LANCASTER UNIVERSITY

DOCTORAL THESIS

Essays in Decision Making Under Risk and Uncertainty

Author:
Jessica Alam

Supervisors:
Konstantinos Georgalos
Eyal Winter

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

January 2026

Declaration of Authorship

I, Jessica Alam, declare that this thesis titled, "Essays in Decision Making Under Risk and Uncertainty" and the work presented in it are my own. I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. I confirm that chapter 2 is jointly co-authored with Dr. Konstantinos Georgalos and Dr. Harrison Rolls has has been published in a journal.

LANCASTER UNIVERSITY

Abstract

Lancaster University Economics Department

Doctor of Philosophy

Essays in Decision Making Under Risk and Uncertainty

by Jessica Alam

This thesis consists of three chapters that focus on decision making under uncertainty. The first of these chapters focuses on decision making for others, the second on different risk attitudes of males and females, and the third on wishful thinking in strategic games.

In the first chapter, I conduct an experiment on decision making on behalf of others, and assess the difference between decision making for oneself and others under perfect and partial information. I use a lottery task to assess the risk preferences in each treatment group, and use the model outlined by Gurdal et al. (2013) to identify the impact of partial information on blame.

The second chapter considers the gender difference in risk attitudes between males and females with inconclusive results. We use use a model that allows for a degree of heterogeneity between individuals, where males and females risk attitudes are drawn from two separate distributions as opposed to one value for males and one for females. We take data from three experiments where gender effects where not found and, using this method we find gender effects. The second chapter was jointly co-authored with Konstantinos Georgalos and Harrison Rolls and has been published in the Journal of Economic Behavior and Organization, 202, 2022, 168-183.

The third and final chapter of this focuses on wishful thinking in strategic games. This adapts an existing belief model used by Nyarko and Schotter (2002), adjusted to account for the possibility of wishful thinking. We then apply this model to the original dataset from Nyarko and Schotter (2002) and a new dataset from Attanasi et al. (2018), which provides data from an alternative game. We find evidence to support the presence of wishful thinking in both games.

Acknowledgements

First and foremost, the biggest thank you possible to my supervisors, Konstantinos Georgalos and Eyal Winter for their support throughout my studies, and helping me grow as an economist. In particular, a huge thank you to Konstantinos for encouraging me never to give up, for his patience, insights and frequent contact despite my never ending challenges and questions. I truly would not have managed to complete this thesis without his encouragement and guidance.

I am also grateful to everyone involved in the Lancaster University Economics Department, the lecturers, support staff, and everyone who has been involved in my studies for their support and encouragement with my projects.

A huge thank you to my partner Joshua who has been patient, encouraging and supportive throughout the completion of this thesis. He has also provided solid proofreading services throughout, and accompanied me on many travel adventures when breaks were needed.

I am grateful to my friends Lokti, Holly and Kenna for making my time spent in Lancaster so happy and always supporting me.

I am also grateful to my family, in particular Mum and Richard, having driven the long drive to Lancaster too many times to count. Honorable mentions for Reuben, Peter and Andrew who have also visited and celebrated with me along the way.

A very special thank you to Dr John Higgs, Angela Higgs and Cleo Higgs for their never-ending encouragement and support.

A special thank you to my colleagues at Grant Thornton, who have also supported me throughout working full time and finalising my thesis; particularly Tom, who has helped me grow significantly as an economist, and Emily and Emma for providing a wonderful working environment to help me through!

Contents

List of Figures

List of Tables

Chapter 1

Decision making on behalf of others with asymmetric information

1.1 Introduction

Decision making on behalf of others is an area of literature that has recently developed significantly. For many people, they will make decisions for other during their life, from medical decisions to simple substitutions at the supermarket. The most notorious example of decision making on behalf of others is the principal agent problem; that is when decisions and ultimate benefactors are separated from one another and their incentives are not aligned. For example, a manager may have incentives to perform well enough to satisfy shareholders, but also has incentives to make their own life as easy as possible, whereas shareholders care only about the payoff from the manager's actions.

There are other examples where decisions are made on behalf of someone else, such as investors make decisions on behalf of advisors. Understanding the way in which individuals make decisions when they are not benefiting directly from the decision helps to understand what drives their behaviour, and can help to align payment to ensure the separation of ownership and control does not cause adverse incentives for the manager, advisor or decision maker.

Making decisions on behalf of oneself is different for making a decision on behalf of another. When making decisions on behalf of someone else, there are a number of considerations that one may make. For example, if whom you are making a decision for is risk averse, risk neutral, or risk seeking. Or if they would be disappointed with a lower payoff, or jealous of other individual's higher payoff.

On top of these considerations, one must also consider their own payoff. Experimental design is key with this consideration. For example, if the decision maker's payoff is fixed, then motivation to make a good choice for your partner may be limited. If your payoff is the same as your partner, your decision may mimic your own decision. If your payoff is chosen by your partner, then accountability changes the decisions you would make (Agranov et al., 2014; Andersson et al., 2020).

