Development of a Framework to Describe the Association Between Hypothetical and Real-World Decisions: A Systematic Concept Review and Cognitive Interviews with Decision Aid Developers

---

Tavis Hayes, B.A. (Hons), GradD.

Supervised by

Jamie Brehaut, PhD.

and

Doug Coyle, PhD.

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in Partial fulfillment of the requirements for the M.Sc. degree in Epidemiology

Submitted to:

School of Epidemiology, Public Health, and Preventive Medicine

Faculty of Medicine

University of Ottawa

ABSTRACT

Despite recommendations, health services interventions often lack mechanistic evidence. One way of acquiring such evidence is through the use of hypothetical settings. However, there are knowledge gaps about the factors that affect whether hypothetical decisions will correlate to real-world ones. This thesis involves a systematic concept review to identify the factors important to the hypothetical-real-world decision relationship and a case study of the relevance of these factors to the decision aid development context. The 42 identified factors were grouped into 5 categories: Personal Characteristics, Presentation Characteristics, Cognitive Factors, Motivation, and Participant Characteristics. Through cognitive interviews with decision aid developers we found that the interviewees had generally not considered many of the factors thought to affect the external validity of hypothetical data. This thesis contributes a framework that highlights factors for investigators to consider when designing studies in hypothetical settings and that can serve as a foundation for future research.

ACKNOWLEDGEMENTS

I have been very privileged to have been surrounded by an amazing group of people while working on this thesis. I owe them all a sincere "thank you." Without them, this work would not have been possible.

Thank you to my Thesis Advisory Committee: Jamie Brehaut, Doug Coyle, and Ian Graham. I appreciate every piece of feedback and support that you have given me over the past two years. I could not have asked for better mentors.

Thank you to Kelly Carroll for your help conducting interviews, for your edits, and for all of the advice you gave me along the way.

Thank you to my family for always encouraging me and for reminding me that's it's important to enjoy the process.

Thank you to all of my friends. I hope to be able to offer you the same support and encouragement that you have given me.

I owe a special thank you to my best friend and wife Mélanie Hayes. Thank you for your love and support. You remind me every day what being a good partner is about.

CONTENTS

Abstract ii

Acknowledgements iii

List of Tables vii

List of Figures vii

Chapter 1: Introduction 1

1.1 The Relationship between Hypothetical and Real Decisions 2

1.2 Exploring the Hypothetical-Real Relationship 5

1.3 The Hypothetical-Real Relationship in the Health Context 9

1.4 The Relationship between Hypothetical and Real Decisions in the Decision Aid Context 9

1.5 Objectives and Research Question 11

Chapter 2: Methods 13

2.1 Systematic Concept Review: 13

2.1.1 Objectives: 13

2.1.2 Search Strategy: 14

2.1.3 Eligibility Criteria 15

2.1.4 Study Selection 15

2.1.5 Data Collection 16

2.1.6 Data Analysis 18

2.2 Pilot Survey Development 19

2.2.1 Survey Rationale....19

2.2.2 Pilot Testing....20

2.2.3 Cognitive Interviews....20

2.2.4 Cognitive Interview Administration....21

2.2.5 Data Analysis....23

Chapter 3: Results....24

3.1 Systematic Concept Review....24

3.1.1 Studies Included....24

3.1.2 Study Methods....25

3.1.3 Factors Identified....31

3.2 Cognitive Interviews....67

3.2.1 Themes elicited from Interviews: External validity issues....67

3.2.2 Relevant Factors....68

3.2.3 Factors with Unclear Relevance....69

3.2.4 Irrelevant Factors....70

Chapter 4: Discussion....72

4.1 Overview....72

4.2 Framework....74

4.3 Systematic Concept Review....77

4.3.1 The Role of Financial Incentives in Modeling health care decisions....77

4.3.2 Relevance of the Factors to The decision Aid Development Context 78

4.3.3 Summary of Factors 82

4.3.4 Participant Characteristics 82

4.3.5 Personal Characteristics 83

4.3.6 Presentation Characteristics 84

4.3.7 Motivation Factors 86

4.3.8 Cognitive Factors 88

4.4 Limitations & Strengths 90

4.5 Contributions 92

4.6 Future Directions 94

4.7 Overall Conclusions 96

References 99

Appendix 1: Search Strategy for PsycINFO 118

Appendix 2: Survey Methods 119

Sample Frame 119

Contact Process 120

Incentives 122

Appendix 3: Draft Survey 123

List of Tables

Table 1 Examples of Research Methods used in the studies included in the systematic concept review, organized by field of research.....26

Table 2 Definitions of Personal Characteristics factors (n=9).....35

Table 3 Definitions of Presentation Characteristics factors (n=8).....43

Table 4 Definitions of the Cognitive factors (n=17).....54

Table 5 Definitions of Motivation factors (n=7).....62

Table 6 Definition of Participant Characteristics Factor (n=1).....64

List of Figures

Figure 1 Study Selection Process.....29

Figure 2 Proportion of included studies by field (n=59).....30

Figure 3 Included studies by date of publication (n=59).....30

Figure 4 Fields from which each factor was identified. Cells are green if a factor was identified by at least one study from a particular field.....65

Figure 5 Framework depicting the 42 factors thought to affect the consistency between hypothetical and real-world decisions, grouped into 5 categories.....74

CHAPTER 1: INTRODUCTION

Much of health research is supported by a mountain of data generated through sources and settings that are not identical to real-world applications. For example, while determinations about drug effectiveness ultimately fall on late stage randomized controlled trials (RCT) because of their design features, it is easy to forget about the amount of preclinical and pilot work that enabled the final trial.[1] Without this work, which inevitably involves laboratory settings, animal models, healthy volunteers, and other components that differ from the real-world application, investigators would not have the requisite supportive mechanistic evidence to ensure that the trial is appropriate. [1,2]

Health services research is a multidisciplinary field of science that studies how personal, organizational, technological, and systemic factors affect access to, quality, and cost of health-care.[3] Clinical epidemiology contributes to health services research by providing data about diagnosis, prognosis, and treatment of human diseases.[4] One of the major differences between the field of pharmaceutics and the field of health services research is in the extent to which they are supported by systematic study of mechanisms involving hypothetical settings and participants. Pharmaceutics research is supported by systematic, iterative preclinical and early-phase studies that evaluate mechanisms of action, rule out alternative hypotheses, identify mediating factors, etc. This often involves testing new products in vitro or with animal models, phase I studies in healthy humans to ascertain the maximally tolerated dose and drug activity, and phase II studies to evaluate whether the drug has an effect in specific subgroups of patients in order to obtain an estimate of the probability of success of a full-scale trial.[5] This type of experimentation leading to a full scale trial has not been the norm in health services research. For example, a systematic review by Colquhoun et al[6] found

that in audit and feedback trials there was sparse use of theory to ensure that new interventions were optimized for success.

The UK Medical Research Council (MRC) identifies the lack of rigorous mechanistic experiments in health services research as a problem and stresses the need for more work.[7] The framework outlines different stages of complex health intervention development: Theory, Modeling, Exploratory Trial, Definitive Trial, and Long term implementation.[7] The UK MRC framework recommends that investigators ensure that the mechanisms underlying their interventions have been thoroughly examined in order to optimize the chances of successful new interventions.[7] This work can take place in many different contexts, involving varying degrees of 'distance' from the real-world. The framework identifies a Modelling stage meant to determine the components of the intervention and the mechanisms through which they will produce outcomes. At this Modelling stage, methodologies may or may not include real patients and settings, as these will not always be necessary to explore the important mechanistic components of an intervention.[7] The current work seeks to understand the conditions under which this important mechanistic work can most effectively be conducted to develop a framework that describes the association between hypothetical and real-world decisions.

1.1 The Relationship between Hypothetical and Real Decisions

In health services research, investigators cannot use animal or in vitro models to test new interventions. Most preliminary work must be carried out on people. These research participants might be healthy volunteers with no relationship to the health condition of the target population, or they could be patients with similar but not identical health issues. The environment in which the

preliminary work is conducted might be very similar or very different from where the intervention will actually be used. We can call these studies that somehow differ from the real-world application of an intervention "hypothetical" studies. In the process of drug development, the ways in which animals can serve to model humans for hypothetical studies is well understood.[5] Conversely, in health services research, the circumstances under which data from hypothetical decision makers and circumstances can predict real-world performance are *not* well understood. In order to be able to provide mechanistic evidence to help us understand the mechanism through which a health service intervention affects a targeted outcome, we need to understand the circumstances under which we expect hypothetical data to generalize to the real-world.

Because the relationship between hypothetical and real-world data has not been sufficiently explored in health services research, researchers hold implicit, untested beliefs about the factors affecting hypothetical data. Many systematic reviews exclude any studies involving hypothetical choices.[8] While the reasoning behind these exclusions is often not explicitly presented, hypothetical data are likely excluded because of an implicit belief that such data cannot reliably predict real-world decisions. Other reviews may include simulated or hypothetical studies while noting that data from these studies are less realistic than data from real-world studies. Investigators have been cautioned to avoid conducting studies that assess the informed consent process under hypothetical conditions, where participants are asked to evaluate consent documents for trials in which they will not ultimately participate, because these studies can be "unrealistic".[9]

Different disciplines have approached the problem of eliciting data from hypothetical settings in different ways, revealing underlying assumptions about how close to the real-world the data must be to be valid. Many laboratory-based studies involve undergraduates required to imagine how they would behave in a real setting. In contrast, designers of high-tech simulation rooms for

training medical students go to great lengths to ensure a high level of similarity between the hypothetical setting being used and what the future physicians will actually experience in the real-world. These hypothetical scenarios differ in the amount of resources they devote to emulating the real-world, but they all intend to generate data that will generalize to real-world situations.

By conducting research in hypothetical settings, investigators are able to obtain data that would otherwise be infeasible to collect. Some types of trials would be unethical to conduct with real life patients. In the case of health services research, it might be unethical to expose real patients to a new tool, such as tools to provide information about treatment options, that could be potentially harmful by affecting their decision in ways that were not intended. In this case, we would need to collect preliminary data by involving a less vulnerable population, perhaps clinicians, other researchers or volunteer participants, in the assessment of the new tool before testing it in a real-life setting. There might also be a limited pool of potential participants, as would be the case if we are studying an intervention for a rare disease. In this case, we would need to employ hypothetical aspects in order to ensure that the new intervention has been evaluated by a sufficient number of people – for example, we might seek out the assistance of people who have the target condition but who are not actively involved in the situation that the tool was designed to assist. It is unclear whether these factors have implications for the external validity of the data collected.

It is important to distinguish between internal and external validity when considering the purpose of research in hypothetical settings. Some pilot studies might focus only on issues that are related to the internal validity of the study, (i.e. whether the results of the study are valid for the study population).[10]More pertinent to the discussion of generalizability of hypothetical decisions however, are those pilot studies that also seek to assess the external validity of their study, (i.e. whether the results can be generalized to broader populations).

When designing a study in a hypothetical setting, many factors have the potential to affect the study's external validity.[5] In the undergraduate example above, there is a (usually implicit) assumption that the setting where a decision takes place, and the differences between undergraduate students and the target population, are not salient enough to bias responses away from those that would occur in the real-world. In the medical simulation setting, it is considered important for the study location to closely resemble the real-world context where the decision making will take place. We can improve the process of collecting data from hypothetical settings by clarifying which factors are important and necessary to consider in order for data to be internally and externally valid.

1.2 Exploring the Hypothetical-Real Relationship

While the factors affecting whether decisions made under hypothetical conditions reflect decisions made under real-world conditions have not been assessed in health services research, there are several fields where hypothetical settings are commonly used in research and where researchers are actively seeking to understand how to most effectively collect and employ hypothetical data. Although the specific research questions and objectives differ between such fields as behavioral economics, moral decision making, and social psychology, these areas all commonly employ hypothetical components in their research and have thought about when the use of these components is appropriate.

Behavioral economics is a significant contributor to the literature related to understanding the distinction between hypothetical and real. One of the research areas within the field of behavioral economics involves studying how people make important, real-world monetary decisions. However, it is expensive and infeasible to conduct studies involving large sums of real money. There is therefore great interest in understanding whether experimental conditions can be created to mimic

real-world decisions. Often, these experiments take the form of assessing how people value hypothetical currency and whether smaller amounts of real money can be used to approximate decisions involving larger sums.

Much of the economics literature addressing the problem of the relationship between hypothetical and real decisions examines how people value various rewards, generally discussed in terms of 'discounting'. Delay discounting studies involve asking people's preference between an immediate and delayed reward. The investigator then adjusts the value of the immediate reward and repeats the process until the immediate and delayed rewards are of equal subjective value; this provides the investigators with an indication of the rate of discounting of the delayed reward.[11] A similar procedure can be employed in order to determine rates of probability discounting. This involves comparing different gambles, a lower-risk/lower-reward option to a higher-risk/higher-reward option, and increasing the probability of the higher-reward option to determine when participants are willing to accept the risk in order to obtain potentially greater rewards.[12] Both types of studies can help investigators understand whether the rate of discounting is predictably associated with the realness of the situation and whether there are modifiable factors that affect the rate of discounting in order to better predict real-world decisions.

The field of moral reasoning is another area where the hypothetical-real association has been explored. Questions in this area of literature relate to factors that affect important, moral decisions, often involving participants weighing different types of harm to others; often these questions cannot ethically be implemented in the real-world. As a result, it is important for investigators to understand whether the way people respond to hypothetical moral dilemmas is indicative of how they would behave in real life.

Researchers in this area have developed approaches for studying moral dilemmas experimentally. One approach involves the use of simulated ethical challenges. A commonly explored question is the trolley problem. In this dilemma, people are tasked with deciding whether to allow a trolley to follow a track which will cause it to kill five people, or flipping a switch to divert it onto a different track which will only result in the death of one person.[13] Since this decision can obviously not be carried out in real life with the real outcomes, various approaches have been employed to create experimental models that 'simulate' real life in different ways. Some involve simply presenting the task in a written format, which may not elicit the same emotional impact as real life moral decision.[14] Other efforts involve investigators asking participants to recall how they behaved in past moral decisions that they actually experienced, an approach that has been argued to be affected by recall bias.[15] Other design studies replace the risk of death in the trolley problem with a decision about providing a donation of real money to orphans. This hypothetical scenario was designed to elicit the same sort of reaction as the life-or-death trolley problem. [13] Still other investigators have employed virtual reality to make the decision making situation more realistic.[14] Virtual reality simulations of the trolley problem could be more emotionally evocative than simply reading textual descriptions of the same problem, and thereby elicit decisions that are more similar to real life.[14]

This literature has identified several factors related to predicting whether hypothetical moral reasoning decisions will be consistent with real-world ones. For example, level of moral reasoning employed when resolving a hypothetical vs. a real moral dilemma is one such factor. Level of Moral Reasoning is often measured using Kohlberg's Stages of Moral Development. This involves three levels: The 'pre-conventional' level where the decision-maker reasons according to the punishment or rewards that will result from their decision; the 'conventional' level where the decision-maker reasons according to the rules and expectations of the family, group, or nation; and the 'post-

conventional level' where the decision-maker reasons according to moral values and principles.[16] Investigators find that hypothetical dilemmas often involve higher levels of moral reasoning (i.e. closer to the post-conventional level) compared to participants' reasoning about dilemmas that they had experienced in real-life. [15] Another factor that can affect moral reasoning is the affective response caused by a real-life decision. The affective response elicited by contextually detailed virtual reality experiences caused participants to reason about the moral dilemma differently than they did when simply reading the problem, suggesting that affective responses may decrease consistency between hypothetical and real-world moral decisions.[14]

Social psychology is a third field of research that has explored the question of how hypothetical decisions apply to real life. This type of research typically involves comparing how people state they would interact with others in hypothetical dilemmas to real life dealings with other people. For example, FeldmanHall et al asked participants to read textual scenarios asking them how much money they would be willing to lose to prevent someone else from receiving an electric shock. After assessing how participants responded in the hypothetical textual scenario the participants were paid with real money and were told that they had the option of spending some of this money to prevent their partner from receiving a painful electric shock. If the participant was not willing to spend the required amount to prevent the painful stimulus, they had to watch a video of their partner receiving an electric shock. The participants believed their partner's pain was real, but they were actually watching actors simulate the painful experience. The authors observed that the potential for real gain causes people to be more willing to allow others to experience harm in real life.[17] This suggests that when there is a potential for gain in the real-world setting that is not correctly modeled by the hypothetical setting, it may decrease the correlation between hypothetical and real-world decisions.

1.3 The Hypothetical-Real Relationship in the Health Context

Relatively little research has focused on factors affecting the relationship between hypothetical and real-world health care decisions. There is good evidence to suggest that people are poor at predicting their feelings about future health states and how they would behave if they were to become ill. People making predictions about their feelings towards future health states are making a hypothetical decision, while people stating their feelings about an existing health state are describing their real life experiences.[18] In other settings, it is known that real decisions have an important emotional impact that is not always modeled in hypothetical settings.[19,20,21,22,23] This may be true in health services research as well, but the factors affecting hypothetical health decisions have not been explored. Until the factors important to ensuring a close relationship between hypothetical and real health decisions are clarified, health service researchers will continue to be guided by implicit assumptions about these factors, or simply fail to conduct this important mechanistic work.

1.4 The Relationship between Hypothetical and Real Decisions in the Decision Aid Context

We have chosen to focus on patient decision aid development as one of the many types of decisions where this issue appears to be relevant. Decision aids are decision making tools for helping patients make difficult health care decisions. These types of decisions often have no clear, clinically correct, course of action and instead must be made on the basis of a consultation between patient and provider, with patient preferences for different outcomes as an important factor in the decision.[24]

Imagine that you were developing a decision aid for prostate cancer treatment in men over age 50. Would obtaining the design team's impressions of the tool be sufficient before introducing it to real patients, or would this lead to some unanticipated harm? If we decide to test the tool with proxy patients, who would serve? Would a prostate cancer survivor make decisions the same way or be able to imagine making the decision as someone who is being actively treated for prostate cancer? When we begin testing the decision aid with real patients, do we need to test the tool in the same location as where it will eventually be used? These considerations are important, but decision aid developers do not currently have any guidance, and therefore need to rely on their own untested assumptions about hypothetical decisions.

The process of developing a decision aid involves many design considerations that the developers must take into account when seeking to provide the best tool possible for patients. The International Patient Decision Aids Standards (IPDAS) provides detailed recommendations about the components of good quality decision aids.[24] Investigators must consider design decisions such as what wordings to use, which numbers need to be provided, how to design the graphics, and examples to help patients understand the types of decisions that will need to be made along the way. Developers need to decide how much or how little information to include. It would be impossible for developers to conduct a real-world trial for each of these design decisions, yet data ideally should be collected on these issues in order to design an effective tool. Therefore, data is often collected using hypothetical settings and participants that are intended to provide data about how the decision aid will function in the real-world. However, there is uncertainty about external validity of these hypothetical data and where in the development process it is appropriate to use these data.[24] The uncertainty about how to utilize hypothetical data is demonstrated by the fact that the most prominent systematic review in this field excluded any studies from hypothetical settings.[8]

Both hypothetical and real-world studies have advantages and disadvantages. Real-world studies involve real life decisions by patients, but involve a level of time, effort, and resources that far exceed hypothetical studies. Even real-world trials may have validity and bias problems in some cases because the real-world can never be fully replicated; in the case of drug trials, the drug will often be applied in a wider range of patients and situations than it was in the trial. In contrast, hypothetical studies can be less resource intensive than real-world studies. They might also be ethically required as the untested intervention could have some unknown impact on the target patient population. However, unless we understand the factors that affect the validity of the data coming from these hypothetical settings, we cannot be sure that they are free from bias and accurately measure real-world decisions.

1.5 Objectives and Research Question

There is currently no framework that describes whether data generated from a study with hypothetical components will accurately predict real-world health behaviour and/or decisions. When developing decision aids, investigators need to ask themselves questions like: does commitment to acting on the decision matter? Is it important to devote resources to making the hypothetical setting resemble the real-world? Under what circumstances? Developing a framework that summarises the factors that affect the relationship between hypothetical and real decisions would provide useful guidance for developers wanting to use their resources as efficiently as possible to develop new health services interventions such as decision aids. It would assist trialists, implementation researchers, and anyone else wishing to design health interventions by helping them to determine ways to collect data that are both practical and externally valid. Additionally, it may be useful for systematic reviewers because it would prevent the erroneous exclusion of data by describing the factors necessary for hypothetical research to accurately predict real life. Such a framework would

have to draw on a range of literatures that explore the association between hypothetical and real-world decisions.

The overall goal of this project is to increase understanding of the factors that affect the association between hypothetical and real-world decisions. We have identified three specific research objectives. The first is to conduct a systematic concept review of studies describing the factors that distinguish hypothetical and real-world decisions and behaviours. The second objective will involve consulting developers of decision-aids in discussions about how and when hypothetical data informed the development of their decision aids, and what factors they believe affect the validity of these data. By reviewing the literature and obtaining the opinions of decision aid development experts, we will be able to identify and evaluate the factors that are important considerations when using hypothetical data in the decision aid process. The third objective will be to develop a framework describing the factors that may affect the relationship between hypothetical and real-world decisions, informed by the results of the systematic concept review and consultations with decision aid developers.

CHAPTER 2: METHODS

The overall goal of this thesis was to develop a framework to describe the factors that are important to the association between hypothetical and real-world decisions. To accomplish this, we have identified two specific objectives. The first is to conduct a systematic concept review to identify and describe the factors that may affect the correlation between hypothetical and real-world decisions. The second objective is to develop a survey of decision-aid authors to explore which factors in the framework apply to decision aid development. In this chapter we describe the methods associated with the two specific objectives in turn.

2.1 SYSTEMATIC CONCEPT REVIEW:

Although no specific reporting guidelines exist for systematic concept reviews, we have modeled our reporting on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (The PRISMA Statement).[25] We developed and followed a protocol for this systematic concept review that is unpublished but available on request.

2.1.1 OBJECTIVES:

Our research question for this review was: "What are the factors that affect the association between hypothetical and real-world decisions?" As identified in the Introduction of this thesis, the factors that are important to this relationship have been explored in a variety of research areas. Therefore, our literature review needed to be designed in a way that would allow us to successfully obtain information from diverse research fields. These factors were identified in order to develop a framework to predict the correlation between hypothetical and real-world decisions and to understand the important considerations when collecting and using data from hypothetical settings.

2.1.2 Search Strategy:

The author and his supervisor (TH & JB) hand searched the literature in order to identify a set of target articles that could serve as the foundation for this review. The nine target articles we found all identified multiple factors that affect the relationship between hypothetical and real decisions and were indexed in PsycINFO and/or Medline.[17,20,26–32] We consulted with a health science librarian from the University of Ottawa, who helped us develop an initial search strategy based on inclusion of the target articles and keyword and title searches for the terms: decision making, hypothetical/real and synonyms, relevant MESH headings, etc. We then consulted with a second health science librarian from the University of Ottawa who peer reviewed our search strategy and helped us to develop a final strategy. Appendix A describes the search strategy that was initially developed for PsycINFO and then modified for Medline. Our protocol originally stated that we would search PubMed, Embase, and CINAHL in addition to the aforementioned databases, but after further consultation with the librarians and the thesis advisory committee we determined that this would result in a workload that was beyond the scope of this project so these databases were not searched.

Our search strategy included both key words and MeSH terms. We identified search terms related to these themes using the identified target articles as well as suggestions from the librarians. The search strategy involved using the Boolean operator "and" to join three themes: decision making or behaviour (and related concepts e.g. reasoning, risk taking) hypothetical situations (e.g. uncertainty, proxy), and real-world situations (e.g. reality, everyday). We did not put any limitations on date of publication. Our electronic database search was conducted in November 30th 2015 and a supplemental snowball search was conducted in December 9th 2015.

We supplemented the electronic database search with a snowball search of the reference sections of the nine target articles.[17,20,26–32] The bibliographies of these articles were screened for relevant articles. Articles were screened based on their title with a liberal inclusion of any study that might yield factors relevant to the framework. If it was not clear whether the article would be relevant based on the title, then the abstract was screened as well. We also conducted a reverse citation search for studies that cited our target articles, using Scopus and Web of Science, depending on where articles were indexed. Hand search of the reference sections of all included studies yielded one additional article that was mentioned in an included study but not retrieved through the other means. Mendeley was used in order to compile the relevant articles collected by the review.[33] Duplicate citations were removed.

2.1.3 Eligibility Criteria

In order to be included in the review, an article needed to address clearly the relationship between some type of hypothetical decision and a corresponding real decision/behaviour. We did not exclude studies based on their methods; both empirical and commentary articles were included. Due to the financial burden of translation, only studies published in English or in French were included. Studies were not excluded based on the setting, time frame, or date of publication.

2.1.4 Study Selection

Using a pilot search strategy, we conducted a pilot abstract screen of the first 100 articles collected. A full text screen was conducted on the first 53 articles screened, regardless of whether they were included or excluded. We determined that an abstract screen was not an effective way to screen studies, as an excessive number of studies (25/53) met the inclusion criteria of the abstract screen and we found that many of the factors were being repeated in the retained articles. This indicated that our pilot search strategy was not specific enough to be feasible with available

resources. After consulting with library staff, we determined that the search strategy would be refined to be more restrictive and that studies would only be included if they were clearly relevant based on their title alone.

After the pilot screen, we completed the title searches and snowball searches using the refined search strategy and collected all citations into a single Excel spreadsheet. Two reviewers (TH & JB) screened the titles for eligibility independently. This was done in order to ensure that only articles that were explicitly and clearly dealing with the association between hypothetical and real decisions were included. If it was not clear whether the study was eligible based on the title, reviewers screened the abstract. The reviewers were not blinded to the journals or authors of the studies that they screened. Originally, we planned in the protocol to retrieve a study if either reviewer judged that the study should be included. However, this was determined to be excessive after piloting the search strategy and instead any disagreements were resolved through discussion and consensus.

After the completion of the title and abstract screen, the full text versions of the studies were collected and screened independently for eligibility by the two reviewers. At the full text screening stage, reasons for exclusion were tracked using an Excel form. At this stage, studies were only included if they clearly presented a factor that would be relevant to the framework describing the relationship between hypothetical and real-world decisions. The reviewers solved any disagreements at this stage by consensus, with JB acting as the final arbiter. The results of the study selection are presented using a PRISMA flow diagram.[25]

2.1.5 Data Collection

After all of the articles were screened for inclusion, the two reviewers proceeded with the data extraction stage. Both reviewers collected data independently using a standardized data

collection form developed using Excel. The data collection form was developed iteratively during the screening and data collection process. We calibrated the initial data extraction form at the beginning of the data collection process by piloting it using two randomly selected studies. Both reviewers collected data independently; any disagreements were resolved by consensus.

In addition to basic study information (e.g. title, journal, date of publication), we collected data about each study's Research Area, Design, and Research Question. The Research Area was determined by two coders (MP & TH) examining the titles and abstracts of the studies. These were coded into five categories: Behavioral Economics, Psychology of Reasoning, Social Psychology, Health-Related Behavior, and Neuroscience. After reading the titles and abstracts the coders classified each study into its relevant fields, a study could be classified as belonging to more than one field. The first 10 studies were assessed in order to determine the accuracy of the definitions used and to resolve any issues that were unclear. Studies without abstracts were coded based on the full text, read by TH. An "Other" category was included in order to classify any studies that did not fit into at least one of the other five categories. Disagreements were resolved through consensus, and JB was consulted when necessary. One coder, TH had read the full texts of all of the studies and therefore was able to provide clarifications when necessary in order to ensure that the studies were properly coded. The study's Design was paraphrased based on the Methods section. Information collected included the type of participants, the experimental methods, and how hypothetical and real settings were distinguished. The Research Question was collected based on a quote from the article.

We identified the target factors presented in the study by obtaining a quote naming and describing the relevant factor. TH extracted quotes from each study to describe how the factor affects the relationship between hypothetical and real decisions. Then, he summarized the quotes and standardized the language used in the descriptions to produce factor statements. JB supervised

this process and evaluated the factor statements to ensure correctness and consistency. We also rated whether the factors were central or tangential to the study's primary research question. The possible ratings were: 1 – Not Relevant, 2 – Central to the Research Question, 3 - Sub-issue of the Research Question. We also rated how closely each factor seemed to be related to the main goals of our study, i.e. 1 – Factor is related to Hypothetical-Real, 2 – Factor directly targets Hypothetical-Real, 3 – Factor describing Hypothetical-Real in a health context, 4 – Factor describing Hypothetical-Real in a decision-aid context.

2.1.6 Data Analysis

After data extraction was completed, TH paraphrased and summarised the quotes extracted from the articles to clarify the factor and how it was proposed to affect the hypothetical-real relationship. Our approach resembled what Hsieh and Shannon (2005) call a "Conventional Content Analysis."[34] This approach is useful when existing theory around the phenomenon being described is limited, as was our case.[34] In Conventional Content Analysis an inductive category development process is used, where key words are taken directly from the text in order to allow categories to flow directly from the data.[34] These codes are then grouped into meaningful clusters that allow them to be presented in a framework.[34]

Data analysis was an iterative process through which TH and JB standardised the presentation of the identified concepts in terms of a clear statement about factors that would increase the correlation, and those that would decrease it. Then, factors were grouped into common themes through an iterative, consensus-based process. These groups were created in a way that would facilitate their understandability in the context of a user-friendly framework. By using the authors' own vocabulary when describing identified concepts, we sought to avoid issues around missing subtleties of domains for which we did not have access to content experts.

Inter-rater reliability for coding of each study's Research Area was calculated using the R software environment and the "irr" package.[35,36] Unweighted Cohen's Kappas were used.

Because our goal was to identify as wide a range of potential factors as possible, we did not assess the risk of bias in individual studies included in this review, evaluate the quality of empirical support underlying each factor, nor did we assess the risk of bias across studies.

2.2 Pilot Survey Development

2.2.1 Survey Rationale

The second goal of this thesis project was to query decision aid developers about which factors from our framework were relevant in determining whether hypothetical decisions in the context of developing and piloting patient decision aids might serve to predict decisions by actual patients in real-world settings. There are various methods that could have been used in order to attain this objective such as interviews, focus groups, and self-administered surveys. We decided that a self-administered web-based survey might be the most appropriate method for several reasons. First, it allows contact with a broad sample of the study population. Second, e-mail addresses of the potential participants for our survey were easily accessible. Third, decision aid developers tend to be highly computer literate and have sufficient computer skills to be able to respond to an internet survey.[37,38] Self-response surveys are also more convenient for busy professionals compared to telephone or in-person interviews, which is important to ensure a high response rate.[38] Our survey included a fairly long list of questions that were similar in form (i.e. assessing the relevance of factors from the framework). We thought this would be tedious to do in an interview format.[38] Cost is another highly relevant issue for us; this work was not funded so

conducting an internet survey as opposed to a mail survey or interviews is one way of minimising the duration and cost of the study.[37,38]

2.2.2 Pilot Testing

In order to minimize nonresponse, it is important that a survey be well-designed and respondent-friendly.[37] In order to ensure that the questions being asked were relevant to the decision aid development context, TH, JB, and IG made decisions through consensus about which factors clearly were not related to this context; these factors were not included in the survey. We reframed the included factors for clarity and ratability in a survey format. Based on discussions between TH, JB, IG, and DC, we added three items about how participants in the hypothetical group could resemble the target group. These items were based on a factor from the framework ("Samples match target population"), but we determined that this statement needed to be clarified and made specific to the decision aid context for the purpose of our survey.

After conducting an initial set of cognitive interviews, we determined that it would no longer be appropriate to conduct a full survey of decision aid authors. The thesis advisory committee therefore recommended that TH continue conducting cognitive interviews with decision aid developers in order to elicit their feedback about the identified factors, but without the intention of developing a self-administered survey. The remainder of this chapter will be devoted to describing the cognitive interview process, and the results section will describe the results of the cognitive interviews in addition to the results of the systematic concept review. The methods for the survey that we had planned to conduct are presented in Appendix 2.

2.2.3 Cognitive Interviews

Decision aid developers were asked to participate in a think aloud interview by assessing the tool that was originally developed as a pilot survey.[39] Think aloud interviewing involves asking the

participants to read the questions out loud as they work through them, and asking them to voice their thoughts while reading and answering questions.[39] This cognitive method provides the investigator with insights into which questions may be difficult to answer, or do not elicit the types of responses that the investigators intended. Going through the process of think aloud interviews helps the investigators understand how to reformat their survey in a way that increases the understandability of the questions, produces consistent results, and produces the answers that the investigators intended to elicit. In order to further improve the design of our survey, we obtained feedback from local experts in decision aid development through this method of think aloud interviewing. Probing questions were asked when necessary in order to elicit insights about the experts' thought processes.[39] Cognitive interviews were iteratively conducted with 7 experts. After conducting an interview, we corrected any major issues that had been identified. We then verified in further interviews that the changes made resolved the issues and we identified whether there were any other major problems requiring correction. After all pilot interviews were conducted, we analyzed the results to identify any smaller issues that seemed consistent, and then corrected these for the final version.

2.2.4 Cognitive Interview Administration

The first question asks whether participants have been involved in the development of a patient decision aid, and if so, indicates that they should think about their most recent experience with decision aid development for this survey. This screening question serves to ensure that our sample identified the correct respondents. We asked them to provide the reference to the study they were thinking of while completing the survey to provide us with a reference for contextualizing answers - particularly the qualitative, open-ended comments - during analysis.

Questions two and three ask about how many years the participant has been involved in the development of decision aids and how many decision aids they have been involved in developing. This provided us with more information about the respondents and would allow us to assess whether responses to the other questions vary with decision aid developer experience.

Questions three and four ask respondents whether they used data from hypothetical situations when developing their most recent decision aid. We provided the participants with a list of examples of hypothetical situations and asked them to check all that apply (e.g. draft or early version decision aids, proxy decision makers, proxy decision making settings, other). These questions provided us with an understanding of how hypothetical data is actually being collected in the decision aid development process.

Finally, we provided participants with a list of the factors found through the systematic concept review and judged to be relevant to the decision aid development. We asked them to use a 5 level (Strongly Agree, Agree, Disagree, Strongly Disagree, Unclear/Unsure) Likert-like scale to describe their agreement with two statements for each factor: "This factor was relevant in the context of my most recent decision aid development process." and "This factor is relevant in general when decision to employ hypothetical settings for the development of decision aids." In order to simplify the presentation of the survey, the Factor Statements were framed consistently as "Hypothetical choices correlate *more highly* with real-world choices when…" or "Hypothetical choices correlate *less highly* with real-world choices when…" We also provided respondents with space to provide qualitative comments for each individual factor statement as well as for the survey in general.

2.2.5 Data Analysis

The interview process involved two to three interviewers and one or two interviewees. One interviewer (TH) gave an overview of the objectives of the thesis and the cognitive interview before explaining the think aloud interview process. The other interviewers (KC and MP) took written notes about the interviewees' answers and about their own observations during the interviews. They also helped TH prompt the interviewees for additional information when necessary. The interviews were audio recorded. TH extracted data from the interviews by compiling all written notes into one document and extracting important themes. He also listened to audio recordings to transcribe important quotes and extract additional themes that were missed in the written notes.

CHAPTER 3: RESULTS

3.1 SYSTEMATIC CONCEPT REVIEW

3.1.1 STUDIES INCLUDED

Before de-duplication, the online database search resulted in 1482 studies. (Figure 1.) An additional 661 studies were found through a snowball search and another 1 through hand search of studies. This resulted in 1846 studies after duplicates were removed. These articles were screened by title, which resulted in the exclusion of 1764 studies. The full texts of the remaining 82 studies were retrieved and screened for eligibility; 17 studies were excluded because they contained no factor for the framework, 6 were excluded because although they presented a potential factor it was too context specific to be relevant. An example of such a case was one study that identified the fact that hypothetical decisions about criminal behaviours are unlikely to be representative of real-world decisions of criminal behaviours.[40] 59 studies were therefore included in the systematic concept review and retained for qualitative synthesis.

Figure 2 describes the fields represented by the studies included in the review. As described in the methods, inter-rater reliability was calculated using Cohen's Kappa. The kappa scores were: Behavioral Economics: 0.84, Psychology of Reasoning: 0.62, Social Psychology: 0.51, for Neuroscience: 0.88 and Health-Related Behaviors: 0.65. Low Kappa scores were due to misunderstandings of the definitions that did not arise when conducting the pilot classification of the first ten articles. Limited information about reward type in the abstracts of articles also lead to some misclassifications of Behavioral Economics papers as Psychology of Reasoning papers, but these were resolved through discussion between the raters. The majority of studies included (47/59) were classified as Behavioral Economics studies. The next most prominent field among included

studies was Psychology of Reasoning (18/59). The fields of Social Psychology (10/59), Health-Related Behavior (7/59), and Neuroscience (3/59) were also represented in some of the included studies. Note that the number of field classifications (85) is greater than the number of included studies (59) because some of the studies were classified as belonging to more than one field.

Figure 3 describes the date of publication of the studies included in the review. The year from which we collected the most studies was 2013(8) follow by 2012 and 2004 (6). (Figure 4.) We included only one study published in the 60's, from 1969. We found no studies published in the 70's. Publication on the topic of the relationship between hypothetical and real decisions is increasing over time.

3.1.2 Study Methods

Table 1 gives examples of the research methods used in the included studies, organized by research area. The "Methods Used" column refers to how each given factor was identified in the literature. "Empirical Data" means that it came from a study that used experimentation or observation to collect empirical evidence that supported the factor, "Cited-Proposition" means that the article authors did not collect empirical evidence to support the factor, but instead cited another article where it had been identified, and "Meta-Analysis" means that it was identified through the results of a meta-analysis.

Table 1 Examples of Research Methods used in the studies included in the systematic concept review, organized by field of research.

Field of Research	Methods Used	Examples
Behavioral Economics	• Empirical Data (n=38) • Cited Proposition (n=6) • Meta-Analysis (n=3)	• Study examining factors that affect hypothetical and real willingness to pay for goods. [41] • Literature review examining the effect of financial incentives in experiments • Meta-analysis to explore features of experimental protocols that influence disparities between hypothetical and real-world values.[42]
Psychology of Reasoning	• Empirical Data (n=14) • Cited Proposition (n=4)	• Study designed to examine the effect of temporal distance on the correlation between intentions, attitudes, values, and behaviours to hypothetical and real participation in fitness activities and a blood drive.[43] • Literature review comparing different approaches to the study of reasoning in hypothetical and real-world contexts.[27]
Social Psychology	• Empirical Data (n=9) • Cited Proposition (n=1)	• Study where participants were tasked with deciding how to allocate 10 $20 bills between 4 people in hypothetical and real settings.[15] • Literature review of factors affecting hypothetical versus live interpersonal interactions.[44]
Health-related Behaviour	• Empirical Data (n=6) • Cited Proposition (n=1)	• Study assessing participants' hypothetical and real willingness for an asthma management programme.[45]

		• Literature review of factors affecting hypothetical and real medical decisions.[18]
Neuroscience	• Empirical Data (n=3)	• Study assessing neural activation when making hypothetical and real decisions to avoid eating disgusting food items.[46]

The field of Behavioral Economics represented most of the studies included in our review (47/59). Within this field, most of the studies involved Empirical Data Collection (38/47). This included studies that identified factors affecting the relationship between hypothetical and real-world decisions by using methods such as comparing participants' hypothetical vs. real-world willingness to pay for goods.[41] Some (6/47) studies included no empirical data but instead referenced other work making relevant claims through cited propositions to support their identification of factors. For example, we identified a literature review examining the effect of (hypothetical vs. real) financial incentives in various experiments.[47] The field of Behavioral Economics was the only field that included meta-analyses (3/47). Of these three meta-analyses, one involved examining the features of experimental protocols that influence disparities between hypothetical and real-world values.[42]

The field of Psychology of Reasoning represented 18/59 of the studies identified in this review. It included 14 studies involving Empirical Data collection. As an example of this type of study, we identified one experiment that was designed to examine how temporal distance affected the correlation between intentions, attitudes, values, and behaviours towards hypothetical and real participation in fitness activities and a blood drive.[43] 4 studies included no empirical data but cited other work that had previously identified a relevant factor. For example, we identified a literature review comparing various approaches to studying reasoning in hypothetical and real-world contexts.[27]

10 of the 59 studies were coded as falling within the area of Social Psychology. All but one of these (9/10) were experiments involving empirical data collection. These involved studies such as one where participants were tasked with deciding how to allocate 10 $20 bills between 4 people, which must be done unequally, in hypothetical setting, in a play setting which involved distributing "play" money to real people, and a real setting which involved distributing real money to other real people.[15] One literature review cited other studies examining the factors that affect hypothetical and real interpersonal interactions.[44]

The field of Health-related Behaviour represented 7/59 studies included in this review. All but one of these studies involved empirical data collection. As an example, one of these studies involved assessing the factors that affected participants' hypothetical versus real willingness to pay for an asthma management programme.[45] One literature review cited other studies examining the factors that affect hypothetical and real-world medical decisions. [18]

The field of Neuroscience represented 3/59 studies included in this review. All of these studies involved empirical data collection. One example of this type of study was an experiment that involved asking participants in hypothetical and real setting how much they would be willing to pay to avoid having to various eat disgusting foods. An fMRI was used in order to measure differences in neural activation between hypothetical and real decisions.[46]

Figure 1 Study Selection Process

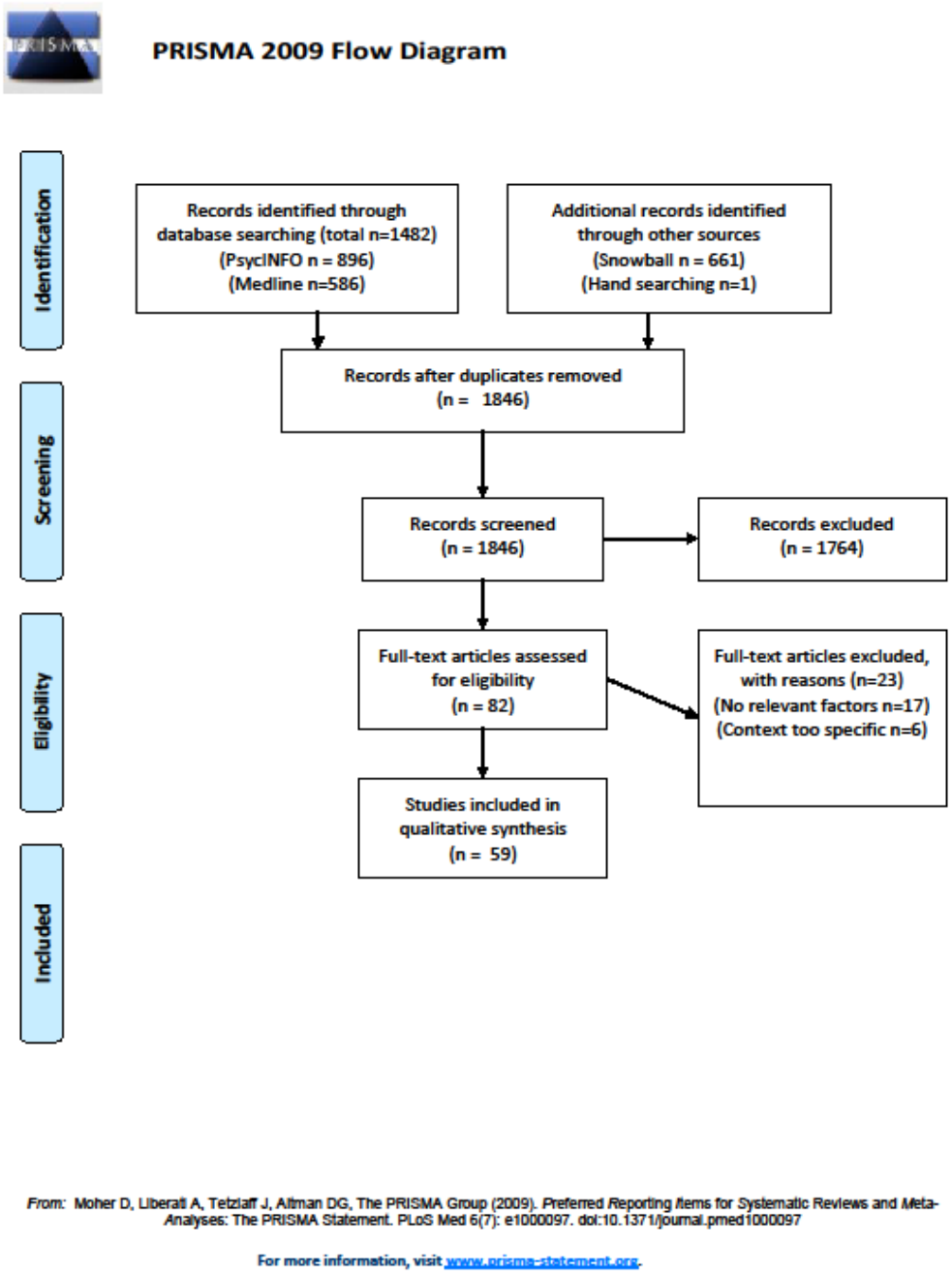


Figure 2 Proportion of included studies by field (n=59)

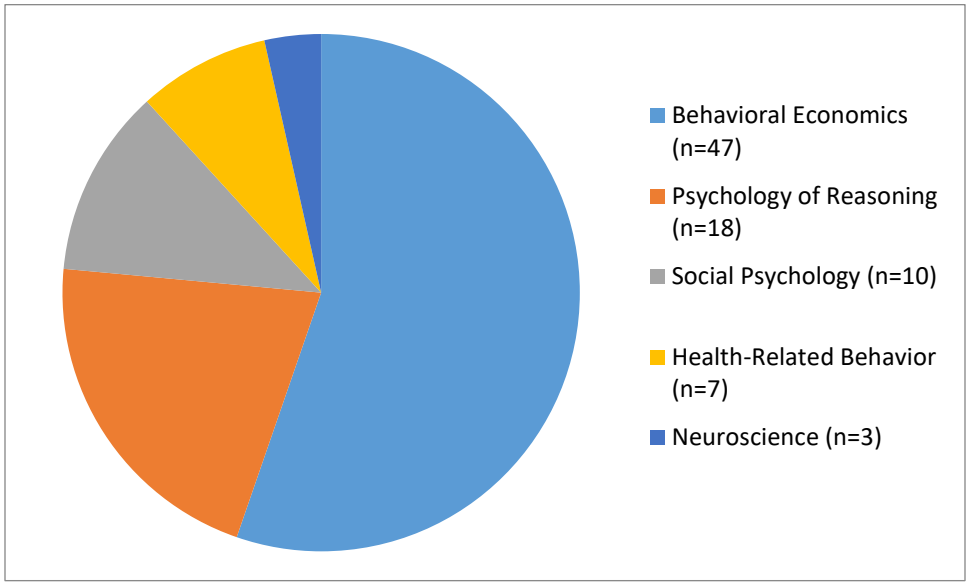
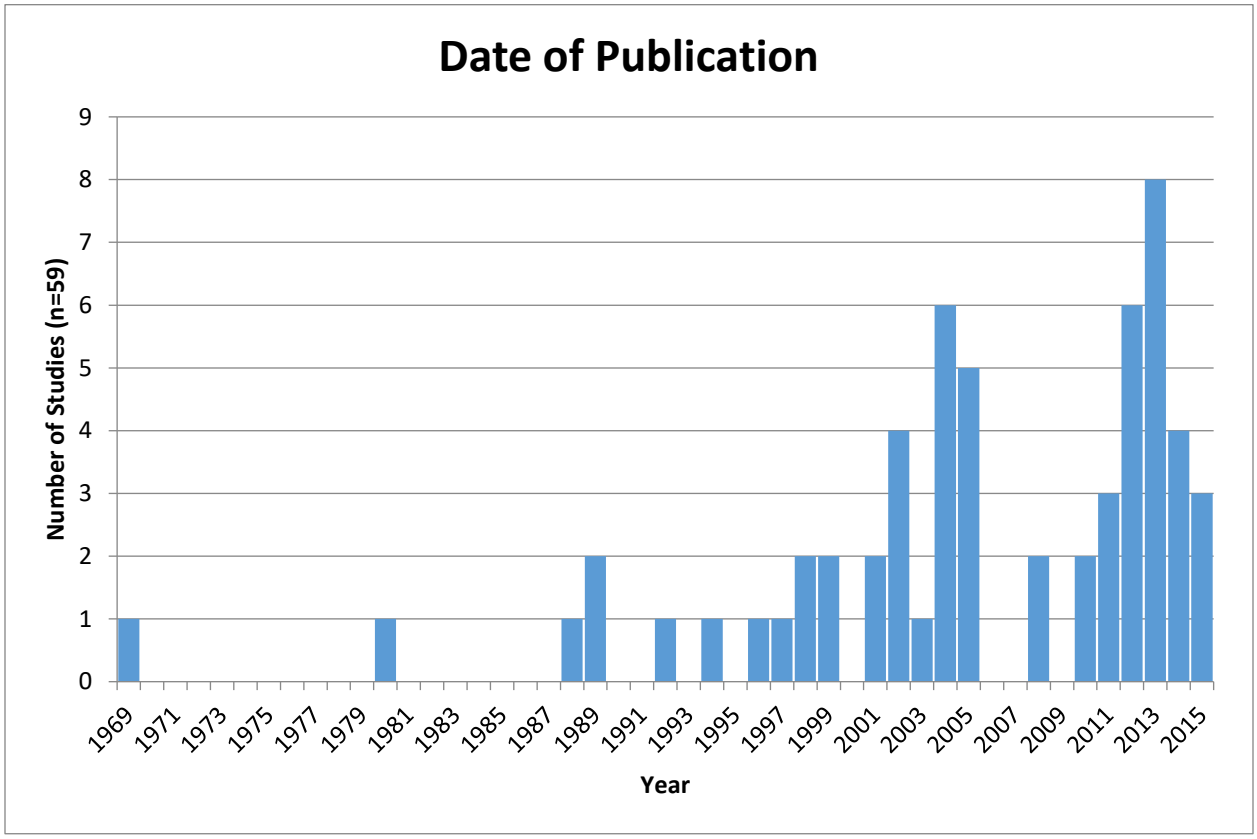


Figure 3 Included studies by date of publication (n=59)



3.1.3 FACTORS IDENTIFIED

We identified a total of 42 relevant factors from the 59 articles that were included in this review. We grouped them into 5 categories: Personal Characteristics, Presentation Characteristics, Cognitive Factors, Motivation, and Participant Characteristics. The following five tables present the definitions of the factors and how they affect the relationship between hypothetical and real-world decisions. (Tables 2-6) The "Factor Statement" is a description of how the factor affects the correlation between hypothetical and real-world decisions. The "Example Scenario" column is an example of the type of decision to which this factor has been found to be relevant in the literature.

Table 2 describes the Personal Characteristics factors identified in the review. These are factors that describe how characteristics of the individual making the decision affect how hypothetical decisions correlate to real-world ones. We identified 9 factors falling into this category.

Moral Reasoning Scores measure the Kohlberg Stage of Moral Reasoning employed by the decision-maker. These are scored using Moral Justice Interviews, which involve asking decision-makers to describe their decision in a moral dilemma. The interviewer then probes the decision-maker about their decision in order to encourage them to consider how their response might differ in other contexts. Finally, the interviewer compares the decision-maker's reasoning to a scoring manual in order to determine their stage of moral reasoning.[48,49] These stages vary from the pre-conventional (focus on the rewards or punishments of a decision), conventional level (focus on the expectations of family, group, or nation), to the post-conventional (reasoning based on abstract principles).[50] Higher moral reasoning scores (i.e. closer to the post-conventional level) are associated with hypothetical decisions compared to real-world decisions.[15,48,51] For example, Trevethan et al suggest that hypothetical dilemmas elicit the decision-maker's highest attained level of moral reasoning, while real dilemmas are associated with lower levels.[48]

Personality Traits are thought to capture people's patterns of thoughts, feelings and behaviours.[52] Different models, such as the Five-Factor and Big Six models of personality, involve questionnaires that evaluate the participants' traits related to: conscientiousness (i.e. organization), extraversion (i.e. sociability), agreeableness (i.e. cooperativeness), neuroticism (i.e. anxiety), and openness to experience (i.e. originality).[53] The Big Six model adds the factor of agency (i.e. dominance) to the other five.[53] *Personality traits* besides openness to experience have been shown to be negatively associated with consistency between hypothetical and real-world decisions. Using the Big Six model of personality, Grebitus found that participants with certain personality traits behaved differently in hypothetical and in real settings. They found that generally, participants favored either auctions or choice experiments depending on their personality type. However, in most cases the effects of a personality trait on real choice experiments was generally the opposite of its effect on hypothetical choice experiments. For example, the personality trait 'agency' was associated with a higher willingness to pay in real choice experiments, but a lower willingness to pay in hypothetical choice experiments.[53] Lonnqvist et al also found that personality traits did not predict decisions in the hypothetical setting, but low neuroticism was predictive of more cooperative decisions when real incentives were offered.[54] Taken together, these results suggests that certain personality traits may lead to inconsistencies between hypothetical and real life decisions.

The *Openness to Experience* trait is the personality trait from the Five-Factor Model of personality that has received the most attention in this context, which is why it has been listed seperately.[55] *Openness to Experience* is measured by assessing participants' answers to a questionnaire evaluating their openness to fantasy, aesthetics, feelings, actions, ideas, and values.[55] The *Openness to Experience* trait has been shown to be positively associated with consistency between hypothetical and real-world decisions. The two studies identifying this factor used the Five-factor model of personality to assess the relationship between hypothetical and real life decisions.[15,54] Their results

may have occurred because the decisions studied involved moral aspects and the ability to apply moral principles has been related to a high O*penness to Experience,* meaning that having high scores in this trait could be related to more moral behaviour.[54]

Cultural Differences as identified in this review specifically refers to differences between Western and Asian people related to their beliefs about intervening vs. letting events run their natural course. Gold et al refer to "Chinese fatalism" which is a set of cultural beliefs that the authors state make people more willing to let events resolve naturally without interference.[56] This is opposed to Western participants' cultural beliefs which make them more willing to intervene in a course of events. These cultural differences were measured using a moral dilemma that involved making a decision to flip a switch to take meals away from 5 orphans or from 1.[56] They found that these *Cultural Differences* lead to different decisions between British and Chinese participants in both hypothetical and real-world moral dilemmas.[56]

Age and Education were both identified in the same study. This study sought to identify socio-economic factors that affect hypothetical bias, defined in their study as the difference between stated willingness to pay and real willingness to pay.[57] Age was measured in years and education was a binary variable: either above or below high school education.[57] They assessed whether demographics characteristics of participants were associated with consistency between hypothetical and real life donation decisions. They found that older and more educated people showed more consistency between hypothetical and real-world decisions. They suggest that this might be related to older and/or more educated participants having a higher appreciation for the need to protect the environment, and increased experience with making pledges to do so.[57]

Cognitive Control (i.e. executive functions) refers to people's ability to concentrate and think about a given task and control their behaviour.[29,58] Morgenstern measured cognitive control using an

electroencephalogram (EEG) measuring event-related potentials. Activity measured from the "[...]fronto-medial part of the anterior cingulate cortex (ACC) and the prefrontal cortex (PFC)", observed about 200 milliseconds after the stimulus is presented, is thought to indicate cognitive control.[29] Morgenstern et al hypothesized that when comparing EEG readings of participants choosing between hypothetical lottery options to those of participants in a real setting, the real decisions would elicit higher cognitive control.[29] However, they found that greater cognitive control was actually observed in hypothetical decisions. This led them to suggest that there is an additional choice criterion (such as the expected value of a lottery) that evokes higher cognitive control for hypothetical decisions. They found that participants were less risk averse when making hypothetical choices compared to real lottery choices, and that this decreased risk aversion was associated with higher cognitive control. This means that cognitive control was found to be negatively associated with consistency between hypothetical and real-world decisions.[29]

Personal Relevance was identified in a study dealing with hypothetical vs. real moral dilemmas. The authors of this study hypothesized that emotional reactions would be stronger for real-life dilemmas than for hypothetical dilemmas because of personal relevance.[21] They measured personal relevance by assessing whether the dilemma involved people with whom the participant had ongoing and long-term relationships.[21] They found that the *Personal Relevance* of a decision may be positively associated with consistency between hypothetical and real-world decisions. When assessing hypothetical and real moral dilemmas, participants became more emotionally aroused when the moral conflict involved people with whom they had long-term relationships than when they assessed moral conflicts involved strangers or conflicts with oneself.[21]

Finally, *Cognitive Ability* was measured in the context of a gamble experiment to measure participants' risk aversion.[31] In this case, *Cognitive Ability* was measured with a nine item test

measuring participants' ability to understand numeric and probabilistic information as well as their problem solving ability.[31] The types of problems included in the cognitive reflective test evaluate participants' System 2 cognitive processes, meaning that they require more effortful concentration to solve.[31] It was found to be negatively associated with consistency between hypothetical and real-world decisions. Taylor found that high cognitive ability was associated with reduced risk aversion in hypothetical gambles, but not in real gambles. Taylor theorized that this might occur because participants with high cognitive ability treated hypothetical tasks as puzzles to solve, rather than answering based on how they would in real life.[31]

The Personal Characteristics factors were all identified in studies collecting empirical data. Factors that were shown to be positively associated with consistency between hypothetical and real-world decisions were: *Moral Reasoning Scores, the Openness to Experience Trait, Age, Education, and Personal Relevance.* The factors that were shown to be negatively associated with consistency between hypothetical and real-world decisions were: *Personality Traits, Cultural Differences, Cognitive Control,* and *Cognitive Ability.*

Table 2 Definitions of Personal Characteristics factors (n=9)

Name of Factor	Definition of factor	Factor Statement	Example Scenario	References
Moral Reasoning Scores	The Kohlberg stage of moral development that the decision-maker employs when making their decision	Higher moral reasoning scores associated with hypothetical moral decisions.	Compared hypothetical moral decisions to previously experienced moral decisions[48]	[15,48,51]
Openness to Experience Trait	Measure of the decision-maker's openness to a variety of ideas and experiences.	Greater openness to experience trait (IASR-B5) is predictive of hypothetical decisions but not real ones.	Compared hypothetical and real decisions about unequally distributing money.[15]	[15,54]

Personality Traits	Personality traits, other than openness to experience, of the decision-maker.	Personality traits, especially extraversion and low neuroticism correlated with inconsistently between hypothetical and real life decisions.	Compared hypothetical and incentivized prisoner's dilemma games.[54]	[53,54]
Cultural Differences	Cultural differences between decision-makers	Cultural differences related to willingness to act affects how people respond to hypothetical and real life moral decisions	Compared hypothetical and real decisions about donating meals.[56]	[56]
Age	Age of the decision-maker	Older decision makers show more consistency between hypothetical and real life decisions	Compared hypothetical and real decisions about making financial donations.[57]	[57]
Education	Educational attainment of the decision-maker	More educated decision makers show more consistency between hypothetical and real decisions	Compared hypothetical and real decisions about making financial donations.[57]	[57]
Cognitive Control	The decision-maker's use of mental processes to concentrate and think	Higher cognitive control is correlated to inconsistency between hypothetical decisions and real decisions	Compared hypothetical and real lottery decisions.[29]	[29]
Personal Relevance	Whether the decision being made is one that involves people with whom the decision-maker has long-term relationships.	Personal relevance of a problem is correlated with consistency between hypothetical and real decisions	Compared hypothetical and self-generated real moral dilemmas.[21]	[21]
Cognitive Ability	Cognitive ability of the decision-maker.	Higher cognitive ability increases risk aversion for hypothetical decisions but not for real life ones	Compared hypothetical and real monetary gambles.[31]	[31]

Table 3 describes the Presentation Characteristics factors identified in the review. These are factors that describe how characteristics of the decision itself and the way that it is presented to the decision maker affect how hypothetical decisions correlate to real-world ones. We identified 8 factors falling in to this category.

The *Study Procedure* factor relates to what information is presented and how the decision is made. The three different studies identifying this factor highlight different components of the study procedure. The Dixon et al study focused on methodological consistency.[59] Their study involved measuring the effect of a "blackout period", a set period of time where the participant is unable to make a decision.[59] The blackout period was used in order to maintain a constant reinforcement rate, meaning that the participant was not simply choosing the immediate rewards to avoid having to experience delays in order to maximise the number of choices they are allowed to make.[59] Müller et al measured study procedure in terms of context matching, that is to say where the hypothetical choice setting matched the "task environment" where the real choice will take place.[60] Ebbesen et al also measured the effect of the study procedure in terms of context matching between hypothetical and real decisions.[61] They also examined the effect of "seemingly insignificant features" of the decision such as, the number of cues given, how concrete the information presented was, and the order in which the information was presented.[61] They concluded that these features can impact participant responses and change the decisions they make.[61] This suggests that it is important that the procedural features of the hypothetical decision resemble those of the target decision in order to elicit the same decision strategies. The other studies identifying this factor reached similar conclusions: Consistent procedural elements may be positively associated with consistency between hypothetical and real-world decisions.

Space for Mental Simulation refers to the degree to which people must imagine the decision making situation and its consequences.[14,17] Both studies identifying this factor measured the space for mental simulation in relation to the amount of contextual information provided; the more contextually rich situations required less mental simulation.[14,17] *Space for Mental Simulation* has been found to be negatively associated with consistency between hypothetical and real-world decisions.[14,17] FeldmanHall et al opposed the goals of self-gain and aversion to harming others by comparing participants' decisions to spend money to prevent harm to another.[17] They compared decisions in a real setting where actors pretended to receive electric shocks when the study participants did not spend real money to prevent them to: a hypothetical scenario with a textual description of the situation, an enriched scenario with a more detailed textual description, a trial-by-trial scenario where subjects completed 20 computerized trials of the scenario, and a near-real setting where the participants met the person who would receive the hypothetical shocks and acted out their decisions receiving hypothetical money.[17] The authors concluded that as the contextual details of the setting increased, the participants had less *Space for Mental Simulation* of their decision, and therefore made decisions that resembled real life more than the decisions for less contextually rich settings. Patil et al conducted a similar study, comparing decisions from textual moral dilemmas to decisions from virtual reality moral dilemmas.[14] They also found that people made decisions that were closer to real life moral decisions in the virtual reality setting because the contextual information in this environment provoked behaviour-motivating emotions and decreased the *Space for Mental Simulation* compared to purely hypothetical dilemmas.[14]

Salience refers to the significance of the decision to the decision-maker.[15] This factor was identified in a study that examined how participants allocated money when they were forced to do so unequally. *Salience* in this context, refers to whether or not the recipients of the allocated money were seen (as opposed to imagined) by the decision-maker.[15] Day stipulates that seeing the people

who will receive the allocated money increases the *Salience* of the decision for the decision-maker.[15] They found that the *Salience* of the decision was positively associated with consistency between hypothetical and real-world decisions. There were three relevant groups in this task: a group thinking about what they would do in this task hypothetically (hypothetical), a group where they were making the decision by distributing fake money to real people (play), and a group where they distributed real money to real people (real). Day found that by increasing the *Salience* of the decision by putting participants face-to-face with the people who would be harmed by the consequences of their decision, they allocation decisions of the play group resembled the decisions of the real group more closely than did those of the hypothetical group.[15]

Definition of Decisions relates to the approach taken to define and study the participants' reasoning. Specifically, it involves whether or not the decision making situation evokes the use of the models/search approach. This approach involves the construction of mental representations of the decision-making situation, and the use of the provided information to test and reach conclusions.[27] It is contrasted with two other approaches: the componential approach (i.e. using cognitive processes to find relationships), and the rules/heuristics approach (i.e. following rules to make inferences).[27] Defining the decision to encourage participants to use the models/search approach to reasoning has been found to be positively associated with consistency between hypothetical and real-world decisions.

Uncertainty is a measure of whether or not explicit statements about the probabilities that the outcomes of the decision will occur were presented to the decision-maker.[62] Making explicit statements about the uncertainty of outcomes has been shown to be positively associated with consistency between hypothetical and real-world decisions. Harrison conducted a review of the literature to determine how decisions differ in hypothetical and in real settings when the outcomes

are uncertain – such as in lotteries where uncertainty is controlled (exogenous lotteries).[62] One of the author's conclusions is that decisions with uncertain outcomes can be affected by hypothetical bias (i.e. less risk aversion in hypothetical than in real decisions). However, hypothetical bias can be reduced by making explicit statements about the *Uncertainty* of the outcomes. This is because these statements cause respondents to answer hypothetical decisions in a way that more closely resembles their real life decisions.[62]

The *Fundamental Attribution Error* refers to how the hypothetical decision is presented to the decision maker. If the decision is presented with the decision-maker as the actor in the situation (e.g. by asking "What would *you* do?") the decision-maker is more likely to believe that there are situational constraints affecting their decision. Whereas if the decision is presented with the participant as an observer (e.g. by asking "What do you believe *someone else* should do?") then they are more likely to state that the other person has more control over their behaviour than they would in that situation.[13] Gold et al tasked participants with making decisions about allocating money to orphans. Participants were randomly allocated to being given the problem as actors (they were the ones making the decision) or observers (they were deciding how another person should act in that situation.) They were tasked with deciding whether they/another should flip a switch or drag a photo of a child into the path of a moving ball, thereby taken a donation away from one orphan but giving it to five others.[13] The authors concluded that when people are making moral decisions about how others should act, they are more inclined to suggest that the difficult action be taken than when they need to make the decision themselves.[13] This suggests that when the hypothetical problem is framed as if the participant is the actor, in order to avoid the *Fundamental Attribution Error,* their decision will be more consistent with what they would decide to do in real life.

The *Framing Effect* refers to how the decision is phrased. This affects the decision maker's initial judgements about the decision.[63] For example, Levin et al found that participants were generally more likely to accept a gamble when it is framed in terms of gains (i.e. positive frame) as opposed to losses (i.e. negative frame.)[63] Although they concluded that participants are affected by the *Framing Effect,* they found that its impact was reduced when there was personal involvement in the decision, as when real money was offered.[63] Therefore, the impact of the frame of the decision has a greater impact in hypothetical decisions than in real-world ones and so the *Framing Effect* may be negatively associated with consistency between hypothetical and real-world decisions.

Time for Reflection is related to how the decision is presented to the decision-maker. This is a measure of the amount of time for that reflection was available to the decision-maker when they were making their decision. This time for reflection relates to the spontaneity of the decision, as more time for reflection allows the decision-maker more time to process the information and think about different outcomes.[23] Differences in the available *Time for Reflection* have been shown to be negatively associated with consistency between hypothetical and real-world decisions. Van Nieuwenhuijze et al compared the decision making of children with mild intellectual disabilities (MID) and externalizing behaviour problems to children with MID and no externalizing behaviour problems. The children were tasked with describing how they would behave in a situation where other children were behaving uncooperatively in games shown in a hypothetical video vignette. Their actual behaviour was then observed by having them experience a game where other children were told to be uncooperative.[23] The authors concluded that because the children had more time to reflect on their decisions in the hypothetical setting, their responses were less spontaneous and therefore differed from their behaviour in the real setting.[23] This suggests that when there is more time for reflection in the hypothetical setting than there will be in real life, decisions may differ between the two.

Most of the Presentation Characteristics factors were identified in studies collecting empirical data. Only the *Uncertainty* and *Definition of Decisions* factors were identified through cited propositions, as the studies identifying these factors were literature reviews.[27,62] Factors that were shown to be positively associated with consistency between hypothetical and real-world decisions were: *Salience, Uncertainty, Study Procedure, and Definition of Decisions*. The factors that were shown to be negatively associated with consistency between hypothetical and real-world decisions were: *Space for Mental Simulation, Fundamental Attribution Error, Framing Effect, and Time for Reflection.*

We identified two studies that disagreed with the conclusions identified by other factors. First, a meta-analysis of factors predicting consistency between hypothetical and real-world decisions found that whether an experiment was conducted in the laboratory or in the field does not lead to a difference in hypothetical compared to real decisions, which is somewhat counter to the conclusions of the *Study Procedure* factor.[64] Second, a lottery study examining participants' aversion to risk found that losses were considered the same way in both hypothetical and real-world decisions, which is somewhat counter to the *Framing Effect* factor.[65]

Table 3 Definitions of Presentation Characteristics factors (n=8)

Name of Factor	Definition of Factor	Factor Statement	Example Scenario	References
Study Procedure	Whether the procedural elements of the decision, how the decision is made and what information is presented, are consistent between the hypothetical decision and the real one.	Procedural elements resembling real life decisions increase how much hypothetical decisions resemble the real-world	Compared decisions between hypothetical and real delay discounted monetary rewards.[59]	[59–61]
Space for Mental Simulation	How much the decision-maker needs to mentally simulate, imagine, features of the decision.	Increasing contextual information decreases the space available for mental simulations and makes hypothetical decisions more like real-world decisions	Compared hypothetical (text) to real (virtual reality) moral dilemmas.[14]	[14,17]
Salience	When the decision seems significant to the decision-maker	Seeing people who will be harmed by the consequences of a decision makes hypothetical decisions more like real life ones.	Compared hypothetical and real decisions about unequally distributing money.[15]	[15]
Definition of Decisions	Whether the decision is defined in a way that encourages the decision-maker to construct a mental representation of the situation and to use contextual information to test and reach conclusions.	Defining hypothetical decisions to elicit the models/search approach increases how much they resemble real-world decisions	Literature review comparing approaches to the study of reasoning.[27]	[27]
Uncertainty	When estimates of the probability of the outcome are explicitly presented to the decision-maker	Providing statements about uncertainty increases consistency between hypothetical decisions and real life	Compared hypothetical and real monetary lottery decisions.[62]	[62]

Fundamental Attribution Error	Whether the decision is worded in a way that asks the decision-maker what they would do or asks what they think someone else should do.	Presenting the hypothetical decision with the decision-maker as the actor (as opposed to an observer) increases consistency between hypothetical and real-world decisions.	Compared hypothetical and real decisions about making financial donations.[13]	[13]
Framing Effect	Whether the decision is framed in a way that is positive (i.e. gains) or negative (i.e. losses)	The framing effect may be larger for hypothetical decisions than for real life ones.	Compared hypothetical and real monetary gambles.[63]	[63]
Time for Reflection	Whether the decision-maker is provided with the same amount of time to make their hypothetical decision as they would have in the real-world	Excessive time for reflection makes hypothetical decisions less like real life ones	Compared hypothetical and real responses to a social problem solving decisions.[23]	[23]

Table 4 describes the Cognitive factors identified in the review. These are factors that describe how the ways people think and make decisions affect how hypothetical decisions correlate to real-world ones. We identified 17 factors falling in to this category.

Discounting refers to the fact that the value of a reward decreases proportionally to the delay to receiving the reward.[66] Delay discounting is measured by asking decision-makers to choose between an immediate reward and a delayed reward of the same value, and then increasing the value of the delayed reward compared to the immediate reward in order to determine the rate of delay discounting.[66] *Discounting* has been shown to be consistent across hypothetical and real-world decisions, and so it has no effect on whether hypothetical and real-world decisions will be consistent across situations. [11,18,66–71] Six studies directly compared participants' choices between immediate and delayed rewards in hypothetical and in real settings. All six of these studies agreed that there was no

systematic difference in delay discounting of hypothetical and real rewards.[11,66–68,72] Two studies compared hypothetical choices between immediate and delayed rewards to other forms of related real behaviour. One of these, a study by Bickel et al, compared delay discounting choices to redemption of vouchers that participants' received for returning opioid-negative urine samples. After providing the researchers with a negative urine sample, participants could decide between redeeming their voucher right away for a reward, and depositing it in an account to receive an increased reward with further negative samples. The authors found that rates of delay discounting were related to the frequency of redeeming vouchers, and therefore that delay discounting was consistent between hypothetical and real-world decisions.[69]

Degree of Certainty is how certain the decision-maker is that they would make the same decision hypothetically as they would in the real-world. This can be measured by asking participants, after they have stated their hypothetical decision, whether they are "definitely sure" or "probably sure" that this is the same decision they would make in the real-world.[45] *Degree of Certainty* has been shown to be positively associated with consistency between hypothetical and real-world-decisions.[32,42,45,73–76] For example, two studies by Blumenschein et al compared decisions in hypothetical and real willingness to pay experiments. In one study, they asked people in the hypothetical setting whether they would be willing to purchase an asthma management programme, and to those who said yes they asked a follow up question about whether they were "definitely sure" or "probably sure" about their answer; participants in the real group were deciding on whether or not they actually wanted to purchase the product. Although participants in the hypothetical group were more likely to state that they would purchase the programme than participants in the real group, the authors were unable to find a statistically significant difference between "real yes" purchase decisions and "definitely sure" hypothetical decisions.[45] The authors had previously found

similar results when conducting a study where participants were instead asked about their willingness to purchase a pair of sunglasses.

The *Hot/Cold Empathy Gap* occurs when someone in a "cold", or hypothetical state, is not able to properly imagine and take into consideration the influence of "hot" emotional or visceral factors.[46] Visceral factors involve the physical changes brought on by strong emotional states that can be difficult to imagine (e.g. feelings of disgust, rapid heartbeat, etc.)[22,46] The *Hot/Cold Empathy Gap* was been shown to be negatively associated with consistency between hypothetical and real-world decisions. One study identifying this factor, by Kang & Camerer, observed the hot/cold empathy gap by asking participants about their willingness to pay money to avoid having to eat disgusting food items. They conducted these experiments in hypothetical and in real settings and found that participants in hypothetical settings were not able to fully appreciate how they would be viscerally affected when confronted with the disgusting food in real life, and therefore underestimated their aversion to eating the food.[46] Kuhberger found similar results when conducting a gambling study in real and hypothetical settings. He suggests that people are not able to adequately imagine the influence of strong emotional changes because these reactions have actual physical reactions that are difficult to imagine. People therefore under predict the influence of visceral factors when making hypothetical decisions and they incorrectly project their current emotional state onto the state they imagine they will be in when making the real-world decision.[20]

Degree of Risk Aversion refers to people's preferences in gambles between smaller, definite rewards versus larger, uncertain rewards. The degree of risk aversion is measured by observing how much the probability of a higher payment option would need to increase for a decision maker to convert from a smaller, sure-thing option to a larger, risky option.[12] *Degree of Risk Aversion* has been shown to be negatively associated with consistency between hypothetical and real-world decisions.

Holt & Laury conducted two studies to assess differences in participants' risk aversion when making hypothetical and real decisions.[12,77] In both studies, participants made choices between paired lottery options in hypothetical and real settings, where higher rewards came with greater risks. In the second study only, they included a second group that only completed either the hypothetical or the real condition. In both studies, they found that risk aversion is significantly higher when the decisions involve real consequences, and that risk aversion increases in proportion to the magnitude of the potential rewards only when the payments are real.[12,77] These studies agreed with the earlier findings by Slovic that in hypothetical conditions participants seek to maximise gains and discount potential losses, but when the decisions are real participants are more risk averse.[30]

Familiarity with a Good refers to whether the decision-maker has had previous experience with the outcome or subject of the decision that they are being asked to make (i.e. the 'good' or 'goods' being considered). This could be because the decision is about something with which they are very familiar, such as a topic that they have studied thoroughly, or involves an activity in which the decision-maker has engaged.[43,57] *Familiarity with a Good* has been shown to be positively associated with consistency between hypothetical and real-world decisions. For example, Mjelde compared hypothetical and real decisions to donate to charities and found that familiarity with the cause of the charity lead to hypothetical decisions that more closely resembled real decisions.[57] The authors suggest that this may be the cause of observed differences in correlation between hypothetical and real private vs. public goods.[57]

Normative Beliefs refer to the decision-maker's thoughts about what others who are important to them, such as family and friends, would think about their decision. This can be measured by asking decision-makers about the strength to which they agree with normative statements (e.g. "My family would think that I should do x") and their motivation to comply with normative beliefs (e.g.

"Generally I do what my family thinks that I should do").[41] The impact of *Normative Beliefs* has been shown to be negatively associated with consistency between hypothetical and real-world decisions. A study by Ajzen et al examined participants' willingness to contribute to a scholarship endowment fund in a hypothetical and in a real situation. Participants were given $10 to complete an unrelated task and then told that they were to participate in a referendum as a group where they would vote on whether they should donate money to a scholarship endowment fund. In the real condition, if the majority voted 'yes' then they would actually have to donate the money; in the hypothetical setting no money would actually be collected but they were told to vote as if it would be.[41] The authors found that a distinguishing feature of hypothetical and real decisions in this problem was that hypothetical decisions were less likely to activate normative beliefs. This means that in the real setting participants were concerned about what their family would think about them giving away their money, and they worried that the other members of the group might not want to donate. These normative beliefs were not activated in the hypothetical condition.[41]

Abstract Construals refers to the features of the decision that are the focus of the decision-maker. Construal-level theory postulates that people conceptualize decisions differently depending on the decision's distance from the here-and now. Distance can be caused by a decision being farther away from an individual's present state on the level of "[…] time, space, social distance, [or] hypotheticality."[78] Using abstract versus concrete construals, that is to say thinking about the general features versus the specifics of a decision, has been shown to be negatively associated with consistency between hypothetical and real-world decisions. According to a review by Eastwick et al, as a decision's hypotheticality increases the way that people think about the decision (i.e. their mental construals) become higher-level and more abstract. They do not focus on the specifics of a decision, but rather its general features. If important specifics about a decision may affect it, then

using only abstract construals in a hypothetical situation may result in differences between hypothetical and real decisions.[44]

Forecasting Emotions refers to the fact that different processes are involved when people experience actual emotions and when they imagine emotions they will experience in the future or have experienced in the past. Because people are unable to perfectly imagine emotions they are not actively experiencing, they rely on emotional schemas, beliefs about the emotions they will feel, when forecasting emotions to decisions that they are not actively making.[44] Forecasting of emotions was also shown in the Eastwick et al review to be negatively associated with consistency between hypothetical and real-world decisions. Because they may rely on forecasted emotions when deciding about how they would react in hypothetical interactions, hypothetical decisions may be less likely to correlate with real-world decisions that rely on experienced emotions.[44]

Mindset refers to the information that the decision-maker is processing to help them determine what their decision will be. The deliberative mindset involves evaluating the pros and cons of different options and whether they will successfully achieve each different goal.[44] The implemental mindset involves focusing on information that is useful for the decision-maker to complete their selected goal; they avoid or ignore information about the attainability of their selected goal.[44] Different contexts may evoke the use different mindsets, for example people generally think about their romantic relationships with an implemental mindset.[44] *Mindset* (deliberative vs. implemental) has been shown to be negatively associated with consistency between hypothetical and real-world decisions. Eastwick suggests that in hypothetical settings people may be more likely to use a deliberative mindset, and in real settings they may more often use an implemental mindset.[44]

Predicted vs. Experienced Utility refers to how people value the outcomes of their decisions. Predicted utility refers to how people expect they will value an outcome when they are making a

decision. Experienced utility refers to how people value an outcome when it is experienced.[18] In the medical context, the value of an outcome can refer to how people predict they will feel towards a health outcome experienced through a treatment versus how they feel when they actually experience that health outcome.[18] Predicted vs. Experienced Utility has been shown to be negatively associated with consistency between hypothetical and real-world decisions. Chapman found that people may be poor at making hypothetical decisions about medical treatments that correspond to what would be their real decision because they have not actually experienced the consequences of the various treatments or the health state that would require their use. Experiencing an adverse health state may affect health preferences in ways that are difficult for individuals to predict. Therefore, their lack of experience with a health state means that their hypothetical decisions may be poor indicators of real-world decisions.[18]

Naturalism bias is the incorrect belief that things that are natural are necessarily good. This can affect decisions when people prefer natural products to artificial products that are functionally identical or more effective.[18] The *Omission Bias* refers to the belief that bad outcomes that occur due to inaction are preferable to the same bad outcomes when they are caused by action. This can affect decisions when people avoid taking an action with potentially beneficial effects (such receiving a vaccine) when that action has a risk of negative consequences that are equal or less than the negative consequences of inaction.[18] *Omission and Naturalism Bias* have been shown to be positively associated with consistency between hypothetical and real-world decisions. The Chapman study is a summary of the literature related to the psychology of medical decision making. The author explains that some biases, such as the *Omission and Naturalism Biases*, exist equally in both real life and hypothetical decisions, and that demonstration of decision biases in hypothetical studies can be linked to the same bias in real-world decisions. They found that people who demonstrate these biases on a

questionnaire were more likely to decline a flu shot in real life; suggesting that demonstration of bias in hypothetical decisions could affect how real-world decisions are made as well.[18]

Social Desirability Bias can affect how the decision-maker interacts with the investigator when their decision is being studied. The decision-maker may attempt to exhibit socially desirable qualities to the investigator through their decisions, for example by making decisions that make them appear generous.[47] *Social Desirability Bias* has been identified as being negatively associated with consistency between hypothetical and real-world decisions. In a literature review of the effects of financial incentives, Camerer found that when making hypothetical decisions, participants may be affected by social desirability bias. This means that they try to act in ways that are socially desirable in their interactions with the experimenter. This is expressed as choosing risker options or options that allow them to appear generous. However, when the decision is made more real by offering real financial incentives, participants are no longer affected by *social desirability bias* and are more likely to make decisions that are less socially desirable but more highly correlated with real life behaviour.[47]

Dispositions refers to the way that people think about making decisions. The study identifying this factor specifically references dispositions that "[…] help a person accept a prolonged state of doubt, enjoy the process of discovering and challenging ideas, and appreciate the need for procedures that reduce bias."[27] These *Dispositions* about thinking have been shown to be positively associated with consistency between hypothetical and real-world decisions. In her review of the literature, Galotti suggests that in order for formal, or hypothetical, reasoning to more closely resemble every day, or real-world, reasoning people need to have dispositions about thinking that allow them to accept the necessary strategies to make hypothetical reasoning more closely resemble real-world reasoning.[27]

Subjective Beliefs refer to the beliefs that people have about the probability of outcomes occurring. If probabilities of outcomes are not explicitly presented and explored with the decision-maker they may have their own implicit beliefs about the probabilities that could affect their ultimate decision.[62] *Subjective beliefs* about the uncertainty around outcomes have been shown to be negatively associated with consistency between hypothetical and real-world decisions. Harrison conducted a literature review of studies that examined hypothetical bias when the outcomes were uncertain. The author found that providing participants with explicit statements about uncertainty increases the correlation between hypothetical and real decisions.[62] They further found that it is important to ask participants about their subjective beliefs about levels of uncertainty. This is because participants may have subjective assessments about the levels of uncertainty around the outcomes being assessed, and by rendering these assessments explicit the experimenter is able to control for them and clarify expected outcomes.[62]

The *Vividness of the Decision Process* factor relates to how easy it is for a decision-maker to imagine the decision process and the consequences of their decision.[79] The *vividness of the decision process* can be affected by whether or not decision makers know that they will actually experience the outcome of their decision.[79] Enhancing the *vividness of the decision* has been shown to be positively associated with consistency between hypothetical and real-world decisions. Beattie & Loomes conducted a study comparing hypothetical and real gambling decisions. They had four different problems with different options with varied probabilities and rewards. Participants were assigned to either: a purely hypothetical group where they answered all four questions but none of them were done for real; a random problem selection procedure group where they answered all four questions and one randomly selected problem was carried out with real rewards; or a real group where they were only given one question to answer but it was played out with real rewards.[79] They concluded that overall providing real rewards does not lead to different decisions than purely hypothetical

scenarios. However, they found that when the *vividness of the decision process* is enhanced, by encouraging subjects to imagine how they would feel if they lost their rewards, decisions may more closely resemble those made in real life.[79]

Attribute Non-Attendance (ANA) is a concept in economics that refers to the fact that when respondents make choices, they sometimes will neglect to fully consider some of the attributes of a decision.[80] *ANA* can be measured by asking decision-makers how much they focus on the different attributes of their choices, or it can be modelled using a mixed logit model.[81] *ANA* has been shown to be negatively associated with consistency between hypothetical and real-world decisions. Morkbak et al assessed participants' *ANA* in choice experiments of preferences for apples using hypothetical and incentivized samples.[81] They found that for real incentivized decisions, respondents' stated *ANA* was higher than in hypothetical decisions. However, when analyses were conducted to infer their *ANA* the investigators determined that *ANA* was lower for real decisions than for hypothetical decisions, especially for the price attribute. This suggests that respondents exaggerate how much they actually ignore attributes, and the attributes of a decision may be more carefully considered for real decisions than for hypothetical ones.[81]

The type of study identifying Cognitive factors in this review varied. 4 factors, *Normative Beliefs, Vividness of Decision Process, Attribute Non-Attendance, and Degree of Risk Aversion*, were identified in studies collecting empirical data. 9 factors were identified in literature reviews through cited propositions: *Social Desirability Bias, Predicted vs. Experienced Utility, Abstract Construals, Forecasting Emotions, Deliberative Mindset, Dispositions, Naturalism Bias, Omission Bias, and Subjective Beliefs*. Four factors were identified by more than one type of study: *Degree of Certainty* (meta-analysis, empirical data, and cited proposition), *Hot/Cold Empathy Gap* (empirical data and cited proposition), *Familiarity with a Good* (empirical data and cited proposition), and *Discounting* (empirical data and cited

proposition). This category contained the three factors that were most frequently identified in our review: *Discounting, Degree of Certainty, and Hot/Cold Empathy Gap.*

The cognitive factors that were shown the be positively associated with consistency between hypothetical and real-world decisions were: *Vividness of the Decision Process, Degree of Certainty, Dispositions, Familiarity with a Good, Naturalism Bias, Omission Bias, and Discounting.* The factors that were shown to be negatively associated with consistency between hypothetical and real-world decisions were: *Normative Beliefs, Social Desirability Bias, Predicted vs. Experienced Utility, Abstract Construals, Forecasting Emotions, Deliberative Mindset, Hot/Cold Empathy Gap, Attribute Non-Attendance, Degree of Risk Aversion, and Subjective Beliefs.*

We identified one study that disagreed with the *Degree of Risk Aversion* factor. This was a lottery choice experiment paired with a survey about risky health-related behaviours. This study concluded that there was no difference between participants' risk preferences experimentally and in real life.[82]

Table 4 Definitions of the Cognitive factors (n=17)

Name of Factor	Definition of Factor	Factor Statement	Example Scenario	References
Discounting	The degree to which a decision-maker prefers a smaller, immediate, reward to a larger, delayed, reward.	Discounting rates in hypothetical decisions are consistent with real life decisions	Compared decisions between hypothetical and real delay discounted monetary rewards.[83]	11,18,66–71,84
Degree of Certainty	Whether the decision-maker is certain that their hypothetical decision is the same as would be their real-world decision	A high degree of certainty about a hypothetical decision makes it more likely to be consistent with a real decision	Compared hypothetical and real decisions to pay for an asthma management programme.[45]	32,42,45,73–76

Hot/Cold Empathy Gap	Whether a person is experiencing the visceral factors caused by being in an emotionally provocative (i.e. hot) state	The impact of visceral reactions is less well appreciated for hypothetical decisions compared to real life ones.	Compared hypothetical and real decisions to pay money to avoid having to eat disgusting foods.[46]	[20–23,46]
Degree of Risk Aversion	Whether the decision-maker prefers options that are less likely, but have greater rewards, to options that are definite but have smaller rewards	Risk aversion is underestimated in hypothetical decisions compared to real life ones.	Compared hypothetical and real monetary lottery decisions.[12]	[12,30,77]
Familiarity with a Good	Whether the decision-maker has direct experience with the object of the decision	Familiarity with a good results in more agreement between hypothetical and real life decisions.	Compared hypothetical and real decisions about making financial donations.[57]	[43,57]
Normative Beliefs	Whether the decision-maker is thinking about what family and others would think about their decision	Normative Beliefs are less likely to be activated for hypothetical decisions than for real ones	Compared hypothetical and real decisions to contribute to a scholarship fund.[41]	[41]
Abstract Construals	Whether the decision-maker is thinking about the general features, versus thinking about the specifics of a decision	Abstract construals of problems are employed more in hypothetical decisions than in real life	Review of psychological perspectives related to hypothetical and real decision making.[44]	[44]
Forecasting Emotions	Whether people are predicting the emotions they think they would feel when making a decision, versus actually experiencing those emotions	Emotions in hypothetical decisions are forecasted more than in real life decisions	Review of psychological perspectives related to hypothetical and real decision making.[44]	[44]

Mindset	Whether participants are evaluating the pros and cons of different options, versus focussing on information that is useful for them to complete a selected goal	The deliberative mindset is used more in hypothetical decisions than in real life	Review of psychological perspectives related to hypothetical and real decision making.[44]	[44]
Predicted vs. Experienced Utility	Whether the decision-maker is predicting how they would feel towards a health state or actually experiencing that health state	Feelings towards future health states are predicted inaccurately in hypothetical decisions compared to real life experience.	Compared hypothetical and real medical treatment option decisions.[18]	[18]
Naturalism Bias	Whether the decision is affected by the (incorrect) beliefs that things that are natural are inherently good	Naturalism bias affects hypothetical and real life decisions equally	Compared demonstration of bias in hypothetical scenarios to real-world health decisions.[18]	[18]
Omission Bias	Whether the decision is affected by a preference for bad outcomes that result from neglecting to act versus bad outcomes that result from action	Omission bias affects hypothetical and real life decisions equally	Compared demonstration of bias in hypothetical scenarios to real-world health decisions.[18]	[18]
Social Desirability Bias	Whether the decision-maker's decision is affected by their desire to conform to the experimenter's beliefs	Social desirability affects hypothetical decisions more than real decisions	Compared hypothetical and real monetary gamble decisions.[47]	[47]

Dispositions	Whether decision-makers have dispositions about thinking that allow them to accept strategies to make hypothetical reasoning resemble real-world reasoning	Being in a state of prolonged doubt increases correlation between hypothetical and real life decisions	Literature review comparing approaches to the study of reasoning.[27]	[27]
Subjective Beliefs	The decision-maker's beliefs about the probability of the outcomes of the decision	Asking about subjective beliefs increases consistency between hypothetical decisions and real life	Compared hypothetical and real monetary lottery decisions.[62]	[62]
Vividness of Decision Process	Whether the decision-maker imagines how they would feel about the consequences of their decision	Thinking about the consequences of hypothetical decisions makes them more like real life decisions	Compared hypothetical and real monetary multi-stage gamble decisions.[79]	[79]
Attribute Non-Attendance	Whether the decision-maker neglects to fully consider some of the attributes of a decision	More attention is paid to the attributes of a real decision than a hypothetical one	Compared hypothetical and real choice preference for purchasing apples.[81]	[81]

Table 5 describes the Motivation factors identified in the review. These are factors that describe how people's motivations for making a decision affect how hypothetical decisions correlate to real-world ones. We identified 7 factors falling in to this category.

Real Consequences refers to whether or not the decision has consequences that will affect the decision maker. These consequences could be the provision of monetary rewards, as in the studies by Hinvest & Anderson and Murphy & Stevens, or the requirement to purchase an item for real, as in the study by Müller et al.[60,76,83] The important feature in the definition of this category is not the

nature of the consequence; it is the fact that the decision has a real impact on the decision-maker's future. Experiencing *Real Consequences* of decisions has been shown to be positively associated with consistency between hypothetical and real-world decisions. For example, Hinvest & Anderson examined whether there was an effect of real (vs. hypothetical) rewards on binary choices between sure-thing (vs. probabilistic) and immediate (vs. delayed) rewards. Although they did not find an impact of providing real rewards on choices between definite or probabilistic outcomes, they did conclude that people made less impulsive choices when there were real rewards offered on the delay discounting task compared to when the rewards were purely hypothetical.[83] Murphy & Stevens conducted a literature review and found similar results.[76] They concluded that when there are only hypothetical consequences, responses to contingent valuation surveys (which assess participants' willingness to pay for goods) can be variable. Offering real consequences in these willingness to pay surveys decreases this variability.[76]

Self-Image relates to how the decision affects the way in which the decision-maker perceives themselves. People seek to preserve a positive-self-image, and to do so they sometimes make decisions that are in-line with their ethical views.[85] Therefore, whether a good has an ethical component or not (e.g. the provision of public goods as opposed to private goods) affects whether a decision is related to the preservation of a positive-self-image.[64,76,85] Decisions related to self-image have been shown to be negatively associated with consistency between hypothetical and real-world decisions. Johansson-Stenman & Svedsater conducted a study where hypothetical and real willingness to pay was compared for a moral good (e.g. donation to the World Wildlife Fund) and an amoral good (e.g. a restaurant voucher).[85] They found that participants' hypothetical willingness to pay was higher than their real willingness to pay for moral goods but not for amoral goods. The authors explain that this is because people are incentivized to overstate their willingness to pay for goods that are concordant with their ethical beliefs as this helps them to preserve their self-image.[85]

List et al found similar results through their meta-analysis. They concluded that the association between hypothetical and real willingness to pay is closer for amoral goods than for moral goods.[64]

Potential for Self-Gain relates to the participants' goals. This factor is defined as decisions where participants can choose to maximise self-gain instead of another option. The decision-maker may choose to focus on *self-gain* even if this carries an increased risk of losses, as in the Etchart-Vincent & L'Haridon study, or causes them to harm other people, as in the FeldmanHall study.[26,65] The *Potential for Self-Gain* has been shown to be positively associated with consistency between hypothetical and real-world decisions. Etchart-Vincent & L'Haridon concluded that participants were less risk averse when the situation involved only hypothetical money because there was no *potential for any gains* from the decisions.[65] FeldmanHall had similar results, concluding that participants' behaviour did differ when the potential rewards were real as they were more willing to inflict pain on others than they stated they would be in the purely hypothetical setting.[26]

High Stakes Rewards is a factor related to the size of the decision-maker's potential reward. This factor may be relevant when the real-world decision has large potential rewards that the hypothetical decision has no possibility of providing to the decision-maker.[62,86] High stakes rewards have been shown to be negatively associated with consistency between hypothetical and real-world decisions. Two literature reviews examined whether the size of monetary incentives offered had any impact on the association between the hypothetical and real decision.[62,86] Harrison concluded that there was no difference in relative risk aversion between hypothetical and real decisions when the rewards are small but relative risk aversion between the two increases with the size of the real monetary rewards.[62] Murphy et al's meta-analysis found consonant results. They found that the primary factor affecting the difference between hypothetical and real decisions was the magnitude of the hypothetical value.[86]

Self-Benefitting Consequences refers to decision-makers' desire to choose options that benefit themselves instead of helping others. This factor was identified in a decision where participants had the option to be generous to others, or to keep more money for their own self-benefit.[15] Providing the decision-maker with real *self-benefiting consequences* has been found to be positively associated with consistency between hypothetical and real-world decisions. Day found that when making moral decisions about allocating money unequally participants' behaviour differed when considering the situation completely hypothetically and when participants actually had to interact with those to whom they would need to distribute the money.[15] However, Day found that this increased generosity was reduced when there were actual self-benefiting consequences to be obtained.[15] This suggests that for a hypothetical decision to be highly correlated with the real decision, not only should those who may be affected by a decision be made more salient by interacting with the decision maker, if the decision involves potential self-benefiting consequences then these should be made salient as well.

Engagement refers how much the way the decision is presented motivates participants to be engaged in decision-making rather than satisficing the question.[87] Decision-makers' *engagement* in their decision has been shown to be positively associated with consistency between hypothetical and real-world decisions. Hainmueller et al compared 5 different types of vignettes describing various people where participants were tasked with deciding whether they would accept the person described if they were a new immigrant.[87] The results of the assessment of these vignettes were then compared to the results of real-world naturalization referendums.[87] They concluded that the paired vignette designs, where two profiles were presented next to each other and the participant is tasked with accepting or rejecting each of the applicants, produced results that were most consistent with the real-world decisions.[87] This was attributed to the fact that participants in these settings were more engaged in their decisions and therefore were not simply satisficing the questionnaire.[87]

Concern measures the degree to which the decision-maker feels concerned about the task. Concern about a task has been shown to be positively associated with consistency between hypothetical and real-world decisions. Irwin et al sought to develop a study that would allow them to assess the effects of hypothetical vs. real rewards while taking into account method, length of the task, and consequences as confounders. They achieved this using a Vickrey, or sealed bid, auction with hypothetical and real money groups – varying the length of the experiment to test for the effects of boredom.[88] They found that auctions with hypothetical money had an increased number of very low and very high bids over auctions with real money and they attributed this to a decreased concern about the task.[88] Therefore, when participants are not concerned about the task (e.g. when there are no real rewards available) their decisions will be less correlated to real-world decisions.

Among the motivation factors, 4 were identified in studies collecting empirical data: *Self-Benefitting Consequences, Potential for Self-Gain, Engagement, and Concern*. Three factors were identified in more than one type of study: *High Stakes Rewards* (meta-analysis, cited proposition), *Real Consequences* (empirical data, cited proposition), and *Self-Image* (meta-analysis, empirical data, and cited proposition).

The motivation factors that were shown to be positively associated with consistency between hypothetical and real-world decisions were: *Real Consequences, Potential for Self-gain, High Stakes Rewards, Self-Benefiting Consequences, Engagement, and Concern*. Only one factor was shown to be negatively associated with consistency between hypothetical and real-world decisions: *Self-Image*.

We identified two studies that disagreed with the Motivation factors. One study involved binary choices between larger delayed or risky options and smaller immediate or certain options. In contrast to the *High Stakes Rewards* factor, this study found that hypothetical and real rewards were statistically equivalent.[89] The other study disagreed with the *Self-Image* factor. This was a meta-analysis

and it concluded that evaluating private (versus public goods) leads to more consistency between hypothetical and real-world decisions.[42]

Table 5 Definitions of Motivation factors (n=7)

Name of Factor	Definition of Factor	Factor Statement	Example Scenario	References
Real Consequences	Whether the decision has real consequences for the decision-maker	Having real consequences makes hypothetical decisions more closely predict real-world ones	Compared hypothetical and real monetary gamble decisions.[83]	[60,74,76,83]
Self-Image	Whether the decision relates to the decision-maker's self-image (e.g. related to their ethical beliefs)	Decisions related to self-image show less consistency between hypothetical and real-world decisions	Compared hypothetical and real decisions about purchasing moral (donation) vs. amoral goods (restaurant vouchers).[62]	[64,76,85]
Potential for Self-Gain	Whether there is the possibility that a decision provides the decision-maker with real rewards	Potential for self-gain affects real-world decisions more than hypothetical decisions	Compared hypothetical and real monetary lottery decisions.[65]	[17,65]
High Stakes Rewards	Whether significant rewards for a decision are available	When large incentives are available, risk aversion is higher for real-world decisions than for hypothetical decisions	Compared hypothetical and real monetary lottery decisions.[62]	[62,86]
Self-Benefitting Consequences	Whether there is an enticement towards choosing an option for self-benefit to the detriment of others	The potential for self-benefitting consequences makes people behave more selfishly in real-world decisions than they do in hypothetical ones.	Compared hypothetical and real decisions about unequally distributing money.[15]	[15]

Engagement	Whether the decision-maker is engaged in the decision, versus satisficing	High engagement makes hypothetical decisions more like real-world ones	Compared decisions about accepting potential immigrants hypothetically to real-world naturalization referrendums.[87]	[87]
Concern	Whether the decision-maker is concerned about the consequences of the decision	Concern about the task is lower for hypothetical decisions compared to real-world ones	Compared hypothetical and real sealed-bid auctions.[28]	[88]

Table 6 describes the Participant Characteristics factor identified in the review. We identified only 1 factor in this category. This factor describes how participants' resemblance to the real-world target population affects how hypothetical decisions correlate to real-world ones.

The two studies that identified this factor measured whether *samples matched the target population* on two levels. First, Hainmueller et al and Kesternich suggest that samples match the target population when they are drawn from a representative sample of the target population and face the same decision.[87,90] Second, Hainmueller et al suggest that the sample should be representative of the target population demographically, as they compared hypothetical decisions of a probability sample of participants drawn from the target population itself to hypothetical decisions made by a convenience sample of students.[87] The studies found that when samples are used that match the target population, consistency between hypothetical and real-world decisions has been shown to be increased. Hainmueller et al found that when they administered hypothetical vignettes to a sample of the target population, they made decisions that resembled real-world decisions.[87] However, they found that when providing the same tools to a sample of students they were not able to replicate the same results. This lead them to conclude that it is important for survey samples to closely match the target population in order for hypothetical decisions to resemble real life ones.[87]

Kesternich et al conducted a study comparing hypothetical and real life decisions of older Americans choosing between health plans.[90] They reached a conclusion resembling that of the Hainmueller study, hypothetical choice data can be similar to real-world decisions when it is obtained from a representative sample who is actively facing the real decision being evaluated.[90] Both of these studies involved the collection of empirical data.

Table 6 Definition of Participant Characteristics Factor (n=1)

Name of Factor	Definition of Factor	Factor Statement	Example Scenario	References
Samples Match Target Population	Whether the sample of people making the hypothetical decision closely resembles the population that faces the real-world decision	When participants in hypothetical situations resemble the target real-world group, hypothetical decisions are more consistent with real-world ones.	Compared hypothetical and real insurance plan decisions.	[87,90]

Figure 4 Fields from which each factor was identified. Cells are green if a factor was identified by at least one study from a particular field.

Factor	Field				
	Behavioral Economics	Psychology of Reasoning	Social Psychology	Health-Related Behavior	Neuroscience
Moral Reasoning Scores	■	■	■		
Openness to Experience Trait	■	■	■		
Personality Traits	■		■		
Cultural Differences	■	■	■		
Age	■				
Education	■				
Cognitive Control	■				■
Personal Relevance		■			
Cognitive Ability	■	■			
Study Procedure	■	■			
Space for Mental Simulation		■	■		
Salience	■	■	■		
Definition of Decisions		■			
Uncertainty	■				
Fundamental Attribution Error	■	■	■		
Framing Effect	■				
Time for Reflection		■	■		
Discounting	■	■		■	

Degree of Certainty					
Hot/Cold Empathy Gap					
Degree of Risk Aversion					
Familiarity with a Good					
Normative Beliefs					
Abstract Construals					
Forecasting Emotions					
Mindset					
Predicted vs. Experienced Utility					
Naturalism Bias					
Omission Bias					
Social Desirability Bias					
Dispositions					
Subjective Beliefs					
Vividness of Decision Process					
Attribute Non-Attendance					
Potential for Self-Gain					
Self-Image					
High Stakes Rewards					
Real Consequences					
Self-Benefitting Consequences					

Engagement					
Concern					
Samples Match Target Population					

3.2 Cognitive Interviews

Cognitive Interviews were conducted with seven decision aid developers. These decision aid developers were asked to talk about their experience in developing DAs in the context of a pilot survey. This survey was designed to elicit opinions about the relevance of the factors identified in our review to the DA development process

3.2.1 Themes elicited from Interviews: External validity issues

Decision aid developers generally did not report thinking about the external validity of their tool when they are developing them in hypothetical settings. Instead, they focused on issues of internal validity. In other words, they reported generally not being concerned with whether hypothetical decisions predict real-world decisions, but instead the focus on whether their tool is easy to understand, the questions are clear to the participants, etc. Many factors did not seem to be concepts that decision aid developers had previously considered. One participant said that: "I think decision aid developers […] develop a decision aid, they'll try to use it in a population that's like as close as they can to real-life. And then when they test it in the end in real life they're hoping it's going to give good results. I'm not sure they would look at all of these [factors]."

Because the decision aid developers were unfamiliar with many of the terms, many of the factors required additional clarification and were difficult to explain in a survey format. Examples of factors that were unclear to interviewees include: *Salience, Self-Image,* and *Abstract Construals*. Thinking

about which factors were relevant to the decision aid developer's own area of research and which were relevant in general was also difficult for respondents; they often said that they would have answered both questions the same way. When interviewees were not able to understand a factor based on the factor statement alone, TH provided further information about the factor. This discussion often helped clarify the factors for the respondents, and they were able to provide suggestions about how to modify the language in order from them to be easier to understand in the decision aid development context. For example, although the factor *Familiarity with a Good* was not clear, decision aid developers were better to able understand when we used the term "options" instead of "goods."

Although the decision aid developer's inexperience with the factors identified made it hard for them to think about the factors' relevance, the interviewees generally felt that a framework describing the important factors would be useful for their practice. One respondent said that they felt the framework would be very useful for them, but that they would need to be provided with the evidence around the factors before being comfortable using the framework.

3.2.2 Relevant Factors

Respondents identified several factors as being particularly relevant to the decision aid development process. These factors required very little discussion or clarification from the interviewers. Relevant factors included: *Personal Relevance* to the participant, providing statements about uncertainty around outcomes (*Uncertainty)*, and ensuring that information is presented the same way in hypothetical and real-world settings (*Study Procedure)*. Participants were able to relate some of these factors back to their own experience with decision aid development, one participant mentioned that when statements about *Uncertainty* around outcomes are provided it is helpful to patients' decision making.

3.2.3 Factors with Unclear Relevance

Some factors were difficult for decision aid developers to determine relevance to decision aid development, either because they had not previously thought about them or because they were not familiar with the concepts described. These factors included: *Subjective Beliefs, Salience, Predicted vs. Experienced Utility, and Mindset*. Some participants indicated they were not familiar with the evidence around certain outcomes. For example, one participant mentioned that they were often answering "Unclear/Uncertain" when they wanted to disagree about the relevance of a factor but did not have any evidence to support their belief.

When discussing the factor *Familiarity with a Good* (framed as: "Participants are familiar with the outcomes being evaluated"), not all participants understood what "outcomes" meant in the context of decision aid development. In the literature from which this factor was drawn, it was clear whether the decision maker had direct experience or not with the good which they were evaluating.[43,57] We clarified that in the context of decision aid development this factor referred to whether the participant was familiar with the different treatment options they were being asked to evaluate. Some respondents said that this factor was relevant as they noticed that hypothetical decisions correlate more highly when participants knew more about the disease and its treatment options. Others said that this factor was not relevant because the decision aid is designed to help participants become more familiar with the outcomes so it did not matter whether or not the participant was familiar with the outcomes at the outset.

The factor *Dispositions*, was framed as "Participants experience some uncertainty/decision conflict when making their decision. Respondents confused this factor with *Degree of Certainty* ("Participants are certain about their responses"). The factor *Dispositions* was drawn from a literature review that discussed participant's dispositions about thinking, and it suggested that reasoning will

more closely resemble the real-world if participants are made to accept experiencing a prolonged state of doubt when making their decision.[27] This is contrasted with *Degree of Certainty* which refers to whether the decision-maker feels certainty that the decision they make in the hypothetical setting is the same as the decision the would make in the real-world.[32,42,45,73–76] Participants were often unable to distinguish between these two factors, even when additional explanation was given.

3.2.4 Irrelevant Factors

Respondents were able to identify some factors that they believed were likely not relevant to the decision aid development process: *Cognitive Ability, Hot/Cold Empathy Gap, Potential for Self-Gain, and High Stakes Rewards.* Respondents' agreed that participants' cognitive ability was less relevant to decision aid development as these tools are developed to be easily understood by all. In the literature in which the *Cognitive Ability* factor was identified, the participants with higher cognitive ability were using a "puzzle-solving" approach (i.e. trying to find the best answer to "solve" the problem) as opposed to describing what they would actually do in real-life.[31] It is not clear whether this would occur in the context of decision-aid development as well.

Respondents also did not believe *Potential for Self-Gain* were relevant to decision aid development. Many interviewees noted that incentives are generally not offered to participants evaluating decision aids and obtaining monetary incentives is generally not part of the decision. One interviewee mentioned that they cannot offer financial incentives to participants evaluating their decision aids, as the ethics review board generally does not allow this.

Finally, respondents generally did not believe that the emotional reactions elicited by the real-world decision would impact how consistent hypothetical decisions would be to the real-world. One participant said that while they did not have any data to support their decision, they found that generally decision aids elicited similar emotional reactions regardless of whether they were being

evaluated hypothetically or in the real-world. They stated that in their practice the participants evaluating the decision aid in hypothetical situation are often patients who, while not making the decision immediately themselves, experience similar emotions to the real-world patients who are actually making the decision. However, this may depend on the type of decision being made. It may be difficult for those in hypothetical situations to accurately imagine the emotions one would feel when actually making a decision with highly impactful consequences (e.g. a decision about a lung transplant as opposed to a decision about taking antibiotics for an ear infection).

Two interviewees believed that when decision aid are designed to help with high-stakes, life-or-death decisions it is difficult for people to imagine hypothetically how they would behave if such a situation were to occur for real. This is somewhat similar to the *High-Stakes Rewards* factor, which states that decision-makers are not able to effectively imagine rewards in hypothetical decisions when they are of large value to them. These interviewees felt that the proposed framework would only be helpful for decision aids that dealt with less risky consequences, such as decisions about medication with minimal negative side effects.

The results of this process demonstrated that this topic may not be appropriate for survey responses. This is because many of the factors identified were unfamiliar to the interviewees and were too difficult to explain using only the factor statements developed for a survey. Instead, this topic should be explored by further qualitative work.

CHAPTER 4: DISCUSSION

4.1 OVERVIEW

The relationship between hypothetical and real-world decisions has important implications for the design of health services research. When testing new interventions, health services researchers inevitably need to conduct preliminary work, while the UK MRC framework recommends practices that would involve even more of this work being done.[7] These preliminary studies often involve people and settings that differ in the extent to which they resemble the real-world.

Without a clear understanding of what aspects of these different study populations and settings affect the validity of preliminary studies, the efficiency of research will be reduced. Investigators will be less likely to use cheaper, more efficient studies without guidance on how to design valid studies in hypothetical settings. This could lead to the implementation of more large scale health services trials before they have had the chance to benefit from preliminary data testing the mechanisms underlying the interventions, leading to more large-scale negative trials, and large amounts of wasted resources.

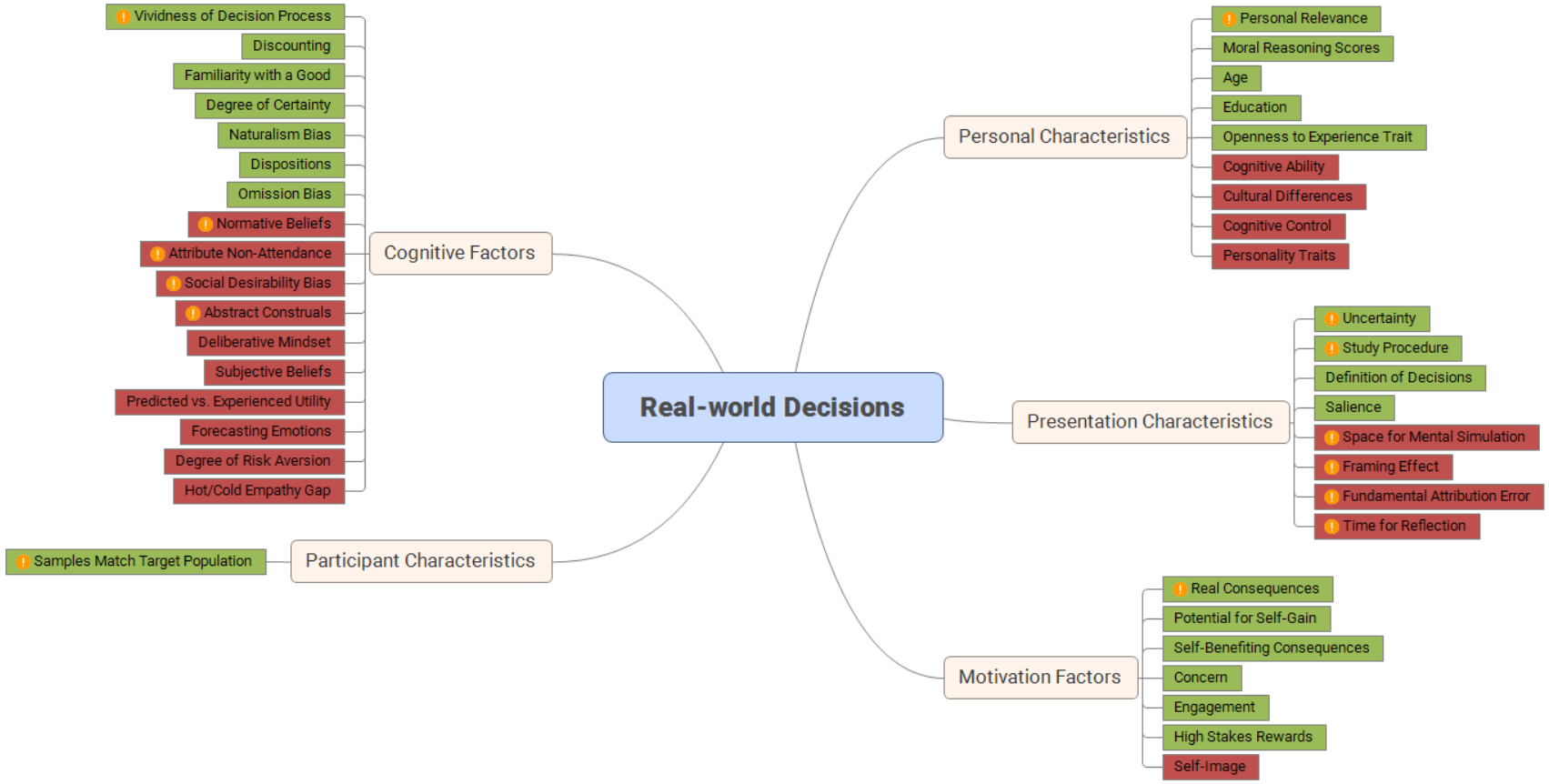
This thesis sought to identify the factors affecting the relationship between hypothetical and real-world decisions. We conducted a systematic concept review to identify factors relevant to the association between hypothetical and real-world decisions. We then organized the identified factors and themes into a framework with the intention of helping researchers understand when hypothetical decisions would generalize to the real-world.

The factors identified in our review came from a wide variety of fields of research including Behavioral Economics, the Psychology of Reasoning, and Social Psychology. These factors have not

been systematically applied to the health services research context, but such an exercise would provide valuable information about their applicability to this area. For example, a recent study measuring physicians' responses about appointment booking used both a self-reported survey and simulated patient approaches.[91] This study found that participants are likely to overestimate their likelihood of booking appointments for new patients.[91] Presseau et al (2012) found similar discrepancies in a study comparing physician and nurse self-reported past diabetes-related behaviours to their simulated behaviours based on responses to hypothetical scenarios.[92] The insights obtained through our study could be applied to such health services research studies by systematically assessing the effect of identified factors such as: a hypothetical decision that involves interactions with simulated patients similar to what the decision-maker would experience in the real-world (i.e. *Study Procedure, Salience, etc.)* and asking the decision-maker about their *Degree of Certainty* after they have made their decision. The exercise of assessing the impact of these factors would help investigators design better hypothetical settings to assess real-world clinician decisions related to these questions.

4.2 Framework

Figure 5 Framework depicting the 42 factors thought to affect the consistency between hypothetical and real-world decisions, grouped into 5 categories[1]



[1] Factors in green in this framework have been found to be positively associated with consistency between hypothetical and real-world decisions. Factors in red have been found to be negatively associated with hypothetical and real-world decisions. Factors marked with an orange exclamation mark may be relevant in the decision aid development context.

Figure 5 depicts our framework describing all factors found to affect consistency between hypothetical and real-world decisions. Connected to the central concept of real-world decisions are five domains corresponding to our five categories of factors. Investigators seeking to design a study with hypothetical components with an eye to yielding decisions that match what would happen in the real-world should consider each of these five domains during development. In the current version of this framework, all five domains are roughly equidistant from the centre, as our systematic concept review did not provide us with the means to assess the relative importance of the different domains. Each domain is comprised of a range of factors. Factors coloured green were shown to be positively associated with consistency between hypothetical and real-world decisions, whereas rectangles coloured red were shown to have be negatively associated.

Factors marked with an orange exclamation point were identified as potentially relevant to decision-aid testing through the process of cognitive interviewing experts in that field. Through these interviews we found that in the context of decision aid development these factors might be more relevant for informing developers about how to test decision aids independently of the actual decision made (i.e. Whether the information presented is balanced, how options are weighted.) However, because most of the factors came from literature where the outcome being evaluated was whether the hypothetical decision actually predicted the real-world decision, the framework might be more natively employed to other research questions with this goal (e.g. Predicting whether someone will participate in a clinical trial.)

While all five categories should always be considered, the relevant factors within the categories will likely differ from one context to the next. For example, factors such as Naturalism and Omission Bias may be relevant considerations in an area such as vaccine development where it has been shown that people who show these biases when making hypothetical decisions will show

them when making real-world decisions as well.[18] However, they may not be relevant considerations in other areas of health services research where they have not been shown to be relevant cognitive factors.

The factors that investigators identify as relevant considerations should be accounted for during the design process of their study. Accounting for the Participant Characteristics category could involve assessing whether it is feasible and ethical to sample participants from the target population of decision-makers, and to do so if possible. Accounting for Personal Characteristics could involve considering which of the factors might be characteristics of the participants in the study and how these will affect consistency between hypothetical and real-world decisions, or taking steps to discuss important personal characteristics with participants to ensure that they do not decrease consistency between hypothetical and real-world decisions. Accounting for Presentation Characteristics could involve considering how to modify the intervention so that the way the decision is presented to the participants involves more of the factors thought the increase consistency with real-world decisions and less of the factors thought to decrease consistency. Accounting for Cognitive Factors could involve considering which of the factors might be affecting the participant making the hypothetical decision, to discuss these factors with the participant in order to render them explicit, or to take steps to ensure that any negative effect of these factors on the consistency between hypothetical and real-world decisions is reduced. Accounting for the Motivation category could involve considering what can be done to ensure that the participant feels the same motivation to make their decision in the hypothetical setting as they would in the real-world.

4.3 Systematic Concept Review

Our systematic search yielded 59 studies that identified one or more factors thought to be related to the relationship between hypothetical and real decisions.

4.3.1 The Role of Financial Incentives in Modeling health care decisions

Of the final 59 studies, most (48/59) came from the Behavioral Economics literature, which has clearly led exploration on this topic. These studies involved a wide range of techniques that all concerned how people make real-world financial decisions. Several studies assessed people's willingness to pay for various products.[41,45,46,53,57,60,73,75,81,85,88,90] This field lends itself to the study of the relationship between hypothetical and real-world decisions because of characteristics such as applicability for many stakeholders and a clear modeling approach. By understanding the factors that affect whether hypothetical willingness to pay decisions accurately predict real-world decisions, for-profit enterprises could more easily conduct studies to assess how to sell their products more effectively. This field also provides a clear approach to modeling other types of decisions because unlike health or moral decisions, decisions involving large sums of money can often be modeled by decisions involving small amounts of money. For example, risk aversion can be studied using imaginary or small amounts of money to assess how people make decisions about whether or not to accept risky gambles.[12,30,77] By determining whether hypothetical or small monetary rewards are valued the same way as important large monetary rewards, investigators can determine whether the conclusions about risk aversion drawn from these studies are relevant when people are facing risky decisions that are actually important to them.

Across the various Behavioral Economics studies, money has served conceptually different roles. For example, in the FeldmanHall (2012) study monetary incentives are used to assess people's willingness to sacrifice self-gain to prevent harm to others.[17] In such a case, money is used as a

moderating factor on the decision the study participant is being asked to make, instead of the subject of the decision itself. This means that investigators use monetary rewards as a way of making a decision more "real" for participants, even when the purpose of the study is to assess factors affecting hypothetical/real-world decisions in contexts other than decisions about money. A further example is the study by Gold et al, which used monetary incentives to operationalize the trolley problem by asking participants whether to divert the path of a moving ball in order to decide whether a real donation was made to 5 orphans or to only 1. This allowed them to assess differences between hypothetical and real-world moral decisions for a problem that traditionally involves deciding between allowing a train to kill a group of innocents or intervening so that it kills only one.[56]

In health decisions, there can be serious ethical concerns around the use of monetary incentives.[93] One of these concerns revolves around the fact that decisions about medical treatment should be made based on informed consent, and there is a concern that offering monetary incentives to patients who opt to receive a certain treatment could be coercive and harmful to the integrity of the medical profession.[93] Because of this, it is unclear how monetary incentives serve to model other factors that are more salient in real-world medical decision making, for example the desperation a patient might feel to receive a last-hope treatment.[94] Additional research to solve the problem of the generalizability of monetary incentives to health care decisions could be beneficial to better understand the factors that affect the association between hypothetical and real-world decisions.

4.3.2 Relevance of the Factors to The decision Aid Development Context

For the purposes of our own project, we chose to focus on the context of decision aid development as a case-study for the application of our framework. This area of health services

research was selected because it is a domain where the factors affecting the validity of results from hypothetical settings are unclear and untested, people are routinely surveyed about their preferences for future health states they have not yet experienced, and where real decisions involve an important emotional impact.

We initially planned to conduct a survey of decision aid developers in order to understand the relevance of the factors identified in the systematic concept review to the decision aid development context. However, after conducting an initial set of cognitive interviews with decision aid development experts in order to pilot the survey, it became clear that researchers in this field are not particularly experienced with assessing or thinking about factors relevant to the association between hypothetical and real-world decisions. Although they become interested in the potential for a framework describing relevant factors, it was difficult for them to understand the factors' applicability based on the survey alone; instead, detailed explanations of the factors under consideration were often required. As a result, we decided to continue conducting cognitive interviews and to use these detailed insights as an initial source of data for the relevance of our framework in the field of decision aid development.

Interviewees generally agreed that *Samples Match Target Population* was a relevant factor. Interviewees suggested that current practice may be simply to try and use a population and setting that is as close as possible to real life in order to avoid the problem of consistency between hypothetical and real-world decisions.

Interviewees also generally agreed that Presentation Characteristics may be relevant, except for the factors *Salience*. It seems that decision aid developers agreed that the content of the hypothetical patient decision aid and the way information is presented should resemble how it is presented in the real-world. Information about uncertainty should be provided, the framing should be consistent

between hypothetical and real situations, and time provided for reflection should resemble the real-world. The decision should be described with enough detail to reduce space for mental simulation. The interviewees were unclear about the meaning of *Salience* of a decision. While some did agree that it is important to present possible harms to participants, they felt that this was a necessary component of decision aids and so it would not affect differences between decisions in hypothetical and real-world contexts. We attempted to make this factor clearer to interviewees by framing *Salience* in terms of the decision-maker's emotional arousal, but they remained uncertain about this point. This suggests that developing measures of the *Salience* of a decision for someone using a decision aid could be an area for future work.

Five of the 17 Cognitive Factors were identified as potentially relevant: *Vividness of Decision Process, Normative Beliefs, Attribute Non-Attendance, Social Desirability Bias,* and *Abstract Construals.* In some cases, such as abstract construals, respondents were not familiar with the terminology used to describe the factors in the literature, but they did agree with the sentiment of the factor. Respondents disagreed that the hot/cold empathy gap was relevant in decision-aid development. This was a somewhat surprising result as this was a factor that was identified in several studies and that was consistently shown to affect the relationship between hypothetical and real-world decisions.[20–23,46] One respondent explained that the hypothetical decision often elicits the same emotional reactions in non-patients as they observe in patients who are making the decision in the real-world. This could simply be due to the nature of health decisions that was not a factor in other decisions examining the hot/cold empathy gap.[20–23,46] Because interviewees were unfamiliar with many of the concepts identified in this category, and it was the category that was the most commonly addressed in the literature, future work in assessing the importance of these factors in the context of decision aid development should be undertaken.

Of the Personal Characteristics and Motivation Factors, only *Personal Relevance* and *Real Consequences* were identified as being potentially relevant. It was somewhat surprising to us that interviewees generally did not agree that Personal Characteristics were relevant; because proxies are often used in decision aid development we had thought that this category of factors would be found relevant. The finding may be due to the fact that interviewees generally were not especially concerned with external validity when employing proxies to develop decision aids, and therefore did not have strong opinions about which Personal Characteristics would be important. This may be an important area for future research, as understanding the relevant Personal Characteristics could provide decision aid developers with the ability to draw conclusions about the external validity of data drawn from decisions made by proxies.

With regards to the factor of *Personal Relevance*, the interviewees may have interpreted this statement more broadly than how it was identified in the literature. The study that identified the factor of *Personal Relevance* defined personally relevant decisions as moral decisions that involve people with whom the decision-maker had close, long-term relationships.[21] Many interviewees agreed with the statement that "the problem has personal relevance to the participant" is relevant to decision aid process without describing in detail how they interpreted "personal relevance". Although interviewees generally agreed that it is important for real consequences of the decision to be explicitly described to participants in order for hypothetical decisions to be consistent with real life, they did not agree that providing monetary rewards or incentives to decisions was a relevant consideration. In the context of decision aid development *Real Consequences* relate to the health implications of a decision. Interviewees agreed that these health implications need to be explicitly presented to participants, but they explained that monetary incentives are not used as a proxy for the *Real Consequences* of a decision in the context of their work, as they have been in other fields.

Two interviewees brought up the idea that perhaps hypothetical decisions are not consistent with real-world ones in the context of decision aid development because it is too difficult for people to imagine how they would decide between options in a life-or-death situation. While this idea is consistent with the factor of *Predicted vs. Experienced Utility*, which suggests that people are poor are predicting future health states, we did not find any data evaluating how people think about very important, life or death, decisions in hypothetical and real-world situations. Comparing how our framework affects life-or-death and less risky decisions in the context of decision aid development may be an important area for future research.

4.3.3 Summary of Factors

We were able to classify the factors into five categories: Participant Characteristics, Personal Characteristics, Presentation Characteristics, Cognitive Factors, and Motivation. The sample may closely resemble the target population (i.e. Participant Characteristics) while the individuals making the decision may have certain characteristics that are related to more or less consistency between hypothetical and real-world decisions (i.e. Personal Characteristics). The way that the decision is presented and the characteristics of the decision itself (i.e. Presentation Characteristics), the way that the participants think about and make their decision (i.e. Cognitive Factors), as well as the participants' motivation for making a decision (i.e. Motivation) could all affect consistency between hypothetical and real-world decisions.

4.3.4 Participant Characteristics

The category of Participant Characteristics included only one factor that was identified by more than one study: Matching samples to the target population increases the consistency between hypothetical and real-world decisions.[87,90] This factor has important implications for hypothetical research. It suggests that as long as the study's participants are drawn from a representative sample

then the data obtained from hypothetical decisions are more likely to match the real-world. Both of these studies described "representative" as a sample of people coming from the target population who have, or will, make the decision being evaluated.[87,90] The key difference between the Participant Characteristics category and the Personal Characteristics category is that the former deals with the participants at the group and sample levels, while the latter deals with participants at the individual level.

Sometimes, it may infeasible to conduct hypothetical studies with a representative sample of the target population (i.e. people who have or will make the target decision). The study may involve treatment for a rare disease and a sufficiently large sample may be impossible to obtain. In those cases, the other categories of factors serve to indicate other characteristics that may be important to match in order for the hypothetical decision to be representative of the real-world decision.

4.3.5 Personal Characteristics

We identified 9 Personal Characteristics, factors that describe how characteristics of the individual decision-maker affect the consistency between hypothetical and real-world decisions. Lower *Moral Reasoning Scores*, higher *Openness to Experience Trait* scores, older *Age*, higher *Education*, and increased *Personal Relevance* were found to be positively associated with consistency between hypothetical and real-world decisions. Higher scores of other *Personality Traits*, *Cultural Differences* between Asian and Western people, increased *Cognitive Control*, and higher *Cognitive Ability* were shown to be negatively associated with consistency between hypothetical and real-world decisions.

Many of these personal characteristics are difficult to modify directly. However, they might suggest intervention tailoring efforts in order to improve the consistency between hypothetical and real-world decisions. For example, corrective entreaties, which involve explaining some of the relevant personal characteristics that affect hypothetical decisions and urging participants to act as if

they were facing the consequences of their decision, have been used in the behavioral economics literature.[41] Hypothetical decisions have been shown to become more representative of real-world decisions after such entreaties.[41] For example, a corrective entreaty could explain to participants that lower levels of *Moral Reasoning* are often used when making real-world decisions, and it could encourage them to think about the consequences of the decision in terms of punishments, rewards, and implications for their family as opposed to more abstract moral principles.[15,48,51] Investigators could also present the problem that people with higher *Cognitive Ability* may try to find the "correct" solution to the problem instead of answering as they would in the real-world.[31] They could remind participants that there may not be a "correct" answer, and that instead they should try to answer as realistically as possible, as if they were going to face the consequences of their decision.[41]

Investigators could further improve the consistency between hypothetical and real-world decisions by modifying other aspects of their study. Participants could be asked to answer demographic questions, as done by Mjelde et al, in order to identify older and more highly educated participants whose hypothetical decisions have been shown to be more consistent with their real-world decisions.[57] When hypothetical moral dilemmas are given to participants, investigators could attempt to make these more personally relevant to the participants by ensuring that they are related to individuals with whom the decision-maker has ongoing long-term relationships.[21] In order to apply the Cultural Differences factor, investigators could avoid using samples of Western participants to draw conclusions about Asian decision-makers and vice-versa for decisions related to intervening in the natural course of an event.[56]

4.3.6 Presentation Characteristics

We identified 8 Presentation Characteristics, factors that describe how characteristics of the decision itself and the way that it is presented to the decision maker affect how hypothetical

decisions correlate to real-world ones. Increased *Salience* of the decision, presenting *Uncertainty* around outcomes, ensuring consistency of the *Study Procedure*, and *Defining the Decision* to encourage participants to use the models/search approach (i.e. reasoning about problems using mental representations of the decision-making situation, using the provided information to test and reach conclusions) have been found to be positively associated with consistency between hypothetical and real-world decisions. Describing the hypothetical situation from the point of view of an observer (i.e. *Fundamental Attribution Error)*, the *Framing Effect*, increasing the *Space Available for Mental Simulation*, and greater provided *Time for Reflection* in the hypothetical setting have been found to be negatively associated with consistency between hypothetical and real-world decisions.

Three presentation characteristics provide investigators with clear guidance that by ensuring that contextual details about the decision are presented, the hypothetical decision will be more consistent with the real-world. *Salience* refers to the fact that seeing the people who may be harmed by the consequences of the decision makes the hypothetical decision correlate more highly with the real-world decision. This is consistent with the idea of *Space for Mental Simulation* as both factors are related to the idea that hypothetical decisions correspond more closely to the real-world when the decision is presented with rich contextual detail. The *Uncertainty* factor implies that investigators should ensure that details about the uncertainty around outcomes are provided to participants when they are available.[62] When participants are presented with such detail about the uncertainty around outcomes, their hypothetical decisions are more consistent with the real-world. This may be because the explicitly presented details may help to correct subjective, possibly incorrect, beliefs that the decision-makers have about the uncertainty of outcomes.[62] Investigators could apply these factors by ensuring that the decision they describe to participants includes personally relevant details that will make the decision seem more "real" to participants. This could include using sensually rich language

in the description of the situation, providing real details about the uncertainty of the outcomes, or using multimedia to increase the salience and decrease mental simulation.

A number of factors identified in this category suggest that it is important that the procedural elements of the study (e.g. how much time they have to think before making the decision, how the decision is described, how the information is ordered) resemble real-life. This is important for investigators as their study of participants' hypothetical decisions may not resemble real-world ones if they are made based on early study materials that do not closely resemble the final version or if participants are allowed more time for reflection than what would be experienced in the real-world. The way that descriptions and questions are phrased should resemble the final version of the tool. It is important to note that if the decision that the decision-maker is being asked to make is affected by positive/negative framing then this may have a greater impact on hypothetical decisions than real decisions.

4.3.7 Motivation Factors

We identified 7 Motivation Factors that describe how people's motivations for making a decision affect how hypothetical decisions correlate to real-world ones. Nearly all of the identified motivation factors were positively associated with consistency between hypothetical and real-world decisions: having *Real Consequences* for the decision, having *Potential for Self-Gain*, when there are *High Stakes Rewards*, when there are *Self-Benefitting Consequences*, when the participants are *Engaged* in the decision, and when they are *Concerned* with the decision. The only motivation factor that was found to be negatively associated with consistency between hypothetical and real-world decisions was whether the decision involved participants' *Self-Image*.

Three factors identified in the motivation category provide signals to investigators that participant's motivation, and the correlation between hypothetical and real-world decisions, may be

affected by the type of decision they are being asked to make. Several authors suggested that when the decision to be made is one that causes participants to seek to preserve their *Self-Image*, hypothetical decisions will less closely resemble real-world decisions.[64,76,85] We identified the factors of *Engagement* in the decision and *Concern* about the task that might also affect whether hypothetical decisions will accurately represent real-world decisions.[87,88]Irwin et al suggest that *Concern* about the task can be increased by providing the participants with real instead of hypothetical financial incentives.[88] Likewise, *Engagement* can be increased by changing the vignette type to a paired vignette design which was found to be a more engaging method than single profile vignettes.[87] Therefore, investigators may want to ensure that their study design is engaging and activates concern for hypothetical participants, as it would in the real-world. If the participants have experience with the topic of the hypothetical decision, it may be more engaging and of more concern to them than to those who do not have experience with the topic, which may increase the consistency between their hypothetical decisions and real-world ones.

Other factors identified in the motivation category may help provide investigators with guidance about rewarding participants in their studies. These factors suggest that participants' behaviour may differ depending on whether there are tangible consequences (e.g. monetary rewards) for the decision. The authors found that when real rewards are available, participants were less likely to make impulsive decisions and that they were more risk averse, selfish, and willing to inflict pain on others. There is some evidence to suggest that these effects may be related to the magnitude of the potential reward. It is important to note that for the studies that identified all of these factors, the real condition involved a financial decision.[15,17,60,62,76,83,86] Therefore, future work should explore the implications of using financial incentives as a proxy for other consequences in health-related decisions.

4.3.8 Cognitive Factors

We identified 17 Cognitive Factors that relate to how the ways people think and make decisions affect how hypothetical decisions correlate to real-world ones. Increased *Vividness of the Decision Process*, higher *Degree of Certainty* about hypothetical decisions, *Dispositions* about thinking, increased *Familiarity with a Good*, the *Naturalism Bias,* the *Omission Bias*, and *Discounting* were all shown to be positively associated with consistency between hypothetical and real-world decisions. The effect of *Normative Beliefs*, *Social Desirability Bias*, predicting instead of experiencing utility of a health state (i.e. *Predicted vs. Experienced Utility*), the use of *Abstract Construals*, *Forecasting Emotions*, the deliberative *Mindset*, the *Hot/Cold Empathy Gap*, not paying sufficient attention to the attributes of a decision (i.e. *Attribute Non-Attendance*), differences in risk preferences (i.e. *Degree of Risk Aversion)*, and *Subjective Beliefs* were shown to be negatively associated with consistency between hypothetical and real-world decisions.

Four of the cognitive factors provide clear guidance to investigators seeking to increase the consistency between hypothetical and real-world decisions. The *Degree of Certainty* factor relates to the fact that when people state that they are more certain about their hypothetical decision, that decision and the real-world one are more likely to be correlated. When decision-makers are not sure about their hypothetical decision, it is less likely to correlate with their real-world decision. This appears to be a factor that would be straightforward to implement into practice. Asking the research participants if they are "definitely sure" their decision would correspond to their real-world decision and excluding those who answer that they are not allows investigators to know which hypothetical decisions may be more likely to correlate with real-world decisions. [32,42,45,73–76]

Investigators may be able to take into account the *Hot/Cold Empathy Gap* and the *Vividness of the Decision Process* in order to improve consistency between hypothetical and real-world decisions.

Decision-makers in "cold" hypothetical states are not experiencing the physical, visceral, sensations related to emotions they would feel making the decision in the real-world are therefore are not easily able to predict the influence that these visceral factors would have on their real decision. However, it may be possible to simulate these visceral factors by encouraging participants to think about the consequences of their decision, thereby making the process more vivid. This has been shown to increase the correlation between hypothetical and real-world decisions in settings such as paying insurance to avoid an aversive outcome.[46] Additionally, investigators may be able to simulate the visceral factors that would affect participants when making the actual decision. This could help to activate the emotions participants would feel and thereby make hypothetical decisions more closely predict real-world ones.

The literature on *Degree of Risk Aversion* may also offer guidance to investigators studying risky decisions. Participants may be less risk averse in hypothetical settings compared to real life.[12,30,77] Investigators have counteracted this decreased risk aversion in hypothetical settings by employing a corrective entreaty. This involves assessing the effect of explicitly telling their research participants that data show that people are generally more risk averse in real life than when they are making hypothetical decisions. Asking participants making hypothetical health care decisions to take this fact into account might make hypothetical decisions correlate more highly with real-world ones.[41,57]

4.4 LIMITATIONS & STRENGTHS

A number of limitations of this work warrant consideration. The purpose of this systematic concept review was to systematically search the literature for studies that identify factors related to our topic. To that end, we initially attempted to design broad search and screening strategies to identify all such factors. Such a broad strategy yielded a non-specific search that included considerable repetition of concepts, because there are whole literatures around some of the concepts (e.g. delay discounting, risk aversion). As with many concept reviews, relevant concepts were not always included in the titles and abstracts of papers, making our original approach require a number of full text searches that was beyond our resources. We adapted our approach by using both a targeted systematic database search and a snowball search to identify other related articles that cited or were cited by articles we knew to be important. This approach may have sacrificed sensitivity of the search in order to achieve feasible specificity. As a result, we cannot be certain that all relevant factors were identified.

The systematic concept review guided us in the production of a framework to identify factors affecting the association between hypothetical and real-world decisions. We did not conduct quality assessment of the data underlying the identified factors. However, we did observe that not all of the factors were equally well supported by evidence. Some factors, such as *Self-Image and High Stakes Rewards* were supported by several studies and meta-analyses.[64,86] Other factors, such as *Discounting*[11,18,66–71,84], *Degree of Certainty* [32,42,45,73–76], and *Hot/Cold Empathy Gap*[20–23,46], were consistently identified in several different studies. Many factors that we identified, however, were only identified in a single study (e.g. *Attribute Non-Attendance*[81], *Subjective Beliefs*[62], *Dispositions*[27], etc.) so it is unclear whether they will be consistently relevant across a range of decisions. Therefore, some factors may

not remain in future iterations of the framework if additional research shows that they are not consistently supported by evidence.

This framework is not a theory; we do not specify the causal pathways along which the identified factors are expected to affect the association between hypothetical and real-world decisions. Nor does it indicate which factors and categories are the most important or how they relate. Instead, it is intended to outline the major factors relevant to this area. Nonetheless, it is an important first step. While we cannot yet tell researchers with precision how well their hypothetical study will product real-world decisions, this framework can help to guide future work in this area.

One issue that arose after piloting our survey with decision aid developers was that the survey was complex for them to understand, and we found that they would not have been able to complete the questions without additional information provided by the interviewer. Therefore, we found that asking decision aid developers about the factors affecting hypothetical and real-world decisions through the format of a survey was infeasible. This meant that we were unable to investigate the broader range of opinions that we were anticipating collecting through a survey.

Based on our interviews, decision aid developers piloting their tools in hypothetical settings are often more interested in the internal validity of the tool (e.g. whether the questions measure what they are supposed to) rather than the external validity (e.g. whether the decisions made in the hypothetical settings are generalizable to the real-world decisions). This suggests that decision aid development may not have been the ideal field to select as a first attempt to employ our framework, since the piloting process often was not primarily concerned with mimicking real-world decision makers. Nevertheless, some interviewees did say that their approach to designing hypothetical settings was to simply try to make them resemble the real-world as closely as possible. This respondent thought that decision aid developers do not think about all of the factors that we

identified. The interviewees said that that they would be interested in an empirically supported framework that outlined the factors that predict the relationship between hypothetical and real-world decisions as this could help them design better hypothetical settings. Therefore, although other fields in the context of health services may be more concerned with ensuring that hypothetical decisions are predictive of real-world decisions, we believe that the framework can still be employed in the context of decision aid development.

It is also important to highlight some of the strengths of this work. As previously existing theory around the association between hypothetical and real-world decisions is limited and diffuse, it was helpful to take an initial exploratory step to assembling the data. The multidisciplinary nature of our work adds to its strength, as we attempted to consolidate the contributions of various fields that have taken different approaches to examining this topic. It is also important to highlight the systematic approach that we took to designing our search strategy as well as collecting and analysing data, as this contributes to ensuring that we assembled a representative picture of the literature and minimised bias.

4.5 CONTRIBUTIONS

Clinical Epidemiology involves studying the diagnosis, prognosis, and management of human disease.[4] It is important when conducting clinical epidemiological research to remember that patients experience not only physiological manifestations of an illness, but also behavioral, social, psychological, and emotional manifestations. This is called the total illness experience.[95] Interventions, such as decision aids, can improve the total illness experience by helping patients and care providers make quality decisions that conform to patients' values and best evidence. As with medications and other interventions that serve to improve the total illness experience, it is important

that complex interventions involving decision making are supported by evidence before full scale implementation in the real-world.[7] One way of collecting such evidence is through the use of hypothetical settings. The framework developed in this thesis serves to indicate the factors that may be important for investigators to consider when collecting data from hypothetical settings in order for decisions made in these settings to be predictive of real-world decisions. This will enrich the design of studies and the use of evidence in clinical epidemiological interventions designed to better patients' total disease experience by improving how important decisions are made.

To the best of the authors' knowledge, this is the first systematic concept review of factors that affect the association between hypothetical and real-world decisions. The framework that was produced by this review seems to be the most comprehensive description of the factors that appear in the literature related to this subject. It spans literature from Behavioral Economics, Psychology of Reasoning, Social Psychology, Health-related Behaviour, and Neuroscience. It will serve as a useful point of departure for investigators who are interested in understanding whether the design of their study will allow them to draw conclusions about the real-world based on hypothetical decisions. This is important as the correlation between hypothetical and real-world decisions is becoming increasingly recognized as an important issue in health care.[7] This is true in the field of decision aid development, but also in other areas of research such as decisions about participating in clinical trials, and piloting knowledge translation interventions about feedback.

Our systematic search identified studies about the association between hypothetical and real-world decisions that were conducted in a variety of different fields. Relatively few of these studies involved the field of health research. Therefore, the results of our study contribute to the knowledge by highlighting the fact that further research must be conducted in health care and health services settings in order to ensure that the factors identified are relevant in those contexts. We discovered

that decision aid developers were interested in developing a better understanding of these factors and they did believe that they would be able to apply a framework designed to present the issues relevant to the association between hypothetical and real-world decisions in the decision aid development context. To this extent, we were able to obtain some initial impressions of the relevant factors from decision aid developers. The results of these cognitive interviews contribute to the understanding of how hypothetical settings are employed in the decision aid context among other fields, and therefore to a better understanding of their value.

4.6 Future Directions

Little research has addressed which factors affect the association between hypothetical and real-world decisions in health fields. As a next step, health services investigators should continue to build upon our framework by including empirical analyses of the impact of the factors we have identified in any future studies involving comparisons of hypothetical and real-world decisions. By assessing the degree to which the factors from our framework affect decisions in the health care context we can further improve the framework by increasing our understanding of the strength of the association of the various factors.

Researchers may also apply our framework to randomized studies assessing how participants' decisions related to health topics differ between hypothetical and real-world settings. Such studies could randomize participants to three groups: an unmodified hypothetical decision group, a hypothetical decision group that was designed using our framework, and a real-world decision group. The investigators could compare decisions made in the unmodified hypothetical decision group and those made in the group for which our framework was employed to the real-

world decisions in order to measure the impact of our framework on making hypothetical decisions more closely correlate to real-world decisions in the health context.

Future investigations could be conducted employing the interview protocol that we developed in other health service research settings. The questions would only need to be slightly modified in order to make them specific to the chosen setting. Conducting such interviews would facilitate the use of our framework, as it could be re-designed to include only factors or descriptions about factors that are specific to the needs of that setting.

It would be potentially beneficial to conduct an economic analysis to assess the cost effectiveness of using hypothetical settings versus real-world settings when developing decision aids or other such interventions. An economic analysis comparing the use of hypothetical versus real-world settings during decision aid development would involve measuring the total cost of the development stage in each. This approach could be combined with other studies that assess whether applying our framework to hypothetical settings improves the correlation between hypothetical and real-world decisions; such a study would compare the incremental cost effectiveness ratios of a decision aid developed with and one without the use of our framework. Once the total cost of developing the decision aids has been measured, their effectiveness can be compared when applied in the real-world. The Cantor et al framework for evaluating the cost-effectiveness of patient decision aids suggests assessing effectiveness based on life years saved because there are established willingness to pay thresholds for this metric.[96] Other options include assessing the patients' knowledge, or their sense of feeling informed and clear about their values using decision conflict scales.[97] Once the incremental cost effectiveness ratios are calculated, future decision aid developers could choose which setting would be most feasible for the development of their tool, taking the trade-off between cost and effectiveness into consideration. This type of study would

need to be done with a variety of preference-sensitive decisions, as the cost of development could vary depending on the type of decision as well as the target patient population.

These cost effectiveness issues extend beyond the area of decision aid development. Improving our understanding of the factors that affect consistency between hypothetical and real-world decisions means that we can improve the efficiency of health services research on a broad scale. A deeper understanding of these factors will assist investigators in conducting valid pilot studies and hypothetical data collection to inform the development of new interventions before proceeding to more expensive real-world trials. The UK Medical Research Council framework encourages investigators developing new complex health interventions to gather more information in order to ensure that the intervention will have the desired effect. This includes steps such as ensuring that the intervention has a theoretical basis, assessing the empirical evidence from previous studies, and conducting qualitative studies to identify the relevant components of a tool.[7] This sort of supportive mechanistic evidence is an important feature of trials to develop new drugs, and our framework will help investigators collect supportive evidence in complex intervention design as well.

4.7 Overall Conclusions

In this thesis, we described the importance of understanding the factors that affect the association between hypothetical and real-world decisions. Forty-two factors were identified through a systematic concept review. These were classified into five fields: Participant Characteristics, Personal Characteristics, Presentation Characteristics, Cognitive Factors and Motivation. These factors were identified by authors of studies from various fields such as Behavioral Economics, the Psychology of Reasoning, and Social Psychology. Comparatively few studies were conducted in health settings.

We were able to achieve our goal of integrating these factors into a framework that describes the association between hypothetical and real-world decisions. Investigators can use this framework when designing studies in hypothetical settings in order to identify factors that maybe relevant for their research and that, if taken into account, may make hypothetical decisions more consistent with real-world decisions.

We were able to use this framework to develop a pilot survey. This survey was designed to obtain information from decision aid developers about the factors affecting the correlation between hypothetical and real-world decisions that are relevant considerations when designing decision aids. Using the pilot survey as a springboard for discussion about our framework through cognitive interviews, we learned that ensuring consistency between hypothetical and real-world decisions is not a primary goal of decision aid studies in hypothetical settings. Because of this, decision aid developers were generally unfamiliar with the factors that we had identified, but many resonated as being potentially important and useful. The implications of the factors for the development of decision aids and other interventions required detailed discussion rather than the short descriptions that we had originally envisioned as part of a survey. Further work in this area should likely involving more qualitative approaches, such as interviews, rather than surveys. Nevertheless, the decision aid developers surveyed did believe that this framework could be a useful tool to help them understand the factors that are truly important to designing studies in hypothetical settings.

Now that an initial framework has been developed, future research should prioritize empirical studies that assess which factors are relevant in various fields of health services research. Such studies could take the form of interviews involving the use of the framework for specific health decisions, assessments of how hypothetical and real-world interventions are affected when the factors identified in the framework are taken into account or further reviews to evaluate the

strength of particular factor associations, for example. Because most of our factors come from Behavioral Economics research, it is unclear whether the identified factors are applicable to health services research or if they were caused by the particularities of Behavioral Economics research. Future investigations should be conducted in health services areas besides decision aid development in order to modify the framework to suit those specific areas. The development of these specialized frameworks will help researchers understand which factors are relevant to ensuring that hypothetical decisions are externally valid measures of real-world decisions.

REFERENCES

1. Bondemark L, Ruf S. Randomized controlled trial: The gold standard or an unobtainable fallacy? *Eur J Orthod*. 2015;37(5):457-461. doi:10.1093/ejo/cjv046.

2. Schulz KF, Grimes DA. Allocation concealment in randomised trials: Defending against deciphering. *Lancet*. 2002;359(9306):614-618. doi:10.1016/S0140-6736(02)07750-4.

3. Health Services Research: An Evolving Definition of the Field. *Health Serv Res*. 2002;37(1):15-17. doi:10.1111/1475-6773.01020.

4. Parfrey PS, Barrett BJ, eds. *Clinical Epidemiology*. Vol 1281. New York, NY: Springer New York; 2015. doi:10.1007/978-1-4939-2428-8.

5. Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. *Fundamentals of Clinical Trials*. Fifth Edit. Cham: Springer International Publishing; 2015. doi:10.1007/978-3-319-18539-2.

6. Colquhoun HL, Brehaut JC, Sales A, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci*. 2013;8:66. doi:10.1186/1748-5908-8-66.

7. Campbell M. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321(7262):694-696. doi:10.1136/bmj.321.7262.694.

8. Stacey D, Légaré F, Nf C, et al. Decision aids for people facing health treatment or screening decisions (Review) Decision aids for people facing health treatment or screening decisions. *Cochrane Libr*. 2014;(1). doi:10.1002/14651858.CD001431.pub4.Copyright.

9. Flory J, Emanuel E. Interventions to improve research participants' understanding in

informed consent for research: a systematic review. *JAMA*. 2004;292(13):1593-1601. doi:10.1001/jama.292.13.1593.

10. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol*. 2002;10(1):42-48. doi:10.1186/1471-2288-10-1.

11. Madden GJ, Begotka AM, Raiff BR, Kastern LL. Delay discounting of real and hypothetical rewards. *Exp Clin Psychopharmacol*. 2003;11(2):139-145. doi:10.1037/1064-1297.11.2.139.

12. Holt CA, Laury S. Risk Aversion and Incentive Effects. *SSRN Electron J*. 2002;92(5):1644-1655. doi:10.2139/ssrn.893797.

13. Gold N, Pulford BD, Colman AM. Do as I say, don't do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. Andreoni Bartels, Berg, Bolton, Cote, Fehr, Fehr, Foot, Frith, Gold, Gold, Gold, Greene, Greene, Greene, Greene, Greene, Greene, Hare, Hare, Harvey, Hauser, Houser, Hsu, Koritzky, Kurzban, Liao, McGuire, Mikhail, Nadelhoffer, Navarrete, Osberg, Petrinovi A, ed. *J Econ Psychol*. 2015;47:50-61. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2015-11172-005.

14. Patil I, Cogoni C, Zangrando N, Chittaro L, Silani G. Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. Avramova Christensen, Cushman, Cushman, Cushman, Dawson, FeldmanHall, Figner, Gilbert, Greene, Greene, Greene, Greene, Haidt, Hauser, Huebner, Kant, Kobayashi, Lombrozo, Mikahil, Mill, Mook, Moretto, Navarrete, Paxton, Rovira, Schwitzgebel, Suter, Tassy, B, ed. *Soc Neurosci*. 2014;9(1):94-107. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2013-45514-009.

15. Day RWC. Relations between moral reasoning, personality traits, and justice decisions on hypothetical and real-life moral dilemmas. *Diss Abstr Int Sect B Sci Eng*. 1998;58(12-B):6795. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1998-95012-014.

16. Kohlberg L. Stages of moral development. *Moral Educ*. 1971:23-92.

17. FeldmanHall O, Mobbs D, Evans D, Hiscox L, Navrady L, Dalgleish T. What we say and what we do: The relationship between real and hypothetical moral choices. Ajzen Batson, Brown, Cikara, Crockett, Cushman, Dunn, Foot, Gilbert, Greene, Greene, Greene, Haidt, Haidt, Isen, Kahane, Kang, Knutson, Kohlberg, Landsberger, Masserman, Milgram, Moll, Russell, Tassy, Teper, Tetlock, Tversky, Tversky, Van Boven, Waldmann B, ed. *Cognition*. 2012;123(3):434-441. doi:10.1016/j.cognition.2012.02.001.

18. Chapman GB. The psychology of medical decision making. Arkes Boyd, Bromily, Cairns, Chapman, Chapman, Chapman, Chapman, Chapman, Chapman, Chapman, Chapman, Chapman, Coppola, Covinsky, Dawson, DeKay, Ditto, Elstein, Fagerlin, Fuchs, Gilbert, Gold, Jansen, Kahneman, Kahneman, Kirby, Lerner, Llewellyn-Thomas, L B, ed. *Blackwell Handb Judgm Decis making*. 2004:587-603. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2004-19929-029.

19. Kang MJ, Rangel A, Camus M, Camerer CF. Hypothetical and real choice differentially activate common valuation areas. Anderson Becker, Blumenschein, Carson, Chandon, Crespi, Cummings, Deichmann, Delgado, Diamond, Duvernoy, Erk, Friston, Gitelman, Gottfried, Greene, Greene, Hare, Hariri, Hopkins, Infosino, Jamieson, Johannesson, Kahneman, Keeter, Kim, Knutson, List, Litt A, ed. *J Neurosci*. 2011;31(2):461-468.

doi:10.1523/JNEUROSCI.1583-10.2011.

20. Kühberger A, Schulte-Mecklenbeck M, Perner J. Framing decisions: Hypothetical and real. Arkes Beattie, Binswanger, Camerer, Camerer, Chapman, Ebbesen, Edwards, Grether, Grether, Harrison, Hertwig, Hertwig, Hogarth, Hsee, Josephs, Kirkpatrick, Klein, Knetsch, Kuhberger, Kuhberger, Levin, Levin, Lichtenstein, Lichtenstein, Loewenstein, Loewen B, ed. *Organ Behav Hum Decis Process*. 2002;89(2):1162-1175. doi:10.1016/S0749-5978(02)00021-3.

21. Skoe EEA, Eisenberg N, Cumberland A. The role of reported emotion in real-life and hypothetical moral dilemmas. Averill Baumeister, Blasi, Blum, Brody, Campos, Caprara, Colby, Crandall, Davis, Dunn, Eisenberg, Eisenberg, Eisenberg, Eisenberg, Eisenberg, Eisenberg, Eisenberg, Fabrigar, Ferguson, Ferguson, Geen, Gilligan, Golombok, Greene, Haidt, Haidt, Haidt, Hoffm B, ed. *Personal Soc Psychol Bull*. 2002;28(7):962-973. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2002-15107-009.

22. Teper R, Tullett AM, Page-Gould E, Inzlicht M. Errors in Moral Forecasting: Perceptions of Affect Shape the Gap Between Moral Behaviors and Moral Forecasts. *Personal Soc Psychol Bull*. 2015;41(7):887-900. doi:10.1177/0146167215583848.

23. van Nieuwenhuijzen M, Bijman ER, Lamberix ICW, et al. Do children do what they say? Responses to hypothetical and real-life social problems in children with mild intellectual disabilities and behaviour problems. Achenbach Coie, Coie, Coie, Crick, Cuperus, Dekker, Dodge, Dodge, Dodge, Dodge, Dodge, Dodge, Dodge, Dodge, Healy, Hunt, Leffert, Linna, Lochman, Lochman, Lochman, Matthys, Matthys, Matthys, Matthys, Mize, Mize, Nisbett, Orobio de Castro, Orobio de Castr A, ed. *J Intellect Disabil Res*. 2005;49(6):419-433. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN

=2005-04972-003.

24. Elwyn G, O'Connor AM, Bennett C, et al. Assessing the quality of decision support technologies using the International Patient Decision Aid Standards instrument (IPDASi). *PLoS One*. 2009;4(3). doi:10.1371/journal.pone.0004705.

25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097.

26. FeldmanHall O, Dalgleish T, Thompson R, Evans D, Schweizer S, Mobbs D. Differential neural circuitry and self-interest in real vs hypothetical moral decisions. Ajzen Amodio, Apperly, Baumgartner, Bechara, Berthoz, Blair, Borg, Bush, Casebeer, Cushman, Dalgleish, Davis, Etkin, Ewbank, FeldmanHall, Fletcher, Gilbert, Glenn, Greene, Greene, Haidt, Hassabis, Hassabis, Higgins, Holroyd, Hooker, Hume, Kang, Kedia, Kr A, ed. *Soc Cogn Affect Neurosci*. 2012;7(7):743-751. doi:10.1093/scan/nss069.

27. Galotti KM. Approaches to studying formal and everyday reasoning. Acock Alexander, Anderson, Baron, Baron, Baron, Barratt, Bartlett, Braine, Braine, Braine, Bruner, Bucci, Buscaglia, Case, Ceci, Ceci, Charness, Cheng, Cheng, Clark, Cole, Damer, Davidson, de Groot, Dellarosa, Detterman, Dewey, Ennis, Erickson, Evans, Fa A, ed. *Psychol Bull*. 1989;105(3):331-351. doi:10.1037//0033-2909.105.3.331.

28. Irwin J. Hypothetical and real consequences in experimental auctions for insurance against low probability risks. *J Behav* …. 1992;5(July 1989):107-116. http://onlinelibrary.wiley.com/doi/10.1002/bdm.3960050203/full. Accessed November 14, 2014.

29. Morgenstern R, Heldmann M, Vogt B. Differences in cognitive control between real and hypothetical payoffs. Battalio Blackburn, Bland, Bush, Camerer, Chatrian, De Neys, Farquhar, Folstein, Greiner, Harrison, Harrison, Harrison, Holt, Holt, Kang, Kuhberger, Paulus, Slovic, Smith, Van Veen, Wendt B, ed. *Theory Decis*. 2014;77(4):557-582. doi:10.1007/s11238-013-9408-x.

30. Slovic P. Differential effects of real versus hypothetical payoffs on choices among gambles. Coombs Feather, Kogan, Preston, Slovic, Slovic, Slovic E, ed. *J Exp Psychol*. 1969;80(3, Pt.1):434-437. doi:10.1037/h0027489.

31. Taylor MP. Bias and brains: Risk aversion and cognitive ability across real and hypothetical settings. Andersen Battalio, Beattie, Benjamin, Branas-Garza, Callen, Camerer, Camerer, Campitelli, Carson, Chilton, Costa-Gomes, Croson, Dave, Dohmen, Eckel, Eckel, Frederick, Gabaix, Green, Harrison, Harrison, Harrison, Harrison, Harrison, Hey, Holt, Holt, Johns A, ed. *J Risk Uncertain*. 2013;46(3):299-320. doi:10.1007/s11166-013-9166-8.

32. Vlaev I. How different are real and hypothetical decisions? Overestimation, contrast and assimilation in social interaction. Anderson Ariely, Birnbaum, Blumenschein, Blumenschein, Camerer, Champ, Chater, Crespi, Delgado, Diamond, Fantino, Fleming, Fudenberg, Greene, Harrison, Hertwig, Johannesson, Johnson, Kang, Kuhberger, Kusev, Laming, Ledyard, List, Little, Locey, Lonnqvist A, ed. *J Econ Psychol*. 2012;33(5):963-972. doi:10.1016/j.joep.2012.05.005.

33. *Mendeley* (for Windows) [computer program]. Version 1.12.1. New York, NY: Mendeley, Inc; 2015.

34. Hsieh H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. *Qual Health Res*. 2005;15(9):1277-1288. doi:10.1177/1049732305276687.

35. Gamer M, Lemon J, Fellows I, Puspendra S. *Various Coefficients of Interrater Reliability and Agreement* [computer program]. Version 0.84. The R Foundation; 2012.

36. Hornik K. The R FAQ. https://cran.r-project.org/doc/FAQ/R-FAQ.html. The R Foundation; 2016.

37. Dillman D. *Mail and Internet Surveys: The Tailored Design Method*. 2nd ed. New York: John Wiley & Sons; 2000.

38. Fowler CFJ. *Survey Research Methods (4th Ed.)*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.; 2009. doi:10.4135/9781452230184.

39. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res*. 2003;12(3):229-238. doi:10.1023/A:1023254226592.

40. Exum ML, Bailey D, Wright EL. False positive and false negative rates in self-reported intentions to offend: A replication and extension. Ajzen Armitage, Armitage, Bachman, Beck, Boster, Bouffard, Bouffard, Bouffard, Campbell, Collins, Exum, Exum, Fishbein, Gendreau, Gottfredson, Grasmick, Green, Higgins, Higgins, Higgins, Hubbard, Kahneman, Kahneman, Kennedy, Kim, Klepper, Laguilles, Lama A, ed. *J Crim Justice*. 2014;42(1):1-9. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-03731-002.

41. Ajzen I, Brown TC, Carvajal F. Explaining the Discrepancy Between Intentions and Actions: The Case of Hypothetical Bias in Contingent Valuation. Ajzen Ajzen, Ajzen, Ajzen, Ajzen, Armitage, Bender, Blackburn, Blumenschein, Blumer, Brown, Brown, Campbell, Champ, Cummings, Cummings, Cummings, Feather, Fishbein, Fishbein, Fishbein, Fisher, Fox, Gollwitzer, Hoehn, Hoehn, Kahneman, Krosnick, LaPiere, L A, ed. *Personal Soc Psychol Bull*.

2004;30(9):1108-1121. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2004-17669-002.

42. Little J, Berrens R. Explaining Disparities between Actual and Hypothetical Stated Values : Further Investigation Using Meta − Analysis. *Econ Bull*. 2004;3(6):1-13. doi:http://www.economicsbulletin.com/.

43. Sagristano MD. The effects of temporal distance on attitude-behavioral intention correspondence. *Diss Abstr Int Sect B Sci Eng*. 2005;65(12-B):6711. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2005-99012-181.

44. Eastwick PW, Hunt LL, Neff L a. External Validity, Why Art Thou Externally Valid? Recent Studies of Attraction Provide Three Theoretical Answers. *Soc Personal Psychol Compass*. 2013;7(June 2015):275-288. doi:10.1111/spc3.12026.

45. Blumenschein K, Johannesson M, Yokoyama KK, Freeman PR. Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. *J Health Econ*. 2001;20(3):441-457. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=11373840.

46. Kang MJ, Camerer CF. fMRI evidence of a hot-cold empathy gap in hypothetical and real aversive choices. *Front Neurosci*. 2013;7(7 JUN):104. doi:10.3389/fnins.2013.00104.

47. Camerer C, Hogarth R. The effects of financial incentives in experiments: A review and capital labor production framework. *J Risk Uncertain*. 1999;19(April):7-42.

doi:10.1023/A:1007850605129.

48. Trevethan SD, Walker LJ. Hypothetical versus real-life moral reasoning among psychopathic and delinquent youth. *Dev Psychopathol*. 1989;1(2):91-103. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1990-19801-001.

49. Lind G. Measuring Moral Judgment: A Review of "The Measurement of Moral Judgment" by Anne Colby and Lawrence Kohlberg. *Hum Dev*. 1989;32(6):388-397. doi:10.1159/000276491.

50. Thackrey M. The Measurement of Moral Judgment. *JAMA J Am Med Assoc*. 1988;260(13):1968. doi:10.1001/jama.1988.03410130176050.

51. Walker LJ, de Vries B, Trevethan SD. Moral stages and moral orientations in real-life and hypothetical dilemmas. *Caring voices women's moral Fram Gilligan's view*. 1994:342-358. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1995-97299-019.

52. Borghans L, Duckworth AL, Heckman JJ, ter Weel B. The Economics and Psychology of Personality Traits. *J Hum Resour*. 2008;43(4):972-1059. doi:10.1353/jhr.2008.0017.

53. Grebitus C, Lusk JL, Nayga RMJ. Explaining differences in real and hypothetical experimental auctions and choice experiments with personality. Almlund Ashton, Ben-Ner, Bettman, Borghans, Bouchard, Butler, Caprara, Carson, Champ, Cherry, Clark, Digman, Dohmen, Epstein, Eysenck, Eysenck, Ferguson, Fischhoff, Fleeson, Funder, Georgantzis, Gill, Goldberg, Gracia, Kagel, Kemp, Lachman, Lichtenstein, A, ed. *J Econ Psychol*. 2013;36:11-26. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc10&NEWS=N&A

N=2013-14650-002.

54. Lonnqvist J-E, Verkasalo M, Walkowitz G. It pays to pay-Big Five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. Balliet Baumeister, Berg, Camerer, Cawley, Cooper, Costa, Dawes, De Raad, Dohmen, Falk, Goerg, Harrison, Hirsh, Hofstee, Knez, Konstabel, Krueger, Lauriola, Levitt, Lonnqvist, Lonnqvist, Mischel, Paulhus, Podsakoff, Pothos, Preacher, Schwartz, Siegel, Sm B, ed. *Pers Individ Dif*. 2011;50(2):300-304. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc8&NEWS=N&AN=2010-26712-029.

55. Konstabel K, Lönnqvist J-E, Walkowitz G, Konstabel K, Verkasalo M. The "Short Five" (S5): Measuring personality traits using comprehensive single items. *Eur J Pers*. 2012;26(1):13-29. doi:10.1002/per.813.

56. Gold N, Colman AM, Pulford BD. Cultural differences in responses to real-life and hypothetical trolley problems. Abarbanell Baron, Bartels, Bond, Brehm, Chan, Chuangsheng, Cushman, Eberhard, Gawronski, Gold, Gold, Gold, Goldberg, Greene, Greene, Greene, Hare, Hauser, Henrich, Hsu, Kamm, Kirkland, Machery, Markus, Mikhail, Moore, Moore, Navarrete, Nisbett, O'Neill, A, ed. *Judgm Decis Mak*. 2014;9(1):65-76. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-03718-006.

57. Mjelde JW, Jin YH, Lee C-K, Kim T-K, Han S-Y. Development of a bias ratio to examine factors influencing hypothetical bias. *J Environ Manage*. 2012;95(1):39-48. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22115509.

58. Diamond A. Executive Functions. *Annu Rev Clin Psychol*. 2014;64:135-168. doi:10.1146/annurev-psych-113011-143750.Executive.

59. Dixon MR, Lik NMK, Green L, Myerson J. Delay discounting of hypothetical and real money: the effect of holding reinforcement rate constant. *J Appl Behav Anal*. 2013;46(2):512-517. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24114165.

60. Müller H, Kroll EB, Vogt B, Muller H, Kroll EB, Vogt B. Do real payments really matter? A re-examination of the compromise effect in hypothetical and binding choice settings. Beattie Braga, Calder, Chaiken, Chang, Chernev, Chernev, Chu, Cox, Dhar, Dhar, Dhar, Doyle, Gierl, Grether, Herrmann, Houghton, Hoyer, Hubbard, Huber, Huber, Kivetz, Knetsch, Kuhberger, Lehmann, Lin, List, McIntyre, Mitchell, Murphy, Nagpal, Nowlis, Pan, B, ed. *Mark Lett*. 2012;23(1):73-92. doi:10.1007/s11002-011-9137-2.

61. Ebbesen EB, Konecni VJ. On the external validity of decision-making research: What do we know about decisions in the real world. *Cogn Process choice Decis Behav*. 1980:21-45.

62. Harrison GW. Hypothetical Bias Over Uncertain Outcomes. *SSRN Electron J*. 2005;(October). doi:10.2139/ssrn.698501.

63. Levin IP, Chapman DP, Johnson RD. Confidence in judgments based on incomplete information: An investigation using both hypothetical and real gambles. *J Behav Decis Mak*. 1988;1(1):29-41. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1989-38838-001.

64. List, John, Gallet, C. What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? *Environ Resour Econ*. 2001;20(3):241-254. doi:10.1023/A:1012791822804.

65. Etchart-Vincent N, L'Haridon O. Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses. Abdellaoui Abdellaoui, Abdellaoui, Abdellaoui, Arkes, Arkes, Baltussen, Bardsley, Battalio, Beattie, Becker, Birnbaum, Bosch-Domenech, Bostic, Bruhin, Camerer, Chateauneuf, Clark, Cox, Cubitt, Davis, Etchart-Vincent, Etchart-Vincent, Fennema, Gibbons, Go A, ed. *J Risk Uncertain*. 2011;42(1):61-83. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc8&NEWS=N&AN=2011-00282-004.

66. Johnson MW, Bickel WK. Within-subject comparison of real and hypothetical money rewards in delay discounting. Ainslie Bickel, Chapman, Chapman, Chapman, Crean, Deluty, Green, Green, Green, Green, Green, Green, Green, Kirby, Kirby, Kirby, Kirby, Lancaster, Logue, Mazur, Meyer, Mischel, Mitchell, Myerson, Petry, Rachlin, Rachlin, Raineri, Richards, Thaler, Vuchini B, ed. *J Exp Anal Behav*. 2002;77(2):129-146. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2002-12588-001.

67. Lagorio CH, Madden GJ. Delay discounting of real and hypothetical rewards III: steady-state assessments, forced-choice trials, and all real rewards. Ainslie Baker, Bickel, Bickel, Coffey, Crean, Critchfield, Deluty, Feather, Giordano, Green, Green, Hyten, Irwin, Johnson, Kirby, Kirby, Lafferty, Lancaster, Lane, Levin, Logue, Madden, Madden, Madden, Madden, Mazur, Mazur, Meyer, Mitchell, Myerson, Myer A, ed. *Behav Processes*. 2005;69(2):173-187.

doi:10.1016/j.beproc.2005.02.003.

68. Lawyer SR, Schoepflin F, Green R, Jenks C. Discounting of hypothetical and potentially real outcomes in nicotine-dependent and nondependent samples. Ainslie Bickel, Bickel, Estle, Field, Green, Heatherton, Heil, Hinvest, Jikko, Johnson, Johnson, Kirby, Kollins, Lagorio, Lawyer, Lawyer, Madden, Madden, Madden, Mazur, McKerchar, McKerchar, Mitchell, Myerson, Myerson, Nasrallah, Odum, Odum, Odum, Ohmura B, ed. *Exp Clin Psychopharmacol.* 2011;19(4):263-274. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc8&NEWS=N&AN=2011-13221-001.

69. Bickel WK, Jones BA, Landes RD, Christensen DR, Jackson L, Mancino M. Hypothetical intertemporal choice and real economic behavior: Delay discounting predicts voucher redemptions during contingency-management procedures. Ainslie Axon, Baker, Bickel, Bickel, Bickel, Bickel, Bickel, Bickel, Bickel, Bjork, Bradford, Budney, Chesson, Coffey, Daugherty, Dom, Green, Green, Heil, Higgins, Hoffman, Johnson, Johnson, Kirby, Kirby, Logue, MacKillop, Mazur, McClure, Mitchell, Mitch A, ed. *Exp Clin Psychopharmacol.* 2010;18(6):546-552. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc7&NEWS=N&AN=2010-26426-010.

70. Madden GJ, Raiff BR, Lagorio CH, et al. Delay Discounting of Potentially Real and Hypothetical Rewards: II. Between- and Within-Subject Comparisons. Ainslie Bickel, Binder, Coffey, Crean, Critchfield, Feather, Green, Green, Hertwig, Hyten, Irwin, Isen, Johnson, Kagel, Kirby, Kirby, Kirby, Lafferty, Lancaster, Lane, Levin, Logue, Logue, Logue, Madden, Madden, Mazur, Mazur, Meyer, Mischel, Mitchell, My B, ed. *Exp Clin Psychopharmacol.*

2004;12(4):251-261. doi:10.1037/1064-1297.12.4.251.

71. Silva FJ, Gross TF. The rich get richer: Students' discounting of hypothetical delayed rewards and real effortful extra credit. Ainslie Critchfield, Davids, Eisenberger, Green, Hardy, Hattie, Hofer, Johnson, Logue, Logue, Logue, Martin, Merrell, Myerson, Pear, Pintrich, Rachlin, Rachlin, Rachlin, Read, Richards, Rothblum, Silva, Steinberg, Vanzile-Tamsen, Winne B, ed. *Psychon Bull Rev*. 2004;11(6):1124-1128. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2005-03508-020.

72. Green L, Myerson J, Holt DD, Slevin JR, Estle SJ. Discounting of delayed food rewards in pigeons and rats: Is there a magnitude effect? Arkes Chapman, Chapman, Grace, Green, Green, Green, Green, Hackenberg, Jackson, Kirby, Koopmans, Loewenstein, Madden, Mazur, Mazur, Myerson, Myerson, Odum, Rachlin, Raineri, Richards, Rodriguez, Simpson, Tobin, Tobin B, ed. *J Exp Anal Behav*. 2004;81(1):39-50. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2004-17658-005.

73. Blumenschein K, Johannesson M, Blomquist GC, Liljas B, O'Conor RM. Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation. *South Econ J*. 1998;65(1):169-177. doi:10.2307/1061360.

74. Harrison GW, Rutström EE. Chapter 81 Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In: *Handbook of Experimental Economics Results*. Vol 1. ; 2008:752-767. doi:10.1016/S1574-0722(07)00081-9.

75. Johannesson, Magnus, et al. Calibrating Hypothetical Willingness to Pay Responses. *J Risk Uncertain*. 1999;18(1):21-32.

76. Murphy JJ, Stevens TH. Contingent valuation, hypothetical bias, and experimental economics. *Agric Resour Econ Rev*. 2004;33(2):182-192.

77. Holt CA, Laury SK. Risk Aversion and Incentive Effects: New Data without Order Effects. *Am Econ Rev*. 2005;95(3):902-904. doi:10.1257/0002828054201459.

78. Trope Y, Liberman N. Construal-level theory of psychological distance. *Psychol Rev*. 2010;117(2):440-463. doi:10.1037/a0018963.

79. Beattie J, Loomes G. The impact of incentives upon risky choice experiments. *J Risk Uncertain*. 1997;14(2):155-168. doi:10.1016/S0167-6687(97)89209-8.

80. Erdem S, Campbell D, Hole AR. Accounting for Attribute-Level Non-Attendance in a Health Choice Experiment: Does it Matter? *Health Econ*. 2015;24(7):773-789. doi:10.1002/hec.3059.

81. Morkbak MR, Olsen SB, Campbell D. Behavioral implications of providing real incentives in stated choice experiments. Akerlof Alfnes, Andreoni, Balcombe, Bierlaire, Broadbent, Brownstone, Cameron, Campbell, Campbell, Campbell, Carlsson, Carlsson, Carlsson, Carlsson, Carson, Chang, Cummings, Gowdy, Gracia, Grebitus, Greene, Hensher, Hensher, Hensher, Hensher, Hensher, He A, ed. *J Econ Psychol*. 2014;45:102-116. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-55169-008.

82. Anderson LR, Mellor JM. Predicting health behaviors with an experimental measure of risk preference. *J Health Econ*. 2008;27(5):1260-1274. doi:10.1016/j.jhealeco.2008.05.011.

83. Hinvest NS, Anderson IM. The effects of real versus hypothetical reward on delay and probability discounting. Allen Chapman, Cherek, Dawe, Elliott, Elliott, Enticott, Evenden,

Green, Green, Herrnstein, Ho, Hyten, Johnson, Kheramin, Kheramin, King, Kringelbach, Lagorio, Lane, Madden, Madden, Madden, Mazur, Mobini, Mobini, Myerson, Navarick, Rachlin, Rachlin, Reyn B, ed. *Q J Exp Psychol*. 2010;63(6):1072-1084. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc7&NEWS=N&AN=2010-10502-005.

84. Wiseman DB, Levin IP, Evin PL. *Comparing Risky Decision Making under Conditions of Real and Hypothetical Consequences*. Vol 66. Netherlands: Elsevier Science; 1996:241-250. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1996-05489-001.

85. Johansson-Stenman O, Svedsater H. Self-image and valuation of moral goods: Stated versus actual willingness to pay. Akerlof Alpizar, Andreoni, Andreoni, Andreoni, Aquino, Ball, Baumeister, Benabou, Benabou, Benabou, Bernheim, Bodner, Bodner, Brekke, Brown, Cameron, Carlsson, Carlsson, Carlsson, Carson, Cummings, Cummings, Cummings, De Oliveira, De Oliveira, Forehand, A, ed. *J Econ Behav Organ*. 2012;84(3):879-891. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc9&NEWS=N&AN=2012-32591-013.

86. Murphy JJ, Allen PG, Stevens TH, Weatherhead D. A meta-analysis of hypothetical bias in stated preference valuation. *Environ Resour Econ*. 2005;30(3):313-325. doi:10.1007/s10640-004-3332-z.

87. Hainmueller J, Hangartner D, Yamamoto T. Validating vignette and conjoint survey experiments against real-world behavior. Alexander Arnold, Benitez-Silva, Berinsky, Bernstein, Bertrand, Diamond, Gelman, Green, Green, Hainmueller, Hainmueller, Hainmueller, Holbrook, Krosnick, Louviere, Louviere, Natter, Neill, Rossi, Schwarz,

Tourangeau, Wallander A, ed. *PNAS Proc Natl Acad Sci United States Am*. 2015;112(8):2395-2400. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2015-12112-004.

88. Irwin JR, McClelland GH, Schulze WD. Hypothetical and real consequences in experimental auctions for insurance against low-probability risks. *J Behav Decis Mak*. 1992;5(2):107-116. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1992-37865-001.

89. Matusiewicz AK, Carter AE, Landes RD, Yi R. Statistical equivalence and test-retest reliability of delay and probability discounting using real and hypothetical rewards. Audrain-McGovern Baumgartner, Beck, Bickel, Bickel, Bickel, Bickel, Davis, Gallistel, Green, Hinvest, Jikko, Jimura, Johnson, Johnson, Johnson, Kirby, Kirby, Krishnan-Sarin, Lagorio, Landes, Lane, Lawyer, Luzvar-Stiffler, Madden, Madden, Madden, Manwarin B, ed. *Behav Processes*. 2013;100:116-122. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc10&NEWS=N&AN=2013-42895-020.

90. Kesternich I, Heiss F, McFadden D, Winter J. Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D. *J Health Econ*. 2013;32(6):1313-1324. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23317633.

91. Coffman JM, Rhodes K V., Fix M, Bindman AB. Testing the Validity of Primary Care Physicians' Self-Reported Acceptance of New Patients by Insurance Status. *Health Serv Res*.

2016;51(4):1515-1532. doi:10.1111/1475-6773.12435.

92. Presseau J, Johnston M, Francis JJ, et al. Theory-based predictors of multiple clinician behaviors in the management of diabetes. *J Behav Med*. 2014;37(4):607-620. doi:10.1007/s10865-013-9513-x.

93. Kuczewski M. Is informed consent enough? Monetary incentives for research participation and the integrity of biomedicine. *Am J Bioeth*. 2001;1(2):49-51. doi:10.1162/152651601300169086.

94. Grady C. Money for research participation: does in jeopardize informed consent? *Am J Bioeth*. 2001;1(2):40-44. doi:10.1162/152651601300169031.

95. Gregory DM, Way CY. Qualitative Research in Clinical Epidemiology. In: ; 2015:301-313. doi:10.1007/978-1-4939-2428-8_18.

96. Cantor SB, Rajan T, Linder SK, Volk RJ. A framework for evaluating the cost-effectiveness of patient decision aids: A case study using colorectal cancer screening. *Prev Med (Baltim)*. 2015;77:168-173. doi:10.1016/j.ypmed.2015.05.003.

97. Stacey D, Légaré F, Col NF, et al. Decision aids for people facing health treatment or screening decisions. In: Stacey D, ed. *Cochrane Database of Systematic Reviews*. Vol 1. Chichester, UK: John Wiley & Sons, Ltd; 2014:CD001431. doi:10.1002/14651858.CD001431.pub4.

98. Dillman D, Smyth J, Christian L. *Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method*. Fourth Edi. Hoboken, NJ: Wiley; 2014.

99. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. In: Edwards PJ, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd; 2009. doi:10.1002/14651858.MR000008.pub4.

100. iPad Air 2. Apple Inc. http://www.apple.com/ca/ipad-air-2/. Published 2016.

Appendix 1: Search Strategy for PsycINFO

Terms for Decision Making or Behaviour:

1 decision making/ or choice behavior/ (69251)
2 Reasoning/ (13804)
3 Behavior/ (22387)
4 (decision* or choos* or choice* or behavio?r*).ti,ab. (960604)
5 Risk Taking/ (10184)
6 (tak* adj2 risk*).ti,ab. (9046)
7 or/1-6 (981737)

Terms for hypothetical situations:

8 Uncertainty/ (5915)
9 hypothetical*.ti,ab. (13298)
10 proxy.ti,ab. (5133)
11 (formal adj3 (reasoning or thinking or decision*)).ti,ab. (875)
12 or/8-11 (25077)

Terms for real-world situations:

13 Reality/ (3961)
14 (real or reality).ti,ab. (108980)
15 everyday.ti,ab. (32567)
16 or/13-15 (139440)
17 7 and 12 and 16 (896)

APPENDIX 2: SURVEY METHODS

SAMPLE FRAME

We proposed to conduct a web-based pilot survey informed by Dillman's Tailored Design Method.[98] Approval was obtained from the Ottawa Health Science Network Research Ethics Board.

In order to reduce coverage error, we selected a list of decision aid authors that was determined to be representative of the survey population. This list of 200 target decision aid developers was identified through Stacey et al.'s systematic review.[8] The Stacey et al. review is the definitive systematic review in the field of decision aid development and was determined to be the best obtainable list of decision aid developers.[97] Using this list allowed us to minimize coverage error. Coverage error occurs when the sample selected is not representative of the survey population. In order to reduce coverage error, the sample frame should be a high quality list that includes everyone in the group to which we intend to generalize our results.

The Stacey et al review lists studies about decision aid development that were included in the review, excluded from the review, awaiting assessment, and ongoing. We removed any duplicate names. The e-mail addresses and contact information of the authors was obtained by retrieving the study listed and consulting the author's contact information, or by cross referencing the bibliography listed on their institutional profile with the reference section of Stacey et al.'s review.

If we were to draw a probability sample from the sample frame, we would need a completed sample size of 65 based on the equation below.[37] This calculation assumes an acceptable sample error of 10%, a 95% confidence level, and maximum heterogeneity between respondents.

$$Ns = \frac{(Np)(p)(1-p)}{(Np-1)\left(\frac{B}{C}\right)^2 + (p)(1-p)}$$
$$= \frac{(200)(0.5)(0.5)}{(200-1)\left(\frac{0.1}{1.96}\right)^2 + (0.5)(0.5)}$$
$$= 65$$

"Where: Ns = completed sample size needed for desired level of precision
Np = size of population
P = proportion of population expected to choose one of the two response categories
B = acceptable amount of sampling error […]
C = Z statistic associated with the confidence level ; 1.96 corresponds to the 95% level"[37]

We determined that in order to avoid sampling error, we would simply contact all of the decision aid developers in our sample frame instead of using probability sampling to generate a random sample. This can be done because our survey will be done online and self-administered so there are little additional costs to expanding the number of contacts from 65 to 200. However, as described by Dillman, it is important to consider all sources of survey error and not neglect the potential for increased nonresponse error when choosing not to use probability sampling.[37]

CONTACT PROCESS

In addition to ensuring that the questionnaire is well-designed and respondent-friendly, Dillman presents four other elements of survey design that serve to increase response rates.[37] These are: making multiple contacts, providing return envelopes with first-class stamps, personalizing correspondence, and providing token prepaid financial incentives.[37]

When making multiple contacts with potential participants, it is important that every type of contact is different. Dillman explains that if the contact documents appear different from each other, they are more powerful than simply re-sending the same document several times.[37]All contacts with participants should be personalized. This includes using official stationary, addressing

the potential participants by name in the salutation, and including a real signature at the end of contact letters.[37] All documents that we intended to send to potential participants were personalized in these ways in order to minimize nonresponse. Dillman also suggests that providing participants with return envelopes with first-class stamps will minimize nonresponse because people are inherently averse to throwing out objects with real value.[37] However, this does not apply to our survey because all contacts were to be done through e-mail.

The process of contacting potential participants was to begin with a prenotification e-mail. This advises them to expect an incoming link to the survey and describes the general purpose of the project. This document serves to build positive salience for the questionnaire, meaning that it increases participants' anticipation for the survey. Two days after prenotification, the potential participants were to be sent a link to the survey, their unique access code, and a more detailed introduction letter. The introduction letter was to describe what the contacts were being asked to do, the purpose of the project, instructions on how to complete the survey, and the potential benefits and end goal of the project.[98] One week after the survey was sent we intended to send a thank-you reminder e-mail to anyone who had not yet responded to the survey. This e-mail would thank those who had completed the survey after the reminder was sent, and it would remind those who had not yet completed the survey to do so. Finally, two weeks after the first reminder was sent, a final plea was to be sent to all non-respondents in order to remind them to complete the survey and to thank them for their time. The final plea would explain that this was the contact's final opportunity to respond to the survey, and that they would not be contacted again.

INCENTIVES

Although Dillman cautions against using offers of prizes to improve response rates, more recent research shows that offering non-monetary rewards and immediately notifying participants about the results of the prize lottery can significantly increase the odds of response.[37,99] A Cochrane Review comparing methods to increase response to electronic surveys found that when non-monetary incentives, such as gift cards or lottery participation, were offered to respondents the odds of response were 1.72 (95% CI 1.09 to 2.72) compared to offering no incentive.[99] The same review also found that the odds of response to an e-questionnaire when respondents were immediately notified of the lottery results were 1.37 (95% CI 1.13 to 1.65) compared to when the notification of the lottery results was delayed. We therefore intended to offer entry into a lottery to win an iPad Air 2 as a way to reward participants for completing the survey and increase the response rate.[100] Respondents were to be immediately notified as to whether they had won the iPad as soon as they submitted their completed survey.

APPENDIX 3: DRAFT SURVEY

OTTAWA HEALTH RESEARCH INSTITUTE
INSTITUT DE RECHERCHE EN SANTÉ D'OTTAWA

Survey of Patient Decision Aid Developers – Factors Describing the Relationship between Hypothetical and Real-World Decisions

Thank you for agreeing to participate in our survey. You have been selected because you were identified as a patient decision aid developer in the Cochrane systematic review "*Decision aids for people facing health treatment or screening decisions".*[2]

We have conducted a review of the literature in order to identify factors known to affect the association between hypothetical and real-world decisions. We are conducting this survey to understand your experience with this issue as a patient decision aid developer.

While one of the goals in developing a decision aid is to ensure that it is clear, internally valid, etc., you may also have considered whether the patients assessing the decision aid in a hypothetical setting are likely to arrive at similar decisions as the patients who use the tool in the real-world. This is the focus of our survey; we hope to learn about how this issue manifests itself during patient decision aid development.

All responses will be kept in confidence and for research purposes only. This study has been approved by the Ottawa Health Science Network Research Ethics Board.

---

[2]Stacey, D., Légaré, F., Col, N. F., Bennett, C. L., Barry, M. J., Eden, K. B., … Wu, J. H. (2014). Decision aids for people facing health treatment or screening decisions. In D. Stacey (Ed.), *Cochrane Database of Systematic Reviews* (Vol. 1, p. CD001431). Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/14651858.CD001431.pub4

For this survey, please think of your most recent experience with patient decision aid development. Throughout this survey we will refer to this as 'your patient decision aid'

1. **Have you been involved in the development of a patient decision aid?**

☐ Yes Please indicate the name of the most recent patient decision aid here:
☐ No

2. **For how many years have you been involved in the development of patient decision aids?**

Number of years:
- □ **1-5**
- □ **6-10**
- □ **11-15**
- □ **16-20**
- □ **21+**

3. **How many patient decision aids have you been involved with developing?**

Number of patient decision aids:
- □ **1-5**
- □ **6-10**
- □ **11-15**
- □ **16-20**
- □ **21+**

In your experience with patient decision aids, did you employ any of the following:

(Check all that apply)

a. Patients with the target condition but who were not actually making the decisions
b. Non-patients who share demographic characteristics with the target group
c. Colleagues
d. Students
e. Convenience samples
f. Other (please specify)

4. In your experience with patient decision aids, did you employ any of the following: (Check all that apply)

a. Drafts or early version patient decision aids
b. Decision making settings other than where the patient decision aid will be used in clinical practice.
c. Other (please specify)

6. We have conducted a systematic review of the literature to find factors thought to affect whether hypothetical decisions will or will not predict real decisions.

Please indicate the extent to which you agree that the following factors might have been relevant in the context of ***your decision aid***.

Factor Statement	**A) Relevant in *the context of my decision aid.***					**Notes:**
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**	**Unclear/ Unsure**	
Hypothetical decisions correlate *more highly* with real-world decisions when…						
the decision has personal relevance to the participant.						
participants are older.						
participants are highly educated.						
participants are familiar with the options being evaluated.						
participants are highly engaged.						
participants are certain about their preferred course of action						
Participants experience some uncertainty/decision conflict when making their decision						
statements about uncertainty are provided. (E.g.by explicitly presenting the probabilities of different outcomes)						
Participants are asked about their beliefs about the probabilities of the different outcomes. (E.g. How likely do *you* believe the outcome is to occur?)						
the decision-making situation is made more emotionally arousing						
the amount of patient specific information is increased.						

Factor Statement	**A) Relevant in *the context of my decision aid.***					**Notes:**
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**	**Unclear/ Unsure**	
the content of the information mimics the real-world.						
the presentation format of the information mimics the real-world.						
the decision is framed as: “What action would *you* take in this situation?” as opposed to: “What action *should* be taken?						
the decision is framed in terms of direct, personal consequences. (As opposed to societal views and expectations related to the decision)						
incentives are offered.						

Hypothetical decisions are *less likely* to correlate with real-world decisions when…

Factor Statement	**A) Relevant in *the context of my decision aid.***					**Notes:**
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**	**Unclear/Unsure**	
participants have high cognitive ability.						
participants are not engaged in the decision						
the real-world decision elicits emotional reactions.						
there is potential for self-gain.						
large incentives are available						
the hypothetical choice is described with limited detail						
there is more time for reflection in the hypothetical setting than in real life						
participants try to answer in a way that is consistent with their self-identity						
others' impressions may be important (I.e. social desirability)						
there are potential risks associated with the decision						
answers could be affected by how a statement is framed (I.e. positive vs. negative frame)						
participants aren't experiencing/haven't experienced the condition						
when the participant is asked to consider the decision broadly. (As opposed to specific details of the decision)						
participants do not pay sufficient attention to the different attributes of a decision.						
What family and others will think about the decision is a factor						
participants have certain personality traits (E.g. extraversion, low neuroticism)						

7. Please indicate the extent to which you agree that the following factors might have been relevant in the context of ***other decision aids***

Factor Statement	**B) Relevant in *the context of other decision aids***					**Notes:**
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**	**Unclear/ Unsure**	
Hypothetical decisions correlate *more highly* with real-world decisions when…						
the decision has personal relevance to the participant.						
participants are older.						
participants are highly educated.						
participants are familiar with the options being evaluated.						
participants are highly engaged.						
participants are certain about their preferred course of action						
Participants experience some uncertainty/decision conflict when making their decision						
statements about uncertainty are provided. (E.g.by explicitly presenting the probabilities of different outcomes)						
Participants are asked about their beliefs about the probabilities of the different outcomes. (E.g. How likely do *you* believe the outcome is to occur?)						
the decision-making situation is made more emotionally arousing						
the amount of patient specific information is increased.						

Factor Statement	**B) Relevant in *the context of other decision aids***					**Notes:**
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**	**Unclear/ Unsure**	
the content of the information mimics the real-world.						
The presentation format of the information mimics the real-world.						
the decision is framed as: “What action would *you* take in this situation?” as opposed to: “What action *should* be taken?”						
the decision is framed in terms of direct, personal consequences. (As opposed to societal views and expectations related to the decision)						
incentives are offered.						

Hypothetical decisions are *less likely* to correlate with real-world decisions when…						
Factor Statement	**A) Relevant in *the context of other decision aids***					**Notes:**
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**	**Unclear/ Unsure**	
participants have high cognitive ability.						
participants are not engaged in the decision						
the real-world decision elicits emotional reactions.						
there is potential for self-gain.						
large incentives are available						
the hypothetical choice is described with limited detail						
there is more time for reflection in the hypothetical setting than in real life						
participants try to answer in a way that is consistent with their self-identity						
others' impressions may be important (I.e. social desirability)						
there are potential risks associated with the decision						
answers could be affected by how a statement is framed (E.g. positive vs. negative frame)						
participants aren't experiencing/haven't experienced the condition						
when the participant is asked to consider the decision broadly. (As opposed to specific details of the decision)						
participants do not pay sufficient attention to the different attributes of a decision.						
What family and others will think about the decision is a factor						
participants have certain personality traits (E.g. extraversion, low neuroticism)						

Please use this space to provide us with any comments you may have on our survey:

Please check here if you would like a copy of the results of this study e-mailed to you upon completion: ☐

Thank you for completing our survey!