigm created is quite malleable to other framing manipulations within the same context, and provides opportunities to explore and extend the relationship of task feedback frames, trial feedback frequency, and their evaluative and behavioral outcomes.

CHAPTER II

STUDY 1B

Recall, Study 1A showed that the experience of trial-by-trial performance feedback can influence subsequent task performance and framing overall task feedback can influence performance evaluations in positive manners. These findings are interesting and informative, but do not sufficiently address the main hypotheses exploring how task feedback frame and trial feedback frequency may interact to influence performance evaluations and subsequent task performance. The potential ability of task feedback frame and the interaction of task feedback frame and trial feedback frequency to influence following behavioral outcomes in task performance is important to further investigate.

The primary goal of Study 1B is to implement a different, more salient task feedback framing manipulation to help verify the relationship between and the importance of task feedback frames and trial feedback frequency for evaluative and behavioral outcomes. Because the task feedback framing manipulation used in Study 1A focused on simple negation of terms, the relatively unobtrusive nature of the frames may have created reference points that were too similar and too subtle to influence behavioral outcomes in addition to evaluative outcomes. While simple negation represents how task feedback may be provided in some situations (e.g., how many points earned versus how many points not earned), this may not reflect the most common or most salient nature of task feedback presentation. Of note, Vollmeyer and Rheinberg (2005) used opposite frame terms (i.e., percent correct versus percent incorrect) in their classroom-based studies finding that positively framed task feedback led better subsequent performance outcomes, so it is highly possible that while simple negation is enough to influence evaluative tendencies, more distinction between task feedback frames (e.g., "You gained XX quarters…" versus "You lost XX quarters…") are required to elicit both evaluative and behavioral outcomes. Additionally, as the receipt of trial feedback (and knowledge of associated normatively correct and incorrect choices) can influence subsequent task performance, a more salient task feedback frame may lend itself to be more easily integrated with this assessment, thereby increasing the likelihood that an interaction between trial feedback frequency and task feedback frame will arise for both evaluative and behavioral outcome measures.

Study 1B also provides an opportunity for investigating whether the noted effects seen in Study 1A are maintained in an online sample and when providing actual task performance scores. This is of particular importance for the noted effects of gain/loss trial order, as these findings were unexpected and unsupported by prior literature. The use of actual task performance scores will allow for better assessment of the generalizability of the findings to more real-world settings and across those individuals with higher or lower initial performances, and it will relieve any potential feelings of suspicion associated with providing false feedback.

Using a more salient task feedback framing manipulation, main effects and interactions of task feedback frame and trial feedback frequency are expected to be found, such that positively framed feedback will show more positive performance evaluations and more improvement in task performance across consecutive rounds of the Cups task, the trial-by-trial feedback condition will show more positive performance evaluations and more improvement in task performance, and that the difference between trial feedback frequency conditions on both performance evaluations and improved task performance will be larger when task feedback is frame positively versus negatively.

Further, a mediated moderation pattern is predicted whereby the anticipated Trial Feedback Frequency X Task Feedback Frame interaction effect on subsequent task performance may be mediated by performance evaluations. While both task feedback frames and frequency of trial feedback may influence performance evaluations to be more positive or negative following initial completion of the Cups task, when people rate their performance more positively as a result of both receiving feedback after every trial (which provides more instructive feedback and ease of integrating trial feedback information) and positively framed task feedback, people use these evaluations to guide their understanding and improve task performance on subsequent trials.

Method

Participants and Design

Two hundred and five participants (*M* = 35.58 years, *SD* = 12.33 years) were recruited via Mechanical Turk and compensated $1.00 for their time. Of the recruited participants, 54.4% were females and 75.6% were Caucasian. Nineteen participants were dropped from analysis due to failure to recall task performance or failure to pass suspicion checks (note: no changes in data patterns or significance levels occurred whether these participants were excluded from the dataset). The study maintained a similar design to Study 1A: 2 (task feedback frame: positive vs. negative) X 2 (trial feedback: no trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) with all factors manipulated between-subjects.

Materials and Procedure

The materials and procedures were similar to Study 1A, with a few key differences: actual performance scores were provided after both rounds of the task, a different task feedback frame manipulation was used, and the study was conducted online. Participants completed the initial round of the Cups task, being told to make the best decisions they can. Throughout the Cups task, participants either received trial-by-trial feedback regarding the outcome of each decision or no trial feedback. Once finished with all trials of the first round, a monetary total reflecting the participant's true overall performance on the task appeared on each participant's screen framed in either a positive or negative manner. To create more distinct differences between task feedback frames, the positive frame represented how much money one gained throughout the task, with higher totals equaling better performance (e.g., “You have gained XX quarters out of a possible 180 for a total of $XX.”), while the negative frame represented how much money one lost throughout the task, with higher totals equaling lower performance (e.g., “You have lost XX quarters out of a possible 180 for a total of $XX.”).

After receiving their overall performance feedback on round one, participants answered the same set of dependent measures as Study 1A. Participants next completed the Cups task for a second round. The participant’s second round of the Cups task presented the same trial feedback and framing conditions as for the first round.

Participants then filled out the same dependent measures as completed after round one. Participants also filled out the abbreviated numeracy scale (Weller et al., 2013) and demographics following the second round of the Cups task. Finally, participants completed manipulation and suspicion checks before being thoroughly debriefed and thanked.

Results

Descriptives and Preliminary Analyses

Looking at the outcome variables of interest, the current sample reported a standard level of overall risk sensitivity (M = 6.40, SD = 5.65). In regards to the specific metrics assessed in the following analyses, descriptive statistics were assessed for each variable. Based on differential risk choices in consecutive rounds of the Cups task, across all conditions, the average risk sensitivity difference score was 0.95 (SD = 4.80), indicating a slight improvement in risk sensitivity. Responses to the five dependent measure questions were combined to form one index of performance evaluation (M = 25.49, SD = 5.21).

Risk Sensitivity Scores

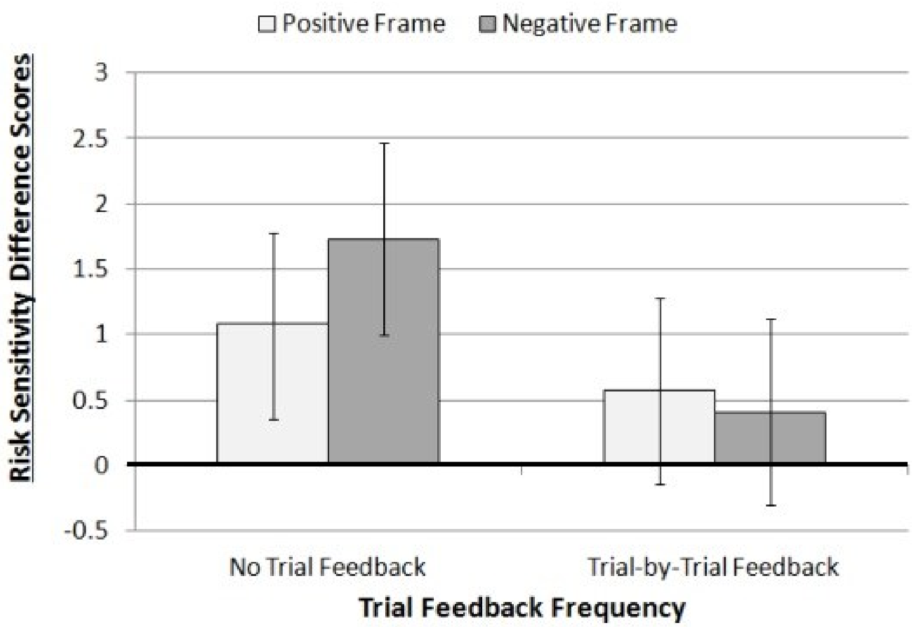
As detailed in Study 1A, risk sensitivity scores were aggregated across trial types and were used to assess the degree of change in task performance, as calculated by finding the difference of participants' risk sensitivity scores between round two and round one of the Cups task. The overall difference scores were submitted to a 2 (task feedback frame: positive vs. negative) X 2 (trial feedback: no trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way independent ANOVA, with all variables as between-subjects factors.

From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. No significant main effect of task feedback frame, trial feedback, or order was found ($p = .77$, $p = .19$, and $p = .82$, respectively), so the hypothesized effects of task feedback frame and trial feedback frequency on risk sensitivity scores and one's tendency to make normatively correct decisions were not supported. In addition, no significant interactions were found (all ps > .19); thus the predicted pattern of Trial Feedback Frequency X Task Feedback Frame interaction on risk sensitivity difference scores was not supported (as seen in Figure 2.1 below).

Performance Evaluations

A between-subjects independent ANOVA with the three main factors (task feedback frame, trial feedback frequency, and order) was used to test differences on the performance evaluation index established in Study 1A. To assess whether task feedback frame and trial feedback frequency effects on judgments and behavioral intentions may exist in addition to behavioral effects, the performance evaluation index for round one was submitted to a 2 (task feedback frame: positive vs. negative) X 2 (trial feedback frequency: no trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way independent ANOVA, with all variables as between-subjects factors.

Figure 2.1 Trial Feedback Frequency X Task Feedback Frame Effect on Risk Sensitivity Difference Scores



Note. The bolded line at the zero mark indicates no change in risk sensitivity scores. Positive scores indicate improvement in risk sensitivity; negative scores indicate decline in risk sensitivity. Error bars indicate ±1 SE.

From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. For the main effect of task feedback frame, it was expected that positively framed overall task feedback would promote more positive performance evaluations, while negatively framed overall task feedback would show less positive (or even negative) performance evaluations. In support of this hypothesis, a significant main effect of task feedback frame was found such that the positive task feedback frame condition (M = 26.33, SD = 4.90) had more positive performance evaluations after round one than the negative task feedback frame condition (*M* = 24.62, *SD* = 5.39), $F(1, 174) = 4.88$, $p = .028$, $\eta^2 = .027$. No other main effects or interactions were found to be significant, *p*s > .15; thus the predicted patterns of more positive ratings when the Cups task supplied trial-by-trial feedback versus no trial feedback and when task feedback frame and trial feedback frequency interact were not supported.

As no effect of interaction between task feedback frame and trial feedback frequency was seen for risk sensitivity difference scores or round one performance evaluations, and the results from this study suggest that no factor had influence on both risk sensitivity and performance evaluations, no mediation analyses were deemed necessary. Therefore, the predicted mediated moderation pattern showing that differences in improved risk sensitivity are a result of more positive or negative performance evaluations, such that participants who receive trial feedback (which provides more instructive feedback and ease of integrating trial feedback information) will use these evaluations to guide their understanding of the task and improve task performance on subsequent trials, was not supported.

Additional Analyses

A 2 (task feedback frame: positive vs. negative) X 2 (trial feedback frequency: trial-by-trial feedback vs. no trial feedback) X 2 (order: gains first vs. losses first) three-way ANCOVA was run to determine whether risk sensitivity differences scores significantly differed among participants who had different feedback conditions when adjusted for numeracy scores as a covariate. Numeracy was found to be a significant predictor of risk sensitivity difference scores ($\beta = 0.358$; $t(177) = 2.13$, $p = .035$, $\eta^2 = .025$), indicating that higher numeracy scores were associated with larger, positive differences in risk sensitivity across consecutive rounds of the Cups task. The pattern of significant main effects and interactions did not change, *p*s > .26; therefore, it does not appear that numeracy scores account for the (lack of) relationship between task feedback framing, trial feedback frequency, and subsequent task performance.

Analyses were run to control for potential demographic moderators, such as gender, and no differences were found, *p*s > .15. When assessing gender as a separate factor in relation to the outcomes of interest, no theoretically interesting results or trends were found in the data; thus, these effects were not pursued further.

To compare how the effects of task feedback frame and trial feedback frequency may differ between those participants who had higher versus lower initial performance, a multiple regression was run including round one performance (measured by risk sensitivity score) as a factor in the overall model predicting improved task performance (measured by risk sensitivity difference scores). When including round one performance as a predictor, the overall model was not significant ($F(12, 173) = .987$, $p = .463$, $R^2 = .064$), and round one performance was not found to be a significant predictor of task performance improvement on consecutive rounds of the Cups task, $p > .57$. No other significant main effects or interactions were observed, *p*s > .11 so it does not appear that having a higher or lower initial task performance influences one's subsequent task performance in positive or negative ways. The same pattern of results was found when predicting performance evaluation index, all *p*s > .13.

Relevant ANOVAs were also assessed for overall risk-taking difference scores and gain/loss domain differences in risk sensitivity and found no theoretically interesting findings; thus, these effects were not pursued further. While a few significant differences in overall risk taking were seen across consecutive rounds of the Cups task, these differences occurred in conjunction with corresponding differences in risk sensitivity scores; therefore, they do not provide additional information about the influence of feedback on decision making beyond what is noted in relation to risk sensitivity difference scores.

Discussion

Overall, the changes made to Study 1B provided little additional support for the predicted effects of different feedback manipulations. Neither task feedback frame nor trial feedback frequency was found to significantly influence behavioral outcomes, as no differences were seen in risk sensitivity scores across consecutive rounds of the Cups task. It was noted that higher numeracy (those with a better ability to use numerical information) was associated with larger differences in risk sensitivity (indicating larger improvements in task performance); however, controlling for this relationship did not alter the relative influence of the different feedback manipulations. Therefore, the predictions that receiving trial-by-trial feedback and positively framed task feedback elicit more improvement in task performance (or more normatively correct decisions) in a risky decision making task were not supported.

In regards to evaluative outcomes, as predicted, the positively framed task feedback elicited significantly more positive performance evaluations than the negatively framed task feedback. It appears the more salient task framing manipulation of gain versus loss was found to influence both task performance and performance evaluations much in the same way as the task framing manipulation in Study 1A (earned vs. failed to earn); therefore, the exact framing used to present overall feedback in the context of risky decision making tasks may not make much difference. Consequently, the following studies will continue to use the gain/loss task feedback frame manipulation as it is more in line with relevant manipulations in the previous literature. Additionally, as no effect of interaction between task feedback frame and trial feedback frequency was seen for subsequent task performance or round one performance evaluations, no meditational relationships were assessed.

The primary goals of Study 1B were to create more effective feedback conditions through increasing the saliency of the task feedback frame manipulation and assess the generalizability of the findings in Study 1A. In regards to these goals, the lack of significant findings in Study 1B makes it harder to assert they were met, but partial progress was made towards both. In support of the goal of salient task feedback frames, the effect on performance evaluations was maintained, but the more salient task feedback frames did not impact behavioral outcomes, indicating that it did not necessarily create more effective feedback conditions. Differences between Study 1A and 1B do go beyond task feedback frame manipulations, so the difference in behavioral outcomes may relate to a difference in methods of data collection or the type of scores provided (fake versus true). As this set of studies represents a novel way to present multiple feedback manipulations in one decision making task, further testing is needed to attend to the inconsistency of task performance outcomes (as will be addressed with Study 2A below).

In addition to seeing similar evaluative outcome effects across Study 1A and 1B, partial support of the goal of generalizability was found in that no differences in significant effects across feedback conditions were noted whether initial task performance was high or low or by numeracy level. While at first glance, this may seem potentially uninteresting, it suggests that what influence feedback manipulations do exhibit is consistent and generalizable across samples and methods of data collection. This has implications for the power of feedback manipulations (in this case, task feedback frame) in influencing thoughts and feelings about one’s performance that may be important for other outcomes in real-world scenarios.

CHAPTER III

STUDY 2A

In Studies 1A and 1B, the focus was on understanding what would happen under different task feedback frames and trial feedback frequency on evaluative and behavioral outcomes during a risky decision making task. In these studies, the issue arose that despite having similar study designs and the expectation that a more salient frame would enhance the results found in Study 1A, Studies 1A and 1B exhibited few similar results.

Given similar effects of task feedback frame on performance evaluations across the two studies, it appears that the most obvious difference between Study 1A and Study 1B is between the type of score provided – fake or actual overall performance outcomes. Study 2A was also conducted with an online sample and included the conditions seen in Study 1B (additional conditions are described below). Conducting Study 2A in this manner can help further address whether the critical feedback manipulations hold any influence on behavioral (in addition to evaluative) outcomes in a more general sample. Verifying what results are consistent and reliable is important in order to understand how different aspects of feedback presentation can be more effective in creating both positive evaluative and behavioral outcomes.

As most studies have found that providing feedback, as compared to no feedback, throughout a task can lead to better outcomes in a variety of settings (e.g., Bangert-Drowns et al., 1991; Kulik & Kulik, 1988), it seems that the advantage found for providing trial feedback on subsequent task performance in Study 1A may be seen in the current study, as well. If a difference is found in which receiving trial feedback leads to better evaluative and behavioral outcomes, it will also be important to see whether adding intermediary conditions of trial feedback frequency may help elucidate these differences.

Little research has investigated trial feedback frequency intervals beyond a general comparison of immediate versus delayed (though see Schmidt, Young, Swinnen

& Shapiro (1989) for one such example) and the effects of immediate versus delayed feedback in educational settings is fraught with inconsistencies and confounds (including different forms, amount of detail, and relevance of feedback across conditions; e.g., Dihoff et al., 2003; Erev et al., 2006); thus, one of the main aims of Study 2A will be to expand on these findings and investigate at what frequency interval trial feedback is distinguished from the no trial feedback condition and leads to more positive evaluations and improved task performance.

As mentioned, additional trial feedback frequency conditions will be included to represent intermediary intervals of feedback presentation. Specifically, the intermediary conditions will include trial feedback presented after every three trials (which represents the three trials of each trial type in the Cups task), after every nine trials, and after each gain/loss trial block (27 trials each). In addition to clarifying designation of immediate versus delay, the trial feedback presented will help control for some of the methodological disparities in the previous research, using the same form, relevance, and amount of detail of feedback across conditions in Study 2A (i.e., number of quarters won or lost on given trials). Any differences arising between conditions should therefore reflect the true nature of the influence of trial feedback frequency.

As noted previously, receiving any amount of experiential information can lead people to discount provided descriptive information and weight the experiential information more heavily in order to guide their decisions (Jessup et al., 2008; Lejarraga & Gonzalez, 2011). Therefore, it is predicted that when adding the intermediary frequency conditions, trial feedback at any presentation interval produces a similar result, providing information that is instructive and useful in guiding subsequent decisions in comparison to receiving no trial feedback throughout the initial task completion. However, it is possible that providing larger blocks of trial feedback less often will result in lessened ability to process the amount of information and directly associate it to one’s choices and outcomes. This may lead to a lessened influence of trial feedback frequency on both evaluative and behavioral outcomes as the frequency interval gets larger.

It is predicted that the partially supported predicted effects of task feedback frame and trial feedback frequency will be found, such that positively framed task feedback will show more positive performance evaluations and the trial-by-trial feedback condition will show more improvement in risk sensitivity scores across consecutive rounds of the Cups task. In addition, it is expected that the intermediary trial feedback frequency conditions (that also provide more instructive information by tying the feedback information to previous choices and outcomes) will not be significantly different from the trial-by-trial feedback condition, but will be significantly different from the no trial feedback condition (which will not be instructive enough to help guide future actions). Therefore, receiving any trial feedback is expected to lead to more positive performance evaluations and more improvement in risk sensitivity scores.

Outcomes to be assessed will also include any effect of task feedback frame beyond more positive performance evaluations, including potential effects on risk sensitivity and a Trial Feedback Frequency X Task Feedback Frame interaction effect for both performance evaluations and risk sensitivity scores. The Trial Feedback Frequency X Task Feedback Frame interaction of interest would show that the difference between performance evaluations and risk sensitivity scores for conditions receiving trial feedback versus no trial feedback will be significantly larger for participants who receive positively framed task feedback than participants who receive negatively framed task feedback.

Further, if a significant interaction occurs, a mediated moderation pattern is predicted whereby the anticipated Trial Feedback Frequency X Task Feedback Frame interaction effect on risk sensitivity scores may be mediated by performance evaluations. While both task feedback frames and trial feedback may influence performance evaluations to be more positive or negative, when people rate their performance more positively as a result of receiving both trial feedback (and thus, providing more instructive feedback and ease of integrating feedback information) and positively framed feedback, they will use these evaluations to guide their understanding of the task and ability to improve task performance on subsequent trials. Participants who receive no trial feedback will not accomplish this due to lack of instructive, easily integrated feedback.

To summarize, the primary goal of Study 2A is to clarify the distinction between what constitutes immediate versus delayed feedback and establish the interval of trial feedback frequency necessary to produce the effects associated with immediate feedback, compared to delayed feedback, including greater performance evaluations and better (more normatively correct) choices during a subsequent decision making task. In doing so, Study 2A will also address issues corresponding to the disparity of results in the previous studies by replicating the cells of Study 1B and helping establish what behavioral and evaluative outcomes the critical feedback manipulations hold in a more generalizable sample.

Method

Participants and Design

Two hundred and ninety-nine participants (*M* = 36.49 years, *SD* = 12.43 years) were recruited via Mechanical Turk and compensated $1.00 for their time. Of the recruited participants, 52.7% were females and 80.2% were Caucasian. Fifteen were dropped from analysis due to failure to recall task performance or failure to pass suspicion checks (note: no changes in data patterns or significance levels occurred whether these participants were excluded from the dataset). The study maintained a similar design to Study 1B: 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first); however, the trial feedback frequency variable included additional levels of feedback frequency. All factors were manipulated between-subjects.

Materials and Procedure

The materials and procedures were the same as in Study 1B with the exception of the additional feedback conditions within the trial feedback frequency variable. As in Study 1B, participants completed the initial round of the Cups task, being told to make the best decisions they can. Participants were counterbalanced across five trial feedback frequency conditions: no trial feedback, block trial feedback, 9 trial feedback, 3 trial feedback, and trial-by-trial feedback. All participants completed the same number and combination of trials throughout the Cups task, but the trial feedback frequency interval varied. The no trial feedback and trial-by-trial feedback conditions kept the same presentation as described in the previous studies. In each of the intermediary conditions, participants viewed quarters as being gained or lost as in the trial-by-trial feedback condition; however, this information was presented at different time points – after every three trials, after every nine trials, and after each gain/loss domain block (27 trials). More specifically, after the designated number of trials in each condition was completed, participants were provided with a screen showing the trials they had responded to since their last feedback update and how many quarters were won or lost on each trial (see Appendix - Figure A3 – parts c, d, and e for more detail).

Once finished with all trials of the first round, as in Study 1B, a monetary total reflecting the participant's true overall performance on the task appeared on the screen framed either a positive or negative manner. After receiving their overall task feedback on round one, participants answered a set of dependent measures regarding evaluations of the task and their performance (see *Materials* - Study 1A for exact measures). Participants next completed the Cups task for a second round. The participant's second round of the Cups task was the same task feedback frame and trial feedback frequency condition as received during the first round.

Participants then filled out the same dependent measures as completed after round one. Participants also filled out the abbreviated numeracy scale (Weller et al., 2013) and demographics following the second round of the Cups task. Finally, participants completed manipulation and suspicion checks before being thoroughly debriefed and thanked.

Results

Descriptives and Preliminary Analyses

The current sample reported a standard level of overall risk sensitivity (*M* = 5.57, *SD* = 4.93). In regards to the specific metrics assessed in the following analyses, descriptive statistics were assessed for each variable. Based on differential risk choices in consecutive rounds of the Cups task, across all conditions, the average risk sensitivity difference score was 0.61 (*SD* = 4.70), indicating a slight improvement in risk sensitivity. Responses to the five dependent measure questions were combined to form one index of performance evaluation (*M* = 25.21, *SD* = 5.14).

Risk Sensitivity Scores

As detailed in Study 1A, risk sensitivity scores were aggregated across trial types and were used to assess the degree of change in task performance, as calculated by finding the difference of participants' risk sensitivity scores between round two and round one of the Cups task. The overall difference scores were submitted to a 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way independent ANOVA, with all variables as between-subjects factors.

From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. No significant main effect of task feedback frame, trial feedback frequency, or order was found ($p = .85$, $p = .84$, and $p = .22$, respectively), so the hypothesized main effects of task feedback frame and trial feedback frequency on subsequent risk sensitivity and one's tendency to make normatively correct decisions were not supported.

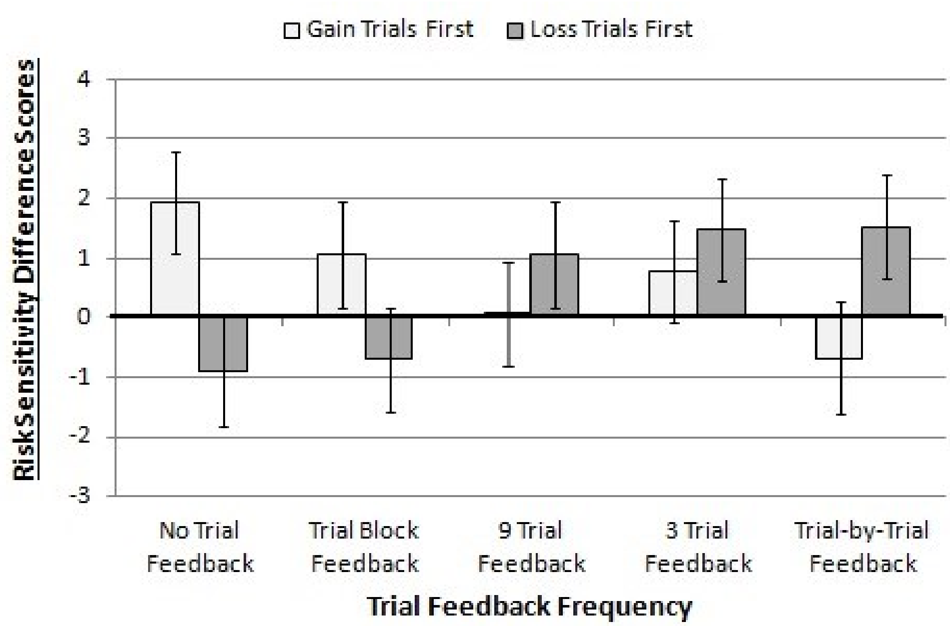
Although no significant main effects on subsequent risk sensitivity were found, a significant two-way interaction of Trial Feedback Frequency X Order was found, $F(4, 264) = 2.71$, $p = .031$, $\eta^2 = .038$ (see Figure 3.1 below). Simple effect tests for order revealed that in the trial-by-trial feedback condition, receiving loss trials first lead to marginal improvements in risk sensitivity on consecutive rounds of the Cups task than receiving gain trials first ($M_{diff} = -2.20$, $SE = 1.28$), $F(1, 274) = 2.96$, $p = .086$, $\eta_p^2 = .011$; however, in the no trial feedback condition, the reverse pattern occurs. Receiving gain trials first lead to larger improvements in risk sensitivity on consecutive rounds of the Cups task than receiving gain trials first ($M_{diff} = 2.81$, $SE = 1.27$), $F(1, 274) = 4.94$, $p = .027$, $\eta_p^2 = .018$. The amount of risk sensitivity improvement associated with the three middle levels of trial feedback frequency (3 trial, 9 trial, trial block conditions) did not differ between whether gain trials or loss trials were presented first, *p*s > .16.

In both cases, when improvement in risk sensitivity scores across consecutive rounds of the Cups task (designated by positive difference scores) was noted, the amount of improvement was marginally different from zero ($t(29) = 1.71$, $p = .098$, $d = .31$ and $t(28) = 2.04$, $p = .051$, $d = .38$, respectively). The lower, negative difference scores were not significantly different from zero, *p*s > .38, indicating no change in risk sensitivity across consecutive rounds of the Cups task.

Simple effects test for trial feedback frequency were not found to be significant, *p*s > .13. Despite the overall nonsignificant simple effects of trial feedback frequency, pairwise comparisons were explored to see if any trends could be seen across the frequency conditions. As can be seen in Table 3.1 below, Fisher LSD post hoc comparisons found that when gain trials were presented first, the trial-by-trial feedback condition had significantly lower risk sensitivity difference scores than the no trial feedback condition. When loss trials were presented first, both the trial-by-trial and three trial feedback conditions had marginally higher risk sensitivity differences scores than the no trial feedback condition. No other comparisons between groups were significant.

No other interactions were found to be significant, *p*s > .14; thus the predicted Trial Feedback Frequency X Task Feedback Frame interaction effect on risk sensitivity difference scores was not supported.

Figure 3.1 Trial Feedback Frequency X Order Effect on Risk Sensitivity Difference Scores



Note. The bolded line at the zero mark indicates no change in risk sensitivity scores. Positive scores indicate improvement in risk sensitivity; negative scores indicate decline in risk sensitivity. Error bars indicate ±1 SE.

Table 3.1 Mean Differences and Standard Errors of Risk Sensitivity Scores Between Trial Feedback Frequency Conditions

(a) Trial Feedback Frequency Conditions Within Gains First Order

	1	**2**	**3**	**4**	**5**
1 – No Trial Feedback	-				
2 – Trial Block Feedback	-1.45 (1.27)	-			
3 – 9 Trial Feedback	-0.75 (1.28)	0.70 (1.22)	-		
4 – 3 Trial Feedback	-1.75 (1.29)	-0.31 (1.23)	-1.00 (1.24)	-	
5 – Trial-by-Trial Feedback	-2.61* (1.27)	-1.17 (1.21)	-1.86 (1.22)	-0.86 (1.23)	-

Note. $^{+}p < .10$, $*p < .05$

(b) Trial Feedback Frequency Conditions Within Losses First Order

	1	**2**	**3**	**4**	**5**
1 – No Trial Feedback	-				
2 – Trial Block Feedback	0.03 (1.21)	-			
3 – 9 Trial Feedback	0.45 (1.24)	0.41 (1.22)	-		
4 – 3 Trial Feedback	2.21^{+} (1.23)	2.17^{+} (1.21)	1.76 (1.24)	-	
5 – Trial-by-Trial Feedback	2.40^{+} (1.28)	2.36^{+} (1.26)	1.95 (1.29)	0.19 (1.28)	-

Note. $^{+}p < .10$, $*p < .05$

Performance Evaluations

A between-subjects independent ANOVA with the three main factors (task feedback frame, trial feedback frequency, and order) was used to test differences on the performance evaluation index established in Study 1A. To assess whether task feedback frame and trial feedback frequency effects on judgments and behavioral intentions may exist in addition to or instead of behavioral effects, the performance evaluation index for round one was submitted to a 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way independent ANOVA, with all variables as between-subjects factors.

From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. For the main effect of task feedback frame, it was expected that positively framed overall task feedback would promote more positive performance evaluations, while negatively framed overall task feedback would show less positive performance evaluations. In support of this hypothesis, a significant main effect of task feedback frame was found such that the positive task feedback frame condition (M = 26.31, SD = 4.14) had more positive performance evaluations after round one than the negative task feedback frame condition (M = 24.07, SD = 5.80), $F(1, 260) = 13.96$, $p < .001$, $\eta^2 = .049$. No other main effects or interactions were found to be significant, ps > .13; thus the predicted patterns of more positive ratings when the Cups task supplied trial feedback versus no trial feedback and when task feedback frame and trial feedback frequency interact were not supported.

As no effect of interaction between task feedback frame and trial feedback frequency was seen for risk sensitivity difference scores or round one performance evaluations, and the results from this study suggest that no factor had influence on both risk sensitivity and performance evaluations, no mediation analyses were pursued. Therefore, the predicted mediated moderation pattern showing that differences in improved risk sensitivity are a result of more positive or negative performance evaluations was not supported.

Additional Analyses

A 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way ANCOVA was run to determine whether risk sensitivity differences scores significantly differed among participants in different feedback conditions when adjusted for numeracy scores as a covariate. No significant main effects or interactions were seen in relation to numeracy scores, *p*s > .14, and the noted Trial Feedback Frequency X Order interaction ($F(4, 263) = 2.62$, $p = .035$, $\eta^2 = .037$) remained significant. Therefore, it does not appear that numeracy scores account for the relationship between task feedback framing, trial feedback frequency, and subsequent task performance.

Analyses were run to control for potential demographic moderators, such as gender, but no differences were found other than those already noted, *p*s > .17. When assessing gender as a separate factor in relation to the outcomes of interest, no theoretically interesting results or trends were found in the data; thus, these effects were not pursued further.

To compare how the effects of task feedback frame and trial feedback frequency may differ between those participants who had higher versus lower initial performance, a multiple regression was run including round one performance (measured by risk sensitivity score) as a factor in the overall model predicting improved task performance (measured by risk sensitivity difference scores). When including round one performance as a predictor, the overall model was marginally significant ($F(39, 244) = 1.35$, $p = .094$, $R^2 = .177$), and round one performance was not found to be a significant predictor ($p > .37$) of task performance improvement on consecutive rounds of the Cups task. No other significant main effects or interactions were observed in relation to round one performance, *p*s > .13 and the noted Trial Feedback Frequency X Order interaction remained as a significant predictor. Therefore, it does not appear that having a higher or lower initial task performance influences one's subsequent task performance in positive or negative ways. A similar pattern of results was found when predicting performance evaluation index, with the main effect of task feedback frame remaining a significant predictor of performance evaluations.

Relevant ANOVAs were also assessed for overall risk-taking difference scores and gain/loss domain differences in risk sensitivity and found no theoretically interesting findings; thus, these effects were not pursued further. While a few significant differences in overall risk taking were seen across consecutive rounds of the Cups task, these differences occurred in conjunction with corresponding differences in risk sensitivity scores; therefore, they do not provide additional information about the influence of feedback on decision making beyond what is noted in relation to risk sensitivity difference scores.

Discussion

Significant findings from Study 2A included a Trial Feedback Frequency X Order interaction. It was seen that when trial outcome feedback was provided following each trial of the Cups task, having loss trials presented first led to more improvement in consecutive rounds, while when no trial feedback was provided, having gain trials presented first led to more improvement in consecutive rounds. Though no predictions were made in regards to this Trial Feedback Frequency X Order interaction and the effects of order were somewhat unexpected, it does seem to suggest that trial feedback frequency, particularly outcome feedback provided after every trial, has some influence on improving subsequent task performance.

It is possible that how one initially comes to rely on trial feedback versus the outcome contingency information available in each trial varies by order of gain/loss trials. For example, the outcome contingency information available in loss trials may be perceived as more difficult to interpret and synthesize (Armstrong et al., 2002; Ledgerwood & Boydstun, 2014), so having outcome feedback after every trial is helpful for understanding the various outcome contingencies and which risks are best to approach or avoid. This acknowledgement of the usefulness of trial feedback may carry over into subsequent rounds of the Cups task, as well, and help produce improvements in task performance as people learn to rely on the feedback to guide their choices.

The outcome contingency information in gain trials may not be as difficult to interpret and synthesize, so if no trial feedback is provided, people may simply focus on the information available in the trial itself to designate which risks are best to approach or avoid. This acknowledgement of the usefulness of outcome contingency information may carry over into subsequent rounds of the Cups task, as well, and help produce improvements in task performance as people learn to rely on the contingency information they are given to guide their choices. Though the above example is speculative in terms of the mechanism through which this interaction may occur, the main point remains that trial feedback frequency is an important variant of performance feedback manipulation to consider for behavioral outcomes.

Indeed, though exploratory, when looking at how task performance improved across the different trial feedback frequency conditions within the interaction, the main distinctions came between the extreme ends of frequency – trial-by-trial versus no trial feedback. Though across the order conditions the relative importance of trial feedback or no trial feedback on improved task performance differed, as both order conditions show this general division, it does offer some support for the initial finding in Study 1A that there are distinct differences between providing trial feedback and not providing feedback for improving one's subsequent performance in a risk decision making task. Of course, the interaction with gain/loss order complicates potential conclusions and requires further research (as will be discussed in the *Future Directions*) to determine the exact differences between and any moderators of immediate versus delayed performance feedback effects.

In addition, the predicted main effect of task feedback frame was again found in relation to evaluative outcomes; the positively framed task feedback elicited significantly more positive performance evaluations than the negatively framed task feedback. In relation to previous research on framing effects and the consistency with which task feedback frame has influenced performance evaluations in the studies thus far, it appears that framing overall task feedback is an important variant of performance feedback for inducing beneficial evaluative outcomes.

As the primary goal of Study 2A was to help delineate the effects of different trial feedback frequency presentations on evaluative and behavioral outcomes, finding no differences within or between the intermediary feedback frequency conditions did not provide support for achieving this goal. As no distinctions were drawn between the intermediary frequency conditions, this suggests that simply providing trial feedback, but not necessarily following each trial, may not be informative enough to influence behavioral or evaluative outcomes in regards to one's performance and normatively correct decision making tendencies. However, a more definitive difference seems to exist between the ends of the trial feedback frequency spectrum, suggesting that an "all or nothing" approach to providing trial feedback is something to consider when constructing feedback presentations. Overall, the predictions surrounding an exact split of which trial feedback frequency conditions should be considered more immediate versus delayed were not supported, but as will be discussed in further detail in Study 2B, this may be due in part to the type of outcome feedback provided in the Cups task.

In addition, some conclusions can be drawn about how generalizable the findings from Study 1A are to broader samples and when using different types of performance outcomes. The reported results from Study 2A suggest that both behavioral and evaluative performance outcomes can be measured in an online sample and when providing true overall outcomes scores. Further evidence of the generalizability of these findings was found in that, again, neither numeracy level nor initial high or low performance seemed to alter these outcomes. Therefore, the significant effects noted here are more likely to be consistent across and generalizable to other outcomes in real-world scenarios.

CHAPTER IV

STUDY 2B

Across Studies 1A, 1B, and 2A, somewhat mixed results were found concerning the effects of task feedback frame and trial feedback frequency on evaluative and behavioral outcomes in a risky decision making task. Interestingly, a trend does appear across the studies discussed thus far whereby different aspects of feedback manipulation correspond with different measured outcomes. Notably, though manipulated within the same decision making task, when trial feedback frequency manipulations produce effects, they tend to be associated more with changes in task performance, while the effects of task feedback frame manipulation tend to be associated more with differences in performance evaluations. Identifying this correspondence between feedback manipulations and outcomes may be a way to establish what the most effective types of feedback are for eliciting different performance outcomes, but as many of the predicted effects have not been noted as significant, it is harder to draw conclusions of the true influence of each feedback manipulation. One possibility, to draw out more of the predicted effects associated with the different aspects of feedback manipulation and to clarify the influential nature of different feedback manipulations is to modify the outcome feedback provided by the Cups task to be more useful and relevant in guiding one's choices.

As discussed previously, trial outcome information in the Cups task indicates normatively correct and incorrect responses, where trials that are risk advantageous necessarily result in more beneficial risk outcomes (i.e., winning in the gain domain and not losing in the loss domain) and trials that are risk disadvantageous necessarily result in more detrimental risk outcomes (i.e., not winning in the gain domain and losing the loss domain). Despite this relationship between trial types and outcomes, the decision choices are binary in the Cups task, so if the riskless option is chosen, the trial outcome information may not convey the advantageous/disadvantageous element of responses as clearly. Therefore, the standard presentation of trial outcome information in the Cups task may be more of a verification of one's response than actual information regarding the relative correctness or incorrectness of that response, and this trial feedback information may not provide enough detail to really guide behaviors beyond the initial task.

Study 2B will help to clarify the issue of whether the trial feedback normally provided in the Cups task is the most relevant and effective for guiding future decisions. Essentially, more detail regarding the context in which choices occurs (i.e., whether the probability of winning/losing a trial was favorable, indicating the normatively correct response) and the decision option one chose on that trial will be provided as part of the Cups task trial feedback. Trial outcome feedback that functions more in line with verifying responses may not be viewed as more useful, as easily integrated into knowledge, or produce as many positive behavioral change as more elaborate, relevant feedback (e.g., Anderson et al., 1971; Bangert-Drowns et al., 1991; Fazio et al., 2010; Pashler et al., 2005). The additional trial feedback information provided in Study 2B can therefore help decision makers determine which choices were better or worse and why that may be in order to guide their future actions.

These more useful aspects of feedback relevance may explain why improved risk sensitivity was seen across a single round of the Cups task but not consistently across consecutive rounds. When assessing performance across three trials in a row, trial outcome feedback may be enough to indicate the relative correctness of one's responses because the exact same context is presented immediately. Comparatively, to apply trial outcome feedback across consecutive rounds may require more elaborate feedback (in a sense, a "why" to the outcome provided) to strengthen the associated between one's response and its outcome.

Increasing the relevance of the trial outcome feedback will create more detailed, corrective feedback useful for guiding one's thoughts and behaviors. As discussed in previous literature, feedback needs to be seen as relevant in order to produce positive outcomes (Bangert-Drowns et al., 1991; Fazio et al., 2010; Pashler et al., 2005); therefore, more relevant trial feedback in the Cups task will enhance the potential for the task feedback frame and trial feedback frequency manipulations to show both behavioral and evaluative outcomes and augment those results (i.e., task feedback framing on performance evaluations) seen more consistently across the previous studies.

Specifically, it is predicted that main effects of task feedback frame and trial feedback frequency will be found such that positively framed task feedback and the trial-by-trial feedback condition will show more positive performance evaluations and more improvement in risk sensitivity scores across consecutive rounds of the Cups task. It is less clear whether the intermediary trial feedback frequency conditions will be more associated with the trial-by-trial feedback condition than the no trial feedback condition; however, a clear distinction in more positive performance evaluations and improved risk sensitivity scores is expected to occur between those receiving trial-by-trial feedback and those receiving no trial feedback. It is also possible that as the trial feedback is more detailed and provides context for making normatively correct or incorrect decisions, providing larger blocks of trial feedback less often will result in lessened ability to process the amount of information and directly associate it to one's choices and outcomes. This may lead to a heightened influence of trial feedback frequency on both evaluative and behavioral outcomes as the frequency interval gets smaller.

The Trial Feedback Frequency X Task Feedback Frame interaction is also expected to be significant. It is predicted that the difference between performance evaluations and risk sensitivity scores for conditions receiving positively versus negatively frame task feedback will be largest for conditions that receive trial-by-trial feedback than conditions that received no trial feedback.

Further, if a significant interaction occurs, a mediated moderation pattern is predicted whereby the anticipated Trial Feedback Frequency X Task Feedback Frame interaction effect on risk sensitivity scores may be mediated by performance evaluations. While both task feedback frames and trial feedback may influence performance evaluations to be more positive or negative, when people rate their performance more positively as a result of receiving both trial feedback (and thus, receive more detailed, instructive feedback) and positively framed feedback, they will use these evaluations to guide their understanding of the task and ability to improve task performance on subsequent trials. Participants who receive no trial feedback will not accomplish this due to lack of detailed, instructive feedback.

To summarize, the primary goal of Study 2B is to test how more relevant and detailed trial outcome feedback can affect evaluative and behavioral performance outcomes, both in detecting more relationships between feedback manipulations and outcomes and augmenting established relationships. Identifying how trial feedback that is deemed more useful can relate to the relative influences of task feedback frame and trial feedback frequency manipulations may also help shed light on how to better determine effective feedback strategies to produce more normatively correct decision making overall, particularly as they relate to various target outcomes. Indeed, implementing more relevant trial outcome feedback will strengthen the potential for both evaluative and behavioral outcomes to emerge and allow for confirmation of the trend between corresponding feedback manipulations and outcome measures (trial feedback frequency and task feedback frame manipulations associated more with influence on behavioral and evaluative outcomes, respectively).

Method

Participants and Design

Three hundred and five participants (M = 35.5 years, SD = 11.81 years) were recruited via Mechanical Turk and compensated $1.00 for their time. Of the recruited participants, 57.5% were females and 77.9% were Caucasian. Sixteen were dropped from analysis due to failure to recall task performance or failure to pass suspicion checks (note: no changes in data patterns or significance levels occurred whether these participants were excluded from the dataset). The study maintained the same design as Study 2A: 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first), with all factors manipulated between-subjects.

Materials and Procedure

The materials and procedures were the same as in Study 2A with the exception of providing the trial feedback in a different format. As in Study 2A, participants completed the initial round of the Cups task, being told to make the best decisions they can. Participants were counterbalanced across the five trial feedback frequency conditions, and all participants completed the same number and combination of trials throughout the Cups task. As in Study 2A, after the designated number of trials in each condition was completed, participants were provided with a screen showing the trials they had responded to since their last feedback update, but instead of standard trial outcome feedback, the feedback stated whether each trial was associated with advantageous or disadvantageous risk taking, their choice, and the number of quarters won/lost (e.g., "On Trial 1, there were 3 cups and the chance to win 5 quarters, and this represented a risk advantageous trial. You chose the risky side and won 5 quarters.").

Once finished with all trials of the first round, a monetary total reflecting the participant's true overall performance on the task appeared on the screen framed in either a positive or negative manner. After receiving their overall task feedback on round one, participants answered a set of dependent measures regarding evaluations of the task and their performance (see *Materials* - Study 1A for exact measures). Participants next completed the Cups task for a second round. The participant's second round of the Cups task was the same task feedback frame and trial feedback frequency condition as received during the first round.

Participants then filled out the same dependent measures as completed after round one. Participants also filled out the abbreviated numeracy scale (Weller et al., 2013) and demographics following the second round of the Cups task. Finally, participants completed manipulation and suspicion checks before being thoroughly debriefed and thanked.

Results

Descriptives and Preliminary Analyses

The current sample reported a standard level of overall risk sensitivity (M = 5.62, SD = 5.35). In regards to the specific metrics assessed in the following analyses, descriptive statistics were assessed for each variable. Based on differential risk choices in consecutive rounds of the Cups task, across all conditions, the average risk sensitivity difference score was 0.51 (SD = 4.95), indicating a slight improvement in risk sensitivity. Responses to the five dependent measure questions were combined to form one index of performance evaluation (M = 25.57, SD = 4.76).

Risk Sensitivity Scores

As detailed in Study 1A, risk sensitivity scores were aggregated across trial types and were used to assess the degree of change in task performance, as calculated by finding the difference of participants' risk sensitivity scores between round two and round one of the Cups task. The overall difference scores were submitted to a 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way independent ANOVA, with all variables as between-subjects factors.

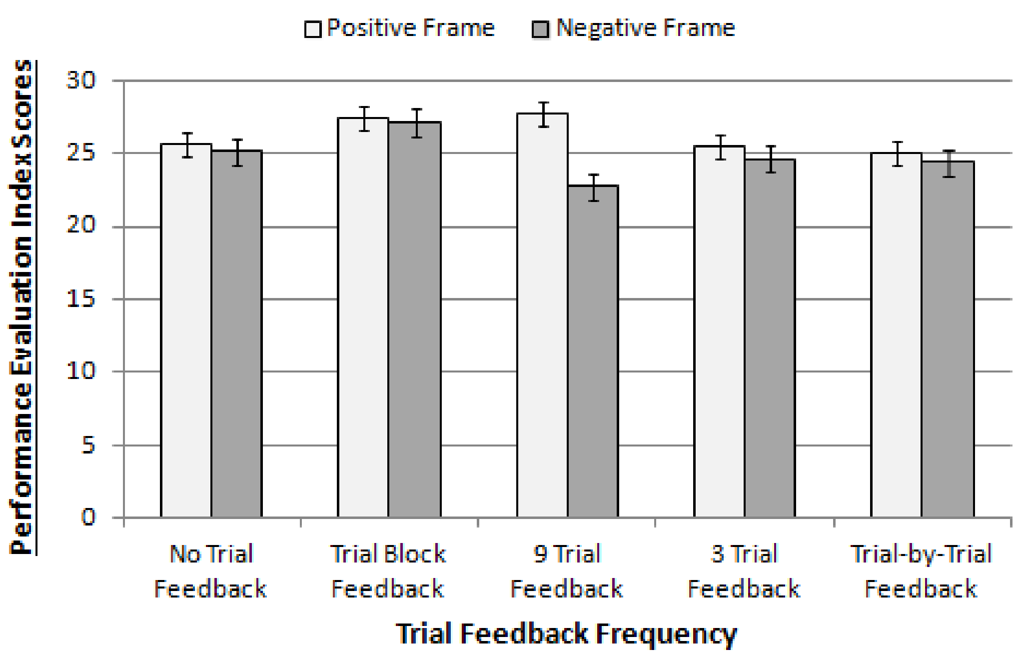
From the ANOVA, main effects of task feedback frame, trial feedback frequency, and order were assessed. No main effect was predicted for the variable of order (gain trials first vs. loss trials first); however, a significant main effect was again found such that those participants who experienced loss trials first (M = 1.25, SD = 5.01) showed larger improvements in risk sensitivity scores across rounds than those participants who experienced gain trials first (M = -0.20, SD = 4.80), $F(1, 269) = 6.29$, $p = .013$, $\eta^2 = .022$. No other main effects or interactions were found to be significant, ps > .23; thus the predicted patterns of improved risk sensitivity when the Cups task supplied trial feedback versus no trial feedback and when task feedback frame and trial feedback frequency interact were not supported.

Performance Evaluations

A between-subjects independent ANOVA with the three main factors (task feedback frame, trial feedback frequency, and order) was used to test differences on the performance evaluation index established in Study 1A. The performance evaluation index for round one was submitted to a 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way independent ANOVA, with all variables as between-subjects factors.

From the ANOVA, main effects of task feedback frame ($F(1, 263) = 6.56$, p = .011, $\eta^2 = .022$) and trial feedback frequency ($F(4, 263) = 2.59$, $p = .037$, $\eta^2 = .035$) were found to be significant. The main effect of task feedback frame was found to be in the hypothesized direction (positively framed overall task feedback promoted more positive performance evaluations, while negatively framed overall task feedback promoted less positive performance evaluations), but this was moderated by trial feedback frequency. As seen in Figure 4.1 below, the significant interaction of Trial Feedback Frequency X Task Feedback Frame suggests that while overall, positively framed task feedback leads to more positive performance evaluations than negatively framed task feedback, the magnitude of this difference varies depending on how often trial feedback is provided ($F(4, 263) = 2.91$, $p = .022$, $\eta^2 = .039$). No other main effects and interactions were found to be significant, *p*s > .49.

Figure 4.1 Trial Feedback Frequency X Task Feedback Frame Effect on Performance Evaluation Index Scores



Note. Higher index scores = more positive performance evaluations. Error bars indicate ±1 SE.

Table 4.1 Mean Differences and Standard Errors of Performance Evaluation Index Scores Between Trial Feedback Frequency Conditions

(a) Trial Feedback Frequency Conditions Within Positive Task Feedback Frame

	1	**2**	**3**	**4**	**5**
1 – No Trial Feedback	-				
2 – Trial Block Feedback	-0.45 (1.21)	-			
3 – 9 Trial Feedback	-2.70* (1.17)	2.25^{+} (1.17)	-		
4 – 3 Trial Feedback	-2.38* (1.18)	-1.93 (1.18)	0.32 (1.14)	-	
5 – Trial-by-Trial Feedback	-0.58 (1.19)	-0.13 (1.19)	2.12^{+} (1.15)	1.79 (1.16)	-

Note. $^{+}p < .10$, $*p < .05$

(b) Trial Feedback Frequency Conditions Within Negative Task Feedback Frame

	1	**2**	**3**	**4**	**5**
1 – No Trial Feedback	-				
2 – Trial Block Feedback	-0.36 (1.30)	-			
3 – 9 Trial Feedback	1.57 (1.26)	1.93 (1.27)	-		
4 – 3 Trial Feedback	-2.85* (1.31)	-2.49^{+} (1.31)	-4.41* (1.28)	-	
5 – Trial-by-Trial Feedback	-0.87 (1.28)	-0.51 (1.28)	-2.43^{+} (1.24)	1.98 (1.29)	-

Note. $^{+}p < .10$, $*p < .05$

Simple effect tests for task feedback frame revealed that the only significant difference between positive and negative task feedback frames was noted for the nine trial feedback condition, $F(1, 273) = 17.81$, $p < .001$, $\eta_p^2 = .061$, in which the positive task feedback frame condition lead to more positive performance evaluations than the negative task feedback frame condition ($M_{diff} = 4.98$, $SE = 1.18$). Performance evaluations associated with the other levels of trial feedback frequency did not differ whether the task feedback frame was positive or negative, *p*s > .52.

Simple effects test for trial feedback frequency were also found to be significant, $F(4, 273) = 2.22$, $p = .067$, $\eta_p^2 = .031$ and $F(4, 273) = 3.09$, $p = .016$, $\eta_p^2 = .043$ for gain first and loss first orders, respectively. Pairwise comparisons were explored to see if any trends could be seen across the frequency conditions, and as can be seen in Table 4.1 above, Fisher LSD post hoc comparisons found several significant differences across both order conditions.

Though the pattern of differences is less consistent and distinct, when looking at how performance evaluations changed across the different trial feedback frequency conditions within the interaction, it appeared that the two most frequent presentations were more similar to each other, while the two least frequent presentations were more similar to each other. This suggests a possible distinction between the most frequent trial feedback conditions and the least frequency trial feedback conditions. The overall trend of this distinction seemed to be that trial feedback presented less frequently resulted in more positive performance evaluations, with a more pronounced increase or decrease for the nine trial condition, depending on the frame valence. Thus, the hypothesized Trial Feedback Frequency X Task Feedback Frame interaction was found, but not necessarily in the predicted direction.

As no effect of interaction between task feedback frame and trial feedback frequency was seen for subsequent task performance, and the results from this study suggest that no factor had influence on both task performance and performance evaluations, no mediation analyses were pursued. Therefore, the predicted mediated moderation pattern showing that differences in improved risk sensitivity are a result of more positive or negative performance evaluations was not supported.

Additional Analyses

A 2 (task feedback frame: positive vs. negative) X 5 (trial feedback frequency: no trial feedback vs. trial block feedback vs. 9 trial feedback vs. 3 trial feedback vs. trial-by-trial feedback) X 2 (order: gains first vs. losses first) three-way ANCOVA was run to determine whether risk sensitivity differences scores significantly differed among participants in different feedback conditions when adjusted for numeracy scores as a covariate. No significant main effects or interactions were seen in relation to numeracy scores, *p*s > .22, and the noted main effect of order ($F(1, 268) = 5.63$, $p = .018$, $\eta^2 = .019$) remained significant. Therefore, it does not appear that numeracy scores account for the relationship between task feedback framing, trial feedback frequency, and subsequent task performance.

To compare how the effects of task feedback frame and trial feedback frequency may differ between those participants who had higher versus lower initial performance, a multiple regression was run including round one performance (measured by risk sensitivity score) as a factor in the overall model predicting improved task performance (measured by risk sensitivity difference scores). When including round one performance as a predictor, the overall model was significant ($F(30, 258) = 2.04$, $p = .002$, $R^2 = .192$), and round one performance was found to be a significant predictor of risk sensitivity differences scores ($\beta = -0.406$, $t(258) = -2.417$, $p = .016$, $\eta^2 = .022$, indicating that higher initial performance was associated with larger, negative differences in risk sensitivity (less improvement in task performance) across consecutive rounds of the Cups task. No other significant main effects or interactions were observed in relation to round one performance, *p*s > .15 and the noted main effect of order remained as a significant predictor. Therefore, it does not appear that higher or lower initial Cups task performance account for the relationship between task feedback framing, trial feedback frequency, and subsequent task performance. A similar pattern of results was found when predicting performance evaluation index, with the main effect of task feedback frame remaining a significant predictor of performance evaluations.

Analyses were run to control for potential demographic moderators, such as gender, but no differences were found other than those already noted, $ps > .22$. When assessing gender as a separate factor in relation to the outcomes of interest, no theoretically interesting results or trends were found in the data; thus, these effects were not pursued further.

Relevant ANOVAs were also assessed for overall risk-taking difference scores and gain/loss domain differences in risk sensitivity and found no theoretically interesting findings; thus, these effects were not pursued further. While a few significant differences in overall risk taking were seen across consecutive rounds of the Cups task, these differences occurred in conjunction with corresponding differences in risk sensitivity scores; therefore, they do not provide additional information about the influence of feedback on decision making beyond what is noted in relation to risk sensitivity difference scores.

Discussion

In Study 2B, the primary goal was to investigate how more relevant and detailed trial outcome feedback may affect evaluative and behavioral performance outcomes, with particular attention on its role to elicit more positive outcomes associated with completing consecutive rounds of a risky decision making task and how various aspects of feedback manipulations may be differentially more effective for these evaluative and behavioral outcomes.

Providing more relevant trial outcome feedback in conjunction with the established feedback manipulations of frame and frequency resulted in a significant main effect of order (gain trials first versus loss trials first) on improved task performance and a significant Trial Feedback Frequency X Task Feedback Frame interaction on performance evaluations. For the main effect of order, receiving losses first was found to lead to more improvement in task performance than receiving gains first. As described previously, this finding is somewhat unexpected, but it is possible that overall, loss trials are perceived more negatively and as more difficult than gain trials when starting the Cups task. Participants may have a harder time figuring out how to approach and synthesize the outcome contingency information in the various trials. It is possible that this enhanced difficulty may result in lower initial performance when loss trials are presented first and simply leave more room for improvement on subsequent task performance, as compared to the relative ease of figuring out how to approach various trials and have higher initial performance when gain trials are presented first.

The more interesting findings result from the significant Trial Feedback Frequency X Task Feedback Frame interaction on performance evaluations. While overall, positive task feedback frames led to more positive performance evaluations than negative task feedback frames, this difference was only significant when trial feedback was presented after every nine trials.

This suggests that the capacity of task feedback frames to promote more positive performance evaluations depends on how frequently trial feedback is provided and how one comes to rely on trial feedback versus the outcome contingency information available in each trial. Specifically, it may be the case that when trial feedback is presented more frequently, people may learn to rely on the trial feedback as an indicator of how well they are doing (e.g., are they winning or losing lots of quarters) on the task. Because a solid sense of how well one is doing has been built by receiving trial feedback frequently throughout the task, when confronted with overall task feedback frame information, the general positive or negative sense from the frame may not sway evaluations of one's performance. Indeed, the receipt of trial feedback (which includes half gain framed outcomes and half loss framed outcomes) may dilute the potential effect of the frame manipulation.

On the other hand, when trial feedback is presented less frequently, people may learn to not rely on trial feedback as an indicator of how well they are doing because they do not receive it very often or at all. Not receiving trial feedback very frequently may result in people not paying as much attention to the feedback, and instead, focusing on the outcome contingency information presented within each trial. Because of this lack of attention given to trial feedback, when confronted with overall task feedback frame information, this feedback information may be disregarded as well, and the general positive or negative sense from the frame may not sway evaluations of one's performance.

In the nine trial condition, the trial feedback is presented on a regular interval, but spaced somewhat further apart and containing a reasonable amount of information to be interpreted and integrated. Under these circumstances, the trial feedback information may be presented frequently enough to be relied upon to provide a sense, though perhaps less clear, of how one is doing on the task. When the overall task feedback information is provided as positive or negative, attention is paid to it to provide further information regarding one's performance, so the general positive or negative feeling from the frame may sway evaluations of one's performance in a valence-consistent direction.

Though a less distinct pattern, when looking at how performance evaluations changed across the different trial feedback frequency conditions within the interaction, it appeared that the two most frequent presentations were more similar to each other, while the two least frequent presentations were more similar to each other. In addition, the overall trend for both positive and negative task feedback frames seemed to suggest that trial feedback presented less frequently resulted in more positive performance evaluations, with a more pronounced increase or decrease for the nine trial condition, depending on the frame valence. Opposite than the predicted pattern of results, this trend is speculated to potentially result from a focus on the task at hand and working through the decision scenarios in a deliberate, mastery-oriented manner. Working in this manner could provide one with a sense of satisfaction of how he/she approached and completed the task (perhaps by feeling like one made the “best” choices), and with little outcome feedback to potentially affect these impressions, this could result in higher performance evaluations (Ames & Archer, 1988; Harackiewicz, Barron, Tauer, Carter, & Eliott, 2000).

Of course, the noted Trial Feedback Frequency X Task Feedback Frame interaction effect on performance evaluations did not occur in the prior studies, so this suggests that the more relevant nature of the trial outcome feedback in the current study (incorporating the context and relative advantageous nature of the decision scenario) had some influence on eliciting more effects than the standard trial outcome feedback (which only provided outcome information).

Therefore, while no predictions were made in regards to this Trial Feedback Frequency X Task Feedback Frame interaction pattern and the above example is quite speculative in terms of the mechanism through which this interaction may occur, it does seem to suggest that the influence of task feedback frame (and to a lesser extent, trial feedback frequency) corresponds more readily to be effective feedback for producing evaluative performance outcomes and that more immediate versus delayed trial feedback may provide some distinction in the pattern of these outcomes.

In response to Study 2B’s goals, applying more relevant trial outcome feedback did not seem to influence behavioral performance outcomes. More relevant trial feedback did seem to be providing some general guidance for performance outcomes compared to the previous trial outcome feedback, as round one performance was found to be a significant predictor of improved task performance, but this did not change the more consistent influence of gain/loss trial order on improved task performance. In noting little influence of the more relevant trial feedback on behavioral performance outcomes, this may indicate that the information chosen to enhance the relevance of the trial feedback in Study 2B, though likely more interpretable and informative than the standard feedback, did not represent the optimal provision of information. While the terms “risk advantageous” and “risk disadvantageous” are the most accurate in describing the trials and providing context for whether one’s choice was better or worse, more simple terminology may be more appropriate and interpretable for participants to use in guiding subsequent choices. Simpler, more interpretable terms (e.g., good bet versus bad bet) may lead to the intended and enhanced effects of the relevance manipulation on behavioral outcomes as predicted in Study 2B.

Increasing the relevance of trial outcome feedback did, however, seem to influence evaluative performance outcomes. In noting these results, a stronger correspondence between specific feedback manipulations (task feedback frame) and outcome categories (evaluative outcomes) was designated. Potential was also seen for investigating a distinction between more and less frequent trial feedback for evaluative outcomes in addition to behavioral outcomes.

CHAPTER V

SUMMARY, IMPLICATIONS, AND FUTURE RESEARCH

Four studies addressed the influencing roles of feedback frame, frequency, and relevance on evaluative and behavioral outcomes in a risky decision making task. Of particular interest was identifying the most effective forms of feedback to elicit more normatively correct decision making – in other words, to promote more positive outcomes by improving one's decision making skills.

Three key aspects of feedback were identified as potentially important for constructing more effective performance feedback from prior literature: the feedback frame (positive or negative), the frequency of feedback, and the relevance of feedback. Specifically, it was predicted that in order for performance feedback to produce more positive evaluations of one's performance and more improvement in task performance on consecutive rounds of a risky decision making task, the overall task feedback should be framed in a positive manner (e.g., how much was earned or gained throughout the task), trial feedback should be provided frequently (e.g., after every trial), and that the trial feedback should be relevant for guiding subsequent behavior (e.g., provide the context of the decision and one's choice simultaneously with the trial outcome). It was also proposed that in combination with each other, trial feedback frequency and task feedback frame should promote the most positive outcomes, and more relevant trial feedback should enhance these effects. Finally, it was proposed that any improvement in task performance may be explained by performance evaluations as a mediating factor. Across the four studies, support for these hypotheses was mixed and not always in the predicted direction. With these mixed results, some effects did consistently arise and other interesting trends occurred that present intriguing follow-up questions to be explored in future research.

Evaluative Outcomes – Performance Evaluations

When assessing evaluative outcomes related to performance evaluations, task feedback frame showed the most consistent results. Framing overall task performance as positive was found to be associated with more positive performance evaluations following the initial round of the Cups task in all four studies (as a main effect) and to interact with trial feedback frequency in Study 2B.

While the main effect of task feedback frame is straightforward, the interaction between the task feedback frame and trial feedback frequency requires a brief explanation. Overall, framing task feedback positively led to more positive performance evaluations than framing task feedback negatively, but this difference was only significant for the intermediary condition in which trial feedback was presented after every nine trials. In addition, a slight trend was found suggesting that the most frequent trial feedback conditions had lower performance evaluations than the least frequent trial feedback conditions. Across all four studies, it appears that trial feedback frequency may have some influence on more positive performance evaluations when the feedback is more relevant, but the most effective factor for producing positive performance evaluations is task feedback frame, particularly positive frames.

Of note, modifying the relevance of the trial feedback in Study 2B did appear to alter the relationship between the specified feedback manipulations and performance evaluations; the main effect of task feedback frame exhibited across all the studies was moderated by the trial feedback frequency manipulation. Trial feedback frequency did not seem to produce any other data patterns of interest nor did gain/loss trial order produce any significant differences in performance evaluations. In addition, across all four studies, neither numeracy level nor initial performance on the Cups task altered the pattern of results.

As the influence of task feedback frame on evaluative outcomes was the only result to remain consistent throughout all four studies, it is important to determine why that may be. It has been previously shown that positive labeling leads to more favorable associations across a variety of dimensions in one's memory, and negative labeling leads to more unfavorable associations across a variety of dimensions in one's memory (Levin & Gaeth, 1988; Levin et al., 1985). A similar association may be occurring in the current studies where, in addition to thinking about the actual score being provided, positive task feedback frames lead to more thoughts about what was done right throughout the task, while negative frames lead to more thoughts about what was done wrong throughout the task. It has also been found that having some idea of what the correct responses are for a task is important to produce positive outcomes from feedback (e.g., Anderson et al., 1971; Bangert-Drowns et al., 1998; Pashler et al., 2005). If positive frames do elicit more thoughts about what responses were correct, then this should lead to more positive performance evaluations. Given that task feedback frame manipulations in all studies were based on attribute framing, it would be expected that these associations produce consistent effects of task feedback frame in each study.

Behavioral Outcomes - Task Performance

When assessing behavioral outcomes related to task performance, trial feedback frequency and gain/loss trial order showed the most consistent results. Presenting outcome feedback after every trial was found to produce more improved task performance in consecutive rounds of the Cups task in both Study 1A (as a main effect) and Study 2A (as an interaction with order). In relation to the order of gain/loss trial presentation, receiving losses trials first was associated with more improved task performance in consecutive rounds of the Cups task in Studies 1A, 2B (both as main effects), and 2A (as the interaction with trial feedback frequency noted above).

While the main effects of both trial feedback frequency and gain/loss order are straightforward, the interaction between the two requires a brief explanation. Specifically, when trial outcome feedback was provided following every trial of the Cups task, having loss trials presented first led to more improvement in consecutive rounds, while when no trial feedback was provided, having gain trials presented first led to more improvement in consecutive rounds. In conjunction with the main effect findings in Study 1A and 2B, this suggests that both gain/loss trial order and trial feedback frequency have influence on improving subsequent task performance, but perhaps particularly so for trial-by-trial feedback and when loss trials are presented first.

Framing overall task feedback as positive or negative did not seem to produce any data patterns of interest, with the only significant effect of frame being moderated by the more consistent influence of gain/loss trial order. Introducing more relevant trial feedback in Study 2B did not appear to enhance or improve the relationship between the specified feedback manipulations and improved task performance, noting the only significant finding to be a similar effect of gain/loss trial order as in Studies 1A and 2A (when provided with the standard, less informative trial feedback). In addition, across all four studies, neither numeracy level nor initial performance on the Cups task altered the pattern of results.

Categorizing Immediate Versus Delayed Feedback

While the effects of trial feedback frequency noted above do not provide exact evidence of how different frequencies of trial feedback can influence improvement in task performance outcomes, they do suggest that trial feedback frequency is important in some capacity for guiding future behaviors. In concurrence with the main effect found in Study 1A that providing trial feedback provides a distinct advantage for improving one's subsequent performance in a risk decision making task, the Trial Feedback Frequency X Order interaction in Study 2A found that across both order conditions, the main distinction between the relative influence on improved task performance came between trial-by-trial feedback versus no trial feedback conditions.

Establishing even initial evidence to be able to distinguish between the effects of trial-by-trial feedback versus no trial feedback help extend the literature on what feedback frequency intervals should be categorized as immediate versus delayed feedback. Recall, the effects associated with more immediate versus delayed feedback have been inconsistent. Part of this issue is the lack of feedback frequency categorization and use of the same terms for different operationalizations. Some studies have considered overall task feedback to be immediate feedback, while others have considered individual trial feedback (similar to the trial-by-trial feedback condition in the current studies) to be immediate, and very few studies have examined the two simultaneously (although see Schmidt et al., 1989 and Erev et al., 2006 for examples of this). In the current studies, it was seen that there were differences between the two intervals of trial feedback, so they should not automatically be considered part of the same category in future research; however, as no differences were seen between the three intermediary conditions, no conclusions can yet be drawn on whether these conditions should be considered more like immediate or delayed feedback.

Order Effects on Task Performance

The other relatively consistent finding in relation to task performance outcomes was that the order in which gain and loss trials were presented had influence on whether task performance improved across consecutive rounds of the Cups task. As can be seen in the summary above, the losses first condition leads to more improvement in task performance in some capacity for Studies 1A, 2A, and 2B, either as a main effect or as part of an interaction.

These findings were unexpected; however, the consistency with which they occurred across independent sets of data warrants some consideration for what may be producing these effects. One potential reason why receiving loss trials first showed more improvement in task performance is that the outcome contingency information available in loss trials may be perceived as more difficult to interpret and synthesize for learning which risks are best to approach or avoid, whereas the outcome contingency information in gain trials may not be as difficult to do so (e.g., Armstrong et al., 2002; Ledgerwood & Boydstun, 2014). If loss trials are indeed more difficult than gain trials, starting out with losses may set the stage for lower performance on the initial round of the Cups task, thereby leaving more opportunity for improvement on later rounds. Through experiencing these loss trials first, any feedback associated with the losses may weigh more heavily in one's mind. As most people strive to avoid losses, one may have to overcome the natural tendency of taking more (sometimes unnecessary or disadvantageous) risks to avoid losses and put more effort in to making better choices on subsequent rounds of the Cups task.

Having outcome feedback after every trial (at any level of relevance) may be viewed as more helpful for interpreting the various outcome contingencies in the case of loss trials because it can be viewed as providing corrective information, which is key to promoting more positive behavioral outcomes (Bangert-Drowns et al., 1991). This reliance on provided trial feedback may guide choices to produce improved task performance. If no trial feedback is provided, the outcome contingency information in gain trials may still be interpretable and be seen as useful in guiding choices for improved task performance. The Trial Feedback Frequency X Order interaction noted in Study 2A may be explained in a similar fashion.

A comparison of task performance between the successive rounds of the Cups task in Studies 1A and 2B supports the general idea that starting with losses in the Cups may be more difficult for people. Indeed, the improvement seen across consecutive rounds when loss trials were presented first served to increase task performance up to the same level of risk sensitivity as those in the gain trials first conditions (which did not differ across rounds). Of course, with no direct measures of how loss versus gain trials are perceived in difficulty and no prior research to provide context, the exact reason for these findings cannot be determined in the current studies, so future research will need to address potential mechanisms and how order effects may be important for effective feedback implementation in other decision making tasks and scenarios where both gain and loss information may occur concurrently (discussed below).

Correspondence Between Feedback Manipulations and Outcome Types

From the significant behavioral and evaluative outcome findings, it can be seen that not all feedback manipulations work in the same way, despite the ability to manipulate them within the same task. Part of this may have to do with the manipulation itself – feedback frame or frequency. Specifically, when task feedback is framed in a positive or negative manner, the frame provides an overall valenced feeling about one's performance (e.g., performed well or did not perform well), but not much detail about what specific choices were made well or poorly throughout the Cups task. This relative lack of detail may have contributed to the comparability of the positive and negative frames; positively framed task feedback is more definitive in how to proceed with further responses by focusing on what was chosen correctly than negatively framed task feedback is by focusing on what was chosen incorrectly. In other words, there may be only one way to respond correctly but multiple ways to respond incorrectly.

While evaluative outcomes and intentions to improve or not may be influenced by the general feelings about one's performance and relative difference in provided information across frames, without specific knowledge of what choices and behaviors led to better or worse performance, there may not be enough information in either of the task feedback frames (but particularly in the negative frame) to prompt change in behavioral outcomes or improve task performance. In addition, it is possible that positively framed task feedback may contribute to feeling that one already knows how to approach the Cups task effectively so less effort is applied on subsequent tasks and no change in behavioral outcomes occurs; however, without a direct measure of effort or intentions of effort, this explanation remains an empirical question.

In contrast, receiving trial feedback more frequently can provide concrete outcome information and a robust sense of how one performed on the Cups task, particularly what choices had desired or undesired outcomes. Having a more explicit knowledge of what choices were more beneficial or detrimental from trial feedback can, and should, influence behavioral outcomes because it is more closely aligned with an understanding of what behaviors can improve task performance. As trial feedback provides information of both desired and undesired outcomes, though, being made aware of the poor choices made may counteract any positive feelings about the good choices made, producing no effect on overall performance evaluations.

In conjunction, though stemming from slightly different reasoning, experiencing frequent trial feedback, of which half the feedback is presented as a gain and half is presented as a loss, may dilute the potential effect of the task feedback frame on behavioral outcomes. In essence, when trial feedback is present, receiving the task feedback frame as part of the overall task performance score may not provide any necessary or pertinent additional information to guide subsequent choices if one has already been exposed to concrete gain and loss outcomes.

Based on seeing stronger correspondence for trial feedback frequency with behavioral outcomes and task feedback frame with evaluative outcomes, one can surmise that clearly designating the goal in providing feedback is important in order to choose the most effective feedback to achieve that goal. If one’s goal is to motivate people or make them feel good about their decision making capabilities, feedback manipulations should center on how overall task feedback is framed; particularly, framing this feedback in a positive light. If one’s goal is to actually improve people’s decision making skills, feedback manipulations should instead center on the frequency with which trial feedback is provided; particularly, providing this feedback more frequently.

Therefore, it may not be enough to feel positive or negative about one's performance in order to promote improved risk sensitivity in consecutive rounds of the Cups task, but evaluations of one's performance could potentially lead to other behavioral downstream consequences. An overall positive perspective of one's performance and associated dimensions of decision making tasks or risk taking in general may make one more inclined to approach these types of activities later on as compared to someone with an overall negative perspective of one's performance and associated dimensions. As one's general stance towards decision scenarios and participation in decision making tasks become more positive, this may encourage the development of better decision making skills in the long run. Future research should focus on including additional behavioral and behavioral intention measures that could address whether this relationship exists.

Even with the noted correspondence between different feedback manipulations and specific outcomes, both evaluative and behavioral outcomes show trends that a distinction may be made between the differential influence of trial feedback frequency at the extreme ends of the spectrum (though in opposite directions). As explained in more detail above, in general, different patterns of behavioral outcomes occur when comparing trial-by-trial feedback to no trial feedback conditions, with some evidence to suggest that presenting trial-by-trial feedback leads to more improvement in task performance than receiving no trial feedback. A similar pattern, though also speculative and less distinct, was found for evaluations of performance when more relevant trial feedback was provided. It appeared that the two most frequent trial feedback presentations had more similar levels of performance evaluations, and the two least frequent presentations had more similar levels of performance evaluations. In addition, the overall trend seemed to suggest that trial feedback presented less frequently resulted in more positive performance evaluations.

While overall, the trends are not strong enough to make clear assertions about where a split between immediate versus delayed feedback categorizations should be (the intermediate conditions did not show much differentiation), simply making a distinction between providing trial feedback for every choice or not at all is important to consider and extends the current literature on this matter.

If these general trends for both task performance and performance evaluations can be replicated in future research, it would also suggest that reinforcement by trial feedback may be needed to create positive outcomes in terms of task performance but reinforcement by trial feedback is not necessarily needed to create positive performance evaluations. Knowing whether trial feedback is beneficial, or even needed, to achieve more positive consequences of one's decision making would provide a deeper understanding as to why there seems to be a disconnection between feedback effects on evaluative and behavioral outcomes.

Implications and Future Directions

Theoretical Implications

These studies provide an initial demonstration that different aspects of feedback (in particular, task feedback frame and trial feedback frequency) can influence both evaluative and behavioral outcomes in an adaptive risky decision making task to produce more normatively correct decisions. This demonstration also provides evidence that each aspect of feedback affects specific types of outcomes more readily, and supplies suggestions for how to construct more effective feedback given one's desire to measure certain outcomes.

These findings have implications for the literature on the role and influence of feedback. While a multitude of studies have been conducted addressing the effects of feedback across a variety of scenarios, there is little consistency within the findings. The current studies address these inconsistencies and offer some clarity as to what types of trial and task feedback are most helpful for producing positive performance evaluations and influencing one's ability to identify better or worse risk taking tendencies so as to improve task performance. Specifically, the current studies designate that more frequent trial feedback intervals (e.g., trial-by-trial feedback) produce the behavioral outcomes associated with immediate feedback, while providing just an overall score at the end of the task (e.g., no trial feedback) produces the behavioral outcomes associated with delayed feedback; thus, they should be categorized as such. As these terms have been frequently used to describe different feedback conditions, the current findings suggest that feedback effects can be and should be classified by their outcomes, not by their relative standing to other conditions in a given study.

The current studies represent an unexplored way to relate trial feedback to choices between relatively advantageous versus disadvantageous options (the Cups task) and extend the knowledge of trial feedback effects to understand the potential interactive relationship with attribute framing on subsequent evaluations and decisions. As many of the previous studies addressing the effects of performance feedback were done in applied settings (e.g., Choe, Lee, Munson, Pratt, & Kientz, 2013; Vollmeyer & Rheinberg, 2000, 2005), they lacked some of the experimental control associated with a lab setting. Even though few significant results were seen, the current studies using strictly manipulated variables in a more controlled laboratory setting help provide more generalizable evidence of how the frame and frequency of feedback can influence performance evaluations and subsequent task performance; however, it remains to be seen whether the noted results can be generalized to other decision making tasks. Further, the findings represent new evidence for determining how the relevance of trial outcome feedback may influence evaluative and behavioral decision making outcomes.

Perhaps most importantly, these studies provide some evidence for a better understanding of what conditions are important for eliciting better decision making in the future. A large majority of judgment and decision making research focuses on decision making deficits and times when people fail to make good decisions. The focus in the current studies on creating and testing conditions that can potentially improve decision making skills provides additional support for movement towards a more positive approach to judgment and decision making research and the potential mechanism through which decision making skills can be improved.

Practical Implications

In addition to informing various areas of basic research, the current studies also have important real world implications. Decision making in the real world is messy and fraught with instances of significant decisions in which the probabilities and outcomes are unclear or not presented ahead of time; however, as people traverse through the decision making process, feedback is received in a variety of formats (e.g., overall positive and negative feelings and specific gain and loss outcomes), and how they receive and interpret this information can influence how they feel about their decisions and future decision making tendencies. Specifically, the results of these studies may be useful for individuals or organizations seeking to improve evaluations and decision making for health and education contexts.

For example, doctors and dieticians may incorporate these results into how they approach working with patients to create diet and exercise plans. It takes time to discover what plan may work best for a patient due to scheduling, finances, and medical history, and oftentimes, plans need to be tweaked and re-evaluated to achieve the optimal strategy. Knowing when to provide positive overall feedback about one’s progress, but also when to provide detailed, instructive feedback about what seems to be or not be working, is important for achieving positive evaluative outcomes and improved behavioral outcomes. Implementing the type of feedback necessary to produce the desired outcomes (more positive self-evaluations or better health behaviors) may guide patients to better decisions about how to approach their diet and exercise plans going forward.

Similar approaches could be taken by educators looking to improve students' performance on various school subjects. Providing feedback information about individual questions or responses that allows a student to see what he/she did correctly and incorrectly rather than a general, overall score at the end of the task may allow for better understanding of what to do differently next time to improve test scores. Future research could build upon these ideas to translate the findings of the studies to a variety of real-world settings.

Future Directions

As the previous sections focus on the potential implications of the current results, this highlights a need for further research on these topics. Though future directions have been briefly noted in the preceding sections, further research on effective feedback presentations should focus on understanding the exact roles of feedback frame, frequency, and relevance and how they may interact with each other to influence evaluative and behavioral outcomes, particularly in ways that enhance performance or decision making skills and in both the Cups task and other decision making tasks. Specifically, future research in this area can aim to provide more clarity to the categorization of immediate versus delayed feedback effects, determine what influence gain/loss order may have outside the current studies, and explore how evaluative and behavioral outcomes can be linked.

In relation to the categorization of immediate versus delayed feedback effects, initial evidence found relatively distinct outcomes associated with the most frequent (trial-by-trial feedback) and least frequent (no trial feedback) trial feedback conditions, but the intermediary frequency conditions were not as distinguishable. However, the evidence for this distinction was somewhat speculative, and the interaction of trial feedback frequency with other variables (gain/loss order and task feedback frame) complicates the conclusions that can be directly drawn from these findings. This presents the opportunity for future testing to verify where the exact differences between trial feedback frequency intervals lie and to look for other potential moderators that could enhance or diminish the extent to which trial feedback frequency conditions produce effects more consistent with immediate or delayed feedback. If found, these additional distinctions can help identify the minimal amount of trial feedback required to produce positive results. In doing so, resources involved with providing feedback, such as time, effort, and money, could be conserved in real-world scenarios.

In relation to gain/loss order effects, without any context of prior results, further investigation is warranted to determine exactly why order effects consistently occurred in the current studies. It is possible, as suggested previously, that loss trials present outcome contingency information that is more difficult for people to utilize than gain trials. Starting out with loss trials, therefore, sets the stage for lower task performance in the initial round, and leaves more room for improvement in successive rounds. Future research should first address whether this explanation is relevant for the noted data patterns, and if this is the case, it will then be important to establish whether gain/loss order effects have any implications outside the laboratory tasks. For example, as many health behavior change programs are assessed across multiple time points and gain and loss information can both be present in these situations, there is potential to manipulate the order of gain/loss information across various time points to differentially promote maintenance of or improvement in target behaviors.

Perhaps the most critical future directions to pursue are why the correspondence between feedback manipulations and specific types of outcome measures occurs and whether there are ways that key aspects of feedback can be manipulated to affect the other type of outcomes to further enhance normatively correct decision making. With particular focus on why task feedback frame did not seem to influence task performance, the arguments made above suggest that trial outcome feedback may provide a better, more concrete sense of previous task performance; thus, the potential effect of the overall task frame manipulation may be diluted and not be enough to override the information provided by the trial outcome feedback. In addition, the task feedback frame manipulation may not have provided enough comparable information across frames to guide subsequent choices. As trial feedback (and the frequency with which it occurred) seemed to be more important for eliciting behavioral outcomes, more prominent framing of each piece of trial feedback to match the overall task feedback can help test whether feedback frame can have an influence on behavioral outcomes, as well as evaluative outcomes.

In the context of the Cups task, each trial is presented as an opportunity to win or lose quarters, so framing individual trial outcome feedback needs to be done with care to avoid unequal amounts of information across frames or contradicting information (e.g., framing the outcome of a gain trial as losing quarters). Framing trial feedback may be best accomplished using simple negation (e.g., “You earned/failed to earn 5 quarters”) rather than alternative frame (e.g., “You won/lost 5 quarters”) terms, and as both styles of framing were seen to affect performance evaluations, using either style should theoretically influence and improve task performance.

It may also be possible to manipulate the frequency of task feedback presentation in order to better connect this type of feedback with behavioral outcomes, but future research would be needed to determine exactly how this could be accomplished and whether it would exhibit more positive or negative influence on task performance. Summary scores (with no individual trial outcomes presented) may be able to be provided more frequently throughout the task as a way to manipulate the frequency of task feedback, but may show detrimental effects on task performance, as one’s ability to associate a choice with the context of the decision and the decision outcome decreases.

Therefore, the effectiveness of more frequent feedback on improving decision making skills may not extend to all types of feedback manipulation.

Furthermore, to address why the more relevant trial feedback presented in Study 2B did not contribute to resolving the disconnection between the key feedback manipulations and evaluative and behavioral outcomes, it is important to examine whether the trial feedback provided was relevant enough to elicit these outcomes. While likely more relevant than in the previous studies, the modification of trial feedback relevance did not alter behavioral outcomes, but did alter evaluative outcomes. This may indicate that the information chosen to enhance the relevance of the trial feedback did not represent the optimal provision of information. Accurate contingency information and descriptive terms (risk advantageous and risk disadvantageous) regarding the relative advantage or disadvantage of choosing the risky option over the riskless option were provided, but more simple terminology (e.g., good bet versus bad bet) may be more appropriate and interpretable for participants to use in guiding subsequent choices. Future research should strive to determine whether simpler, more interpretable terms (or other variations of feedback that may contribute to increased relevance) may help bridge the gap between the effects of relevant feedback manipulations on evaluative and behavioral outcomes. Identifying key components of feedback that contribute to its relevance and usefulness will serve to elucidate what makes the most effective feedback for producing desired behavioral outcomes in addition to evaluative outcomes.

Another potential reason that a disconnection between evaluative and behavioral outcomes exists may be due to a disparity between the level of specificity relating to performance evaluations and improved decision making behavior. The measures pertaining to performance evaluations were fairly global in their nature (e.g., “I feel that I can control my performance in the Cups task”, “I am confident I would show improvement on the Cups task if I were to do it again”), whereas the behavior of interest was fairly specific (e.g., increase risk advantageous choices, decrease risk disadvantageous choices). This mismatch in specificity may have tapped in to separate responses to the decision making task; indeed, it would be possible to simultaneously feel positive about one's performance and know there is room for improvement but not know how to change one's behavior to directly address that assessment. Addressing this issue in future research by modifying the performance evaluation measures to be more specific to the behavior change being assessed (e.g., "To improve on this task, I would need to change my risk taking tendencies", "The choices I made throughout the task generally had positive outcomes") may help produce more consistent and similar effects of feedback across evaluative and behavioral outcomes. Having more similar evaluative and behavioral outcomes produced by the same feedback manipulations may then allow for the anticipated meditational role of performance evaluations on the relationship between various feedback manipulations and improved task performance to be found.

Finally, future research may also want to assess how different moderators may affect feedback framing effects to enhance the likelihood of producing behavioral outcomes. Particular focus should be placed on scouring the previous literature to definitively determine under what circumstances positively versus negatively framed performance feedback leads to more beneficial outcomes and how those moderating variables may relate to performance feedback for decision making tasks. As a specific example, feedback framing effects can be moderated by regulatory focus, such that positively framed feedback leads to more positive evaluative and behavioral outcomes for promotion-based tasks and negatively framed feedback leads to more positive outcomes for prevention-based tasks (Van Dijk & Kluger, 2011); therefore, one potential moderator to include in future studies may be presenting an explicit, promotion-focused goal to win the most money in the Cups task. The design and instruction of the Cups task are for a promotion-focused task, but this may not be explicit to participants, so making that designation more clear could potentially help the task feedback frame manipulation produce more improved task performance, in addition to the positive performance evaluations already seen.

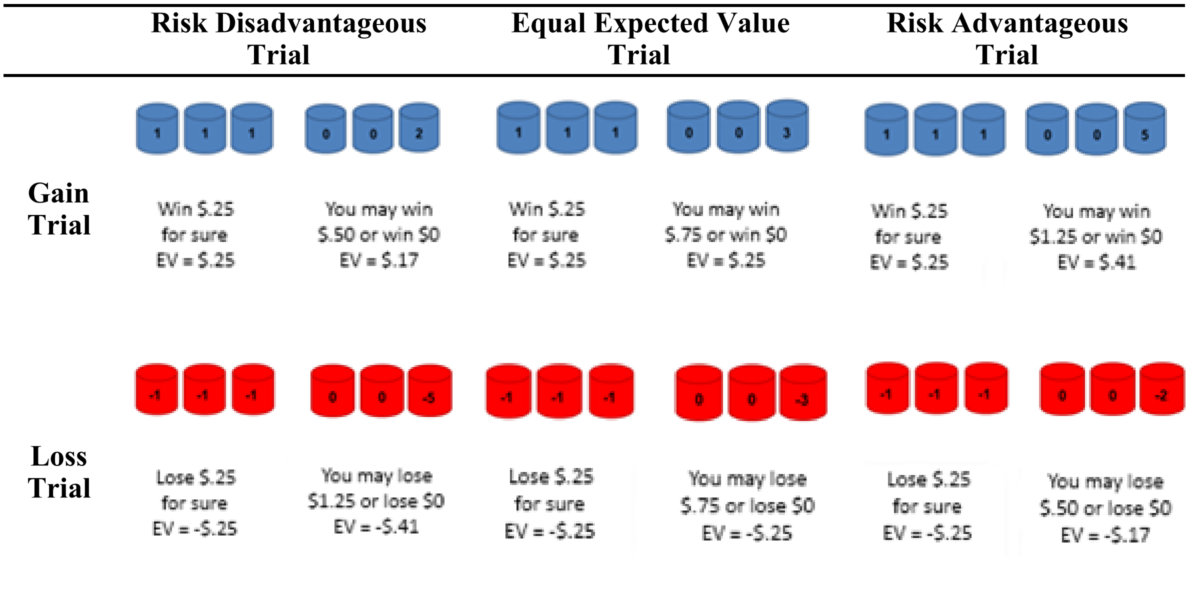
Conclusions

When working to improve decision making skills, feedback is often necessary to provide guidance for future decisions. The relationship between feedback and decision making outcomes is complex because not all types of feedback are equally effective, so it is important to consider what type of feedback will be best for producing beneficial outcomes. Through the current studies, it was found that different feedback manipulations influenced different types of outcomes, particularly that trial feedback frequency was more influential for task performance outcomes, while task feedback frame was more influential for performance evaluation outcomes. Knowing that attitudes and evaluations do not always coincide with behavior, identifying the specific outcome researchers, educators, and practitioners wish to achieve (evaluative versus behavioral) should heavily influence what type, what presentation, and how much of feedback to highlight in order to achieve desired outcomes.

APPENDIX

STUDY DESIGN AND MATERIALS

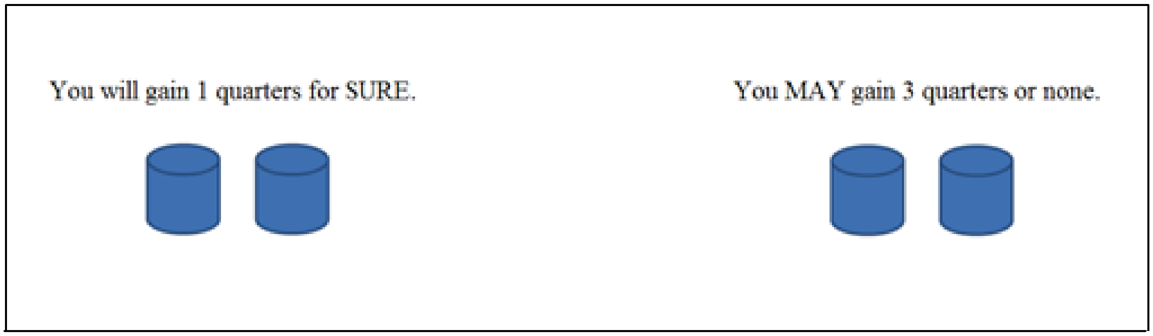
Figure A1 Samples of Trial Types in the Cups Task



Note: In each case, the riskless side is depicted on the left and the risky side is depicted on the right. In the experiments, these were counterbalanced over trials.

Figure A2 Cups Task Trial Example

(a) Gain Trial



(b) Loss Trial

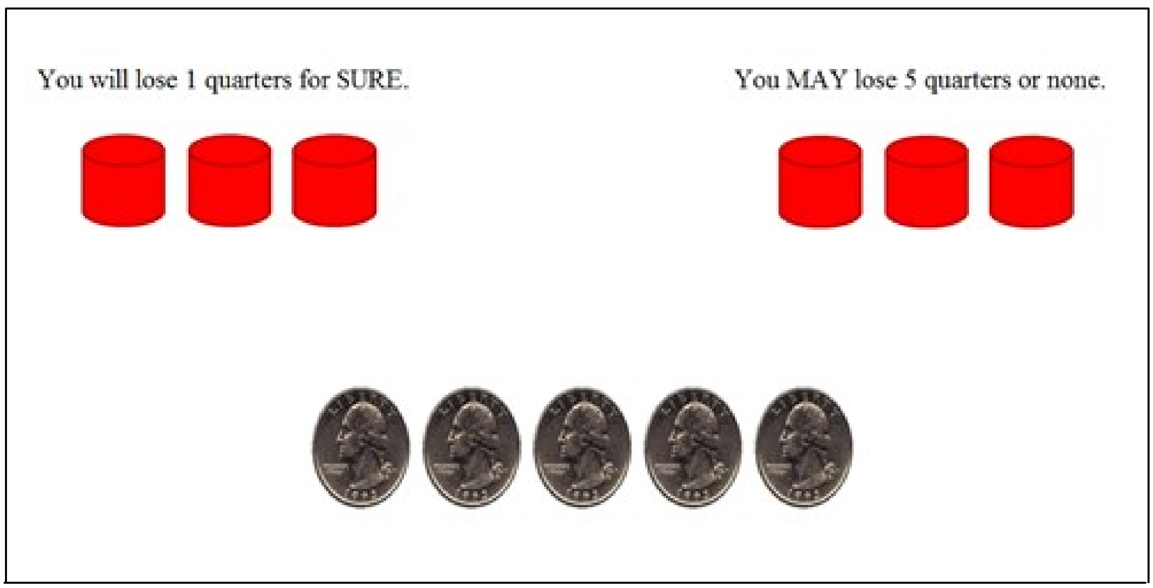
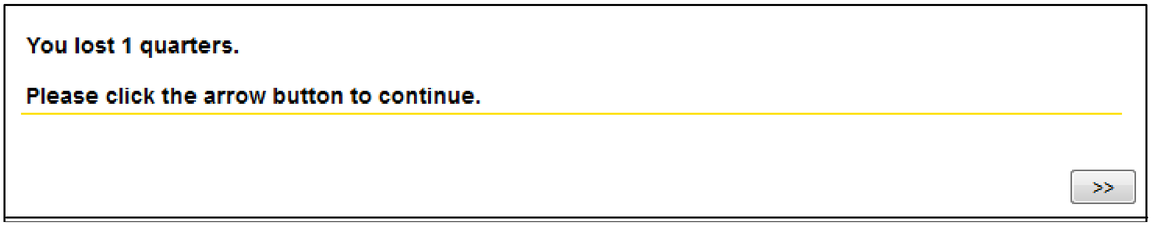
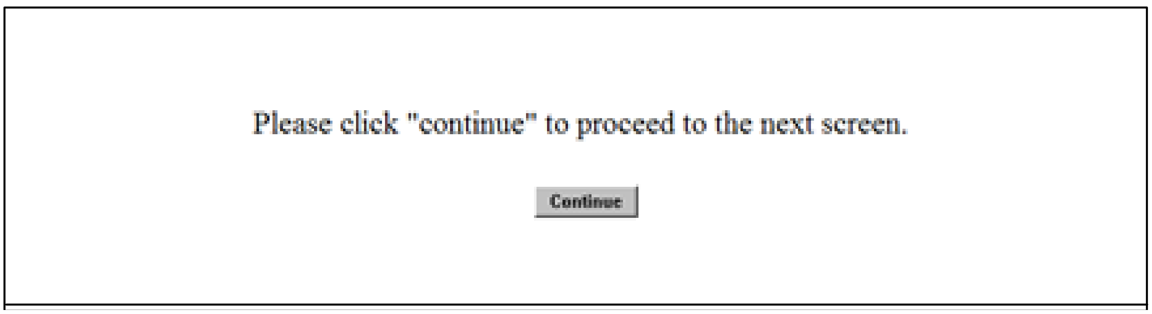


Figure A3 Trial Feedback Frequency Presentation Example

(a) Trial-by-Trial Feedback (Standard Cups Task Feedback)



(b) No Trial Feedback



(c) 3 Trial Feedback

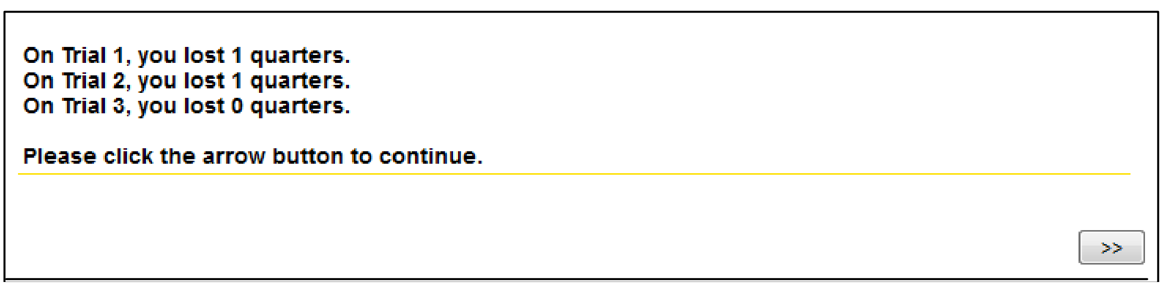


Figure A3 Continued.

(d) 9 Trial Feedback

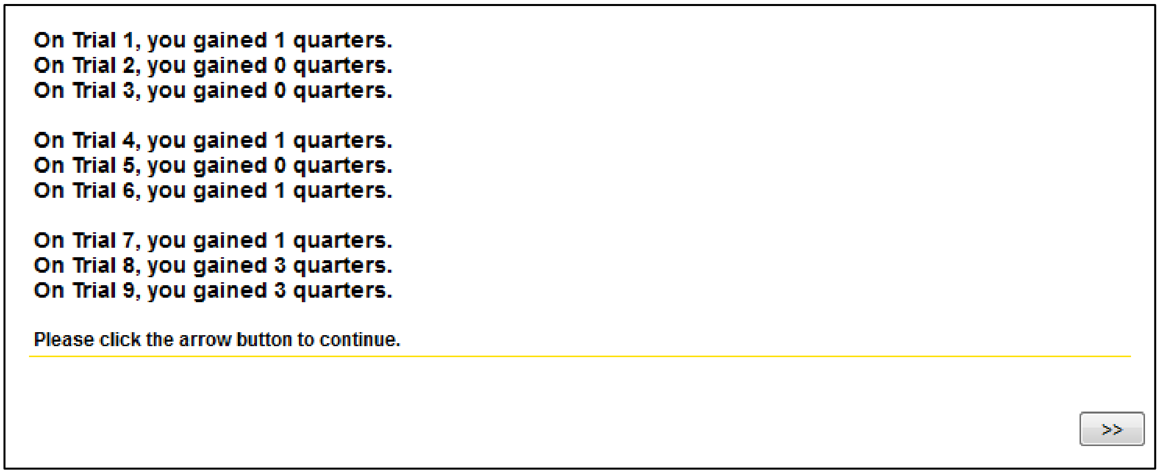


Figure A3 Continued.

(e) Trial Block Feedback

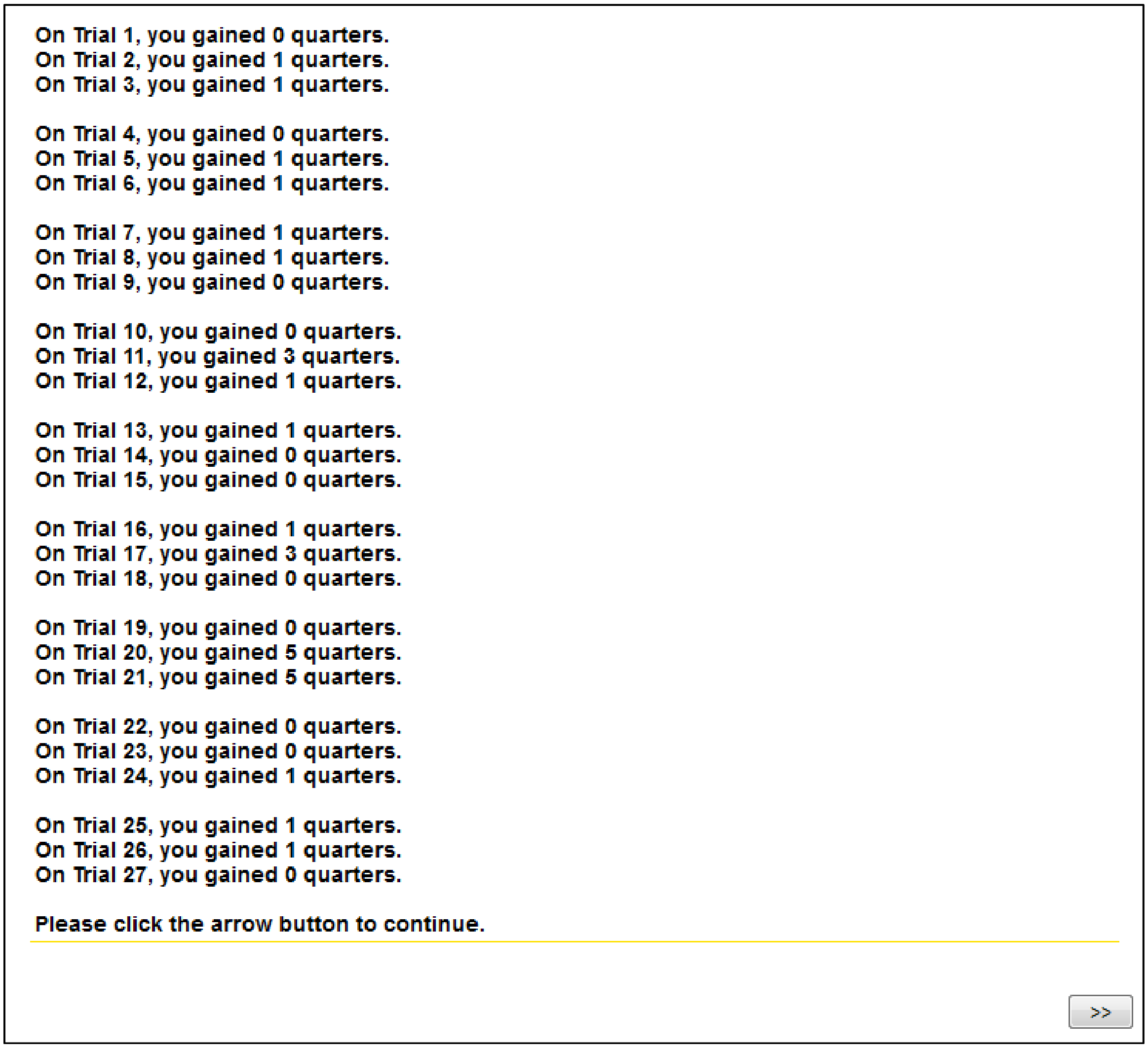


Table A1 Study Designs

	Task Feedback Frame	**Trial Feedback Frequency**	**Feedback Relevance**
Study 1A	Earned vs. Failed to Earn	No Trial Feedback vs. Trial-by-Trial	X
Study 1B	Won vs. Lost	No Trial Feedback vs. Trial-by-Trial	X
Study 2A	Won vs. Lost	No Trial Feedback vs. Trial Block vs. 9 Trial vs. 3 Trial vs. Trial-by-Trial	X
Study 2B	Won vs. Lost	No Trial Feedback vs. Trial Block vs. 9 Trial vs. 3 Trial vs. Trial-by-Trial	"On Trial 1, there were 3 cups and the chance to win 5 quarters, and this represented a risk advantageous trial. You chose the risky side and won 5 quarters."

REFERENCES

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260-267. doi: 10.1037/0022-0663.80.3.260

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology, 62*(2), 148-156. doi:10.1037/h0030766

Armstrong, K., Schwartz, J. S., Fitzgerald, G., Putt, M., & Ubel, P. A. (2002). Effect of framing as gain versus loss on understanding and hypothetical treatment choices: Survival and mortality curves. *Medical Decision Making, 22,* 76-83. Doi: 10.1177/0272989X0202200108

Bangert-Drowns, R. L., Kulik, C. L., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213-237.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16,* 215-233.

Brevers, D., Bechara, A., Cleeremans, A., Kornreich, C., Verbanck, P., & Noel, X. (in press). Impaired decision-making under risk in individuals with alcohol dependence. *Alcoholism: Clinical and Experimental Research*. doi: 10.1111/acer.12447

Brug, J., Glanz, K., van Assema, P., Kok, G., & van Breukelen, G. J. P. (1998). The impact of computer tailored feedback and iterative feedback on fat, fruit, and vegetable intake. *Health Education Behavior*, 25, 357-371.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*(2), 290-298. doi: 10.1037/a0031026

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*(4), 474-482. doi: 10.1037/0022-0663.79.4.474

Campbell, M.K., DeVellis, B.M., Strecher, V.J., Ammerman, A.S., DeVellis, R.F., & Sandler, R.S. (1994). Improving dietary behavior: the effectiveness of tailored messages in primary care settings. *American Journal of Public Health, 84*(5), 783-787.

Choe, E. K., Lee, B., Munson, S., Pratt, W., & Kientz, J. A. (2013). Persuasive performance feedback: The effect of framing on self-efficacy. *AMIA Annual Symposium Proceedings*, 825-833.

Clariana, R. B., Wagner, D., & Roher Murphy, L. C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development, 48*(3), 5-21.

Contento, I., Black, G.I., Bronner, Y.L., Lytle, L.A., Maloney, S.K., Olson, C.M., & Swadener, S.S. (1995). The effectiveness of nutrition education and implications for nutrition education policy, programs, and research: A review of research. *Journal of Nutritional Education, 27,* 277-418.

Culpepper, S. A. (in press). The reliability of linear gain scores as measures of student growth at the classroom level in the presence of measurement bias and student tracking. *Applied Psychological Measurement*. doi: 10.1177/0146621614534763

DiClemente, C., Marinilli, A., Singh, M., & Bellino, L. (2001). The role of feedback in the process of health behavior change. *American Journal of Health Behavior, 25*(3), 217-227.

Dihoff, R., Brosvic, G., & Epstein, M. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record, 53*, 533-548.

Epstein, M. L, & Brosvic, G. M. (2002). Immediate Feedback Assessment Technique: Multiple-choice test that "behaves" like an essay examination. *Psychological Reports, 90*(1), 226.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate Feedback Assessment Technique promotes learning and corrects inaccurate first responses. *Psychological Record, 52*(2), 187-201.

Epstein, S. (1979). The stability of behavior: On predicting most of the people most of the time. *Journal of Personality and Social Psychology, 37*(7), 1097-1126. doi: 10.1037/0022-3514.37.7.1097

Erev, I., Luria, A., & Erev, A. (2006). On the effect of immediate feedback. In Eshet-Alkalai, Y., Caspi, A., & Yair, Y. (Eds.), *Learning in the Technological Era. Proceedings of the Chais Conference, March* (Vol. 1, pp. 26-30).

Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory, 18,* 335-350. doi: 10.1080/09658211003652491

Fedor, D. B. (1991). Recipient responses to performance feedback: A proposed model and its implications. *Research in Personnel and Human Resources Management, 9*, 73-120.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY: McGraw-Hill Book Company.

Fox, C. R., & Hadar, L. (2006). “Decisions from experience” = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber, & Erev (2004). *Judgment and Decision Making, 1*(2), 159-161.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25-42.

Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine, 43*(1), 101-116. doi: 10.1007/s12160-011-9308-7

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92*(2), 316-330. doi: 10.1037/0022-0663.92.2.316

Hattie, J. A., Biggs, J., & Purdie, N. (1996). Effects of learning skills intervention on student learning: A meta-analysis. *Review of Research in Education, 66,* 99-136.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi: 10.3102/003465430298487

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*(5), 493-518. doi: 10.1002/bdm.598

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15(8),* 534-539.

Howard, K., & Salkeld, G. (2009). Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. *Value in Health*, *12*(2), 354-363.

Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology: An International Review, 49,* 550-565. doi: 10.1111/1464-0597.00031

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136-153. doi: 10.1037/0033-295X.93.2.136

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.

Kluger, A. N., & DeNisi, A. (1996). The effect of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284. doi: 10.1037/0033-2909.119.2.254

Kluger, A.N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions In Psychological Science, 7,* 67-72. doi: 10.1111/1467-8721.ep10772989

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes, 75*(1), 23-55. doi: 10.1006/obhd.1998.2781

Kulik, J.A. & Kulik, C.C. (1988). Timing of feedback and verbal learning. *Review of Educational Research,58*(1), 79-97.

Kuvaas, B., & Selart, M. (2004). Effects of attribute framing on cognitive processing and evaluation. *Organizational Behavior and Human Decision Processes, 95*(2), 198-207. doi: 10.1016/j.obhdp.2004.08.001

Lauriola, M., Levin, I. P., Hart, S. S. (2007). Common and distinct factors in decision making under ambiguity and risk: A psychometric study of individual differences. *Organizational Behavior and Human Decision Processes, 104*(2), 130-149. doi: 10.1016/j.obhdp.2007.04.001

Ledgerwood, A., & Boydstun, A. E. (2014). Sticky prospects: Loss frames are cognitively stickier than gain frames. *Journal of Experimental Psychology: General, 143,* 376-385. doi: 10.1037/a0032310

Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes, 116*(2), 286-295. doi: 10.1016/j.obhdp.2011.05.001

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3*, 113-132. doi:10.1016/0959-4752(93)90011-n

Levin, I. P., & Bossard, E. A. (2011). Comparison of children and adult's learning from experience in a descriptive decision making task. Poster presented at the annual meeting of the Society for Judgment and Decision Making, Seattle, WA.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15,* 374-378. doi: 10.1086/209174

Levin, I. P., Hart, S. S., Weller, J. A., & Harshman, L. A. (2007a). Stability of choices in a risky decision making task: A 3-year longitudinal study with children and adults. *Journal of Behavioral Decision Making, 20,* 241-252. doi: 10.1002/bdm.552

Levin, I. P., Johnson, R. D., Russo, C. P., & Deldin, P. J. (1985). Framing effects in judgment tasks with varying amounts of information. *Organizational Behavior and Human Decision Processes*, 36, 362–377. doi: 10.1016/0749-5978(85)90005-6

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76,* 149-188. doi: 10.1006/obhd.1998.2804

Levin, I. P., Weller, J. A., Pederson, A. A., & Harshman, L. A. (2007b). Age-related differences in adaptive decision making: Sensitivity to expected value in risky choice. *Judgment and Decision Making, 2,* 225-233.

Lhyle, K. G., & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology, 79,* 320-322.

Linville, P. W., Fischer, G. W., & Fischhoff, B. (1993). AIDS risk perceptions and decision biases. In J. B. Pryor & G. D. Reeder (Eds.), *The social psychology of HIV infection* (pp. 5–38). Hillsdale, NJ: Lawrence Erlbaum.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37-44. doi: 10.1037/t05138-000

Loke, W. H., & Tan, K. F. (1992). Effects of framing and missing information in expert and novice judgment. *Bulletin of the Psychonomic Society*, 30, 187–190.

Maheswaran, D., & Meyers-Levy, J. (1990). The influence of message framing and issue involvement. *Journal of Marketing Research, 27*(3), 361-367. doi: 10.2307/3172593

McClenaghan, B. A., & Ward, D. S. (1987). Health and physical education. In Charles A. Maher & Susan G. Forman (Eds.). *A behavioral approach to education of children and youth (School Psychology Series)* (pp. 131-151). Hillsdale, NJ: L. Erlbaum Association.

McKenzie, C.R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, *10*(3), 596-602. doi: 10.3758/BF03196520

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96, 165*-173. doi: 10.1037/0022-0663.96.1.165

Nunnally, J.C. (1978). Psychometric Theory (2nd ed.). New York: McGraw Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd Ed.). New York: McGraw-Hill.

Nygren, T. E. (1997). Framing of task performance strategies: Effects on performance in a multiattribute dynamic decision making environment. *Human Factors, 39*(3), 425-437. doi: 10.1518/001872097778827115

Oenema, A., & Brug, J. (2003). Feedback strategies to raise awareness of personal dietary intake: Results of a randomized controlled trial. *Preventative Medicine, 36*, 429-439.

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making, 18*(1), 1-27. doi: 10.1002/bdm.481

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3-8. doi: 10.1037/0278-7393.31.1.3

Peters, E., Hibbard, J. H., Slovic, P., & Dieckmann, N. F. (2007). Numeracy skill and the communication, comprehension, and use of risk and benefit information. *Health Affairs, 26*, 741-748.

Piñon, A., & Gambara, H. (2005). A meta-analytic review of framing effect: Risky, attribute, and goal framing. *Psicothema, 17*(2), 325-331.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*, 1-14. doi: 10.1002/bdm.681

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds.). (1991). Measures of personality and social psychological attitudes. San Diego: Academic Press.

Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate health behavior: The role of message framing, *Psychological Bulletin, 121,* 3-19. doi: 10.1037/0033-2909.121.1.3

Schmidt, R.A., Young, D.E., Swinnen, S. & Shapiro, D.C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 352-359.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67-85. doi: 10.1037/0033-2909.110.1.67

Tversky, A., &Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. doi: 10.1126/science.7455683

Tverksy, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science, 20*(4), 473-479. doi: 10.1111/j.1467-9280.2009.02319.x

Van Dijk, D. & Kluger, A. N. (2011). Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective. *Journal of Organizational Behavior, 32,* 1084–1105. doi: 10.1002/job.725

Vollmeyer, R. & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction, 10*(4), 293-309. doi: 10.1016/S0959-4752(99)00031-6

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction, 15*(6), 589-602. doi: 10.1016/j.learninstruc.2005.08.001

Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111*(2), 430-445. doi: 10.1037/0033-295X.111.2.430

Weller, J. A., Levin, I. P., & Denburg, N. L. (2011). Trajectory of risky decision making for potential gains and losses from ages 5 to 85. *Journal of Behavioral Decision Making, 24*(4), 331-344. doi: 10.1002/bdm.690

Weller, J. A., Levin, I. P., Rose, J. A., & Bossard, E. A. (2012). Assessment of decision-making competence in preadolescence. *Journal of Behavioral Decision Making, 25(4),* 414-426. doi: 10.1002/bdm.744

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A rasch analysis approach. *Journal of Behavioral Decision Making, 26,* 198-212. doi: 10.1002/bdm.1751

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, *20*(1), 59-69. doi: 10.1177/014662169602000106

Wyer, R. S., & Srull, T. K. (1989). *Memory and social cognition in its social content.* Hillsdale, NJ: Lawrence Erlbaum.

Xue, G., Lu, Z., Levin, I. P., & Bechara, A. (2011). An fMRI study of risk-taking following wins and losses: Implications for the gambler's fallacy? *Human Brain Mapping, 32*(2), 271-281. doi: 10.1002/hbm.21015

Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making, 19*(1), 1-16. doi: 10.1002/bdm.509

Yechiam, E., Druyan, M., & Ert, E. (2008). Observing others' behavior and risk taking in decisions from experience. *Judgment and Decision Making, 3*(7), 493-500.

Zahorik, J. A. (1987). Reacting. In M. J. Dunkin (Ed.), *International encyclopedia of teaching and teacher education* (pp. 416-423). Oxford: Pergamon Press.