Whilst making decisions for others is itself complicated, the addition of asymmetric information complicates the decision even more. In the case of investment management, an advisor who makes decisions on behalf of investors may make decisions that the investor only receives partial information about. These scenarios are common in reality, and it is important to understand how the partial information impacts the decisions made.

Current literature has extensively studied the impact of various payment structures on decision making on behalf of others (Andersson et al., 2020)[1], and what determines the blame allocated to the decision maker (Gurdal et al. 2013). Most of the current literature investigates the payment structure in the context of another factor that impacts decisions, for example loss aversion (Füllbrunn and Luhan, 2017), social closeness (Montinari and Rancan, 2013) or risk taking on behalf of others (Eriksen and Kvaløy, 2010).

This paper extends the decision making on behalf of others literature by conducting

[1] Usually, this is decisions for oneself, others and oneself and others, as seen in Füllbrunn and Luhan (2017)

an experiment similar to existing experiments in the literature, but extended to include partial information in an effort to obtain more realistic behaviour of investors decision making process. In reality, in the context of the principal agent problem, the principal does not have all the information on the decision the agent makes. This introduces the main concern of moral hazard, whereby the agent has incentives to change behaviour so they can achieve a "good" outcome without any additional effort. Hence, we introduce some partial information to the experimental task. This partial information is related to the size of outcomes that can occur from both the lottery and sure thing. This means that whilst the investor knows some information about the decision, they do not know the full information. We additionally investigate the role of blame in the context of partial information.

To test the impact of partial information, we conduct an experiment whereby players make decisions on behalf of others under full and partial information, and find that individuals revert to their own risk preferences when making decisions on behalf of others, and that investors struggle to give blame or credit under partial information.

1.2 Literature Review

Existing literature has extensively investigated decision making on behalf of others, in various contexts.

Experimental designs have been used to investigate how decision making for others differs from individual decision making. As noted by Füllbrunn et al. (2020), Existing literature does not reach a consensus on risk taking on behalf of others. When conducting a meta-analysis of the decision making for others literature, Polman and Wu (2020) find that decisions made for others are slightly more risky than decisions made for oneself. However, results across individual studies vary significantly. Most of the existing literature hypotheses that individuals will take less or more risk on behalf of others, depending on the framing of the experimental task. For example, an experiment with accountability incorporated may is likely to lead to less risk taking on behalf of others, as participants may wish to avoid negative consequences.

However, in the context of lotteries with limited information feedback, participants may wish to take more risks to try and impress their partner.

Within existing literature, some papers found evidence of more risk taking on behalf of others (Chakravarty et al., 2011; Pahlke et al., 2012; Agranov et al., 2014; Pollmann et al., 2014; Andersson et al., 2020). However, other experiments find less risk taking on behalf of others (Montinari and Rancan, 2013; Fornasari et al., 2020). Some experiments even find no differences in decision making for oneself and others (Eriksen et al., 2020). The wide range of results are indicative of the different methodologies and tasks chosen by different authors. This is particularly notable, as most experiments within the literature are exploring a secondary topic alongside decision making on behalf of others, for example social distance (Montinari and Rancan, 2013), or responsibility (Charness and Jackson, 2009a).

One such differentiation is the choice of task. Experimental tasks vary significantly, with some authors choosing to use binary choice tasks (Polman, 2012; Andersson et al., 2020; Reynolds et al., 2009). Other task choices include stag hunt games (Charness and Jackson, 2009b), investment tasks similar to Gneezy and Potters (1997) (Eriksen and Kvaløy, 2010; Charness and Gneezy, 2010; Montinari and Rancan, 2013).

Holt and Laury (2002) price lists are also a common choice (Chakravarty et al., 2011; Humphrey and Renner, 2011; Andersson et al., 2016; Füllbrunn and Luhan, 2017), as these tasks make it easy to identify risk aversion. Other choices of task include Risky and save project investment (Agranov et al., 2014), trust game (Kvaløy and riaga, 2014), CE choice lists (Vieider et al., 2016), Ellsburg urn tasks (König-Kersting and Trautmann, 2016; Eckel and Grossman, 2002), gambles (Luzuriaga et al., 2017), choice tasks based on multiple price lists (Fornasari et al., 2020) , and choices between sure things and 50/50 gambles (Pahlke et al., 2012).

Even in extensions of the literature, such as decision making for groups, the tasks very. For example, Reynolds et al. (2009) uses theoretical tasks that do not directly incentivise participants. This provides little confidence for real-world applications, and the addition of hypothetical "others" further reduces the validity of this result. Each of these games has slightly different objectives, and these can lead to individuals making more or less risky decisions based on the context of the game. For example, whilst investment tasks are useful to more closely link to the principal agent problem, participants may be averse to the idea of investments and think about the implications of their participants being averse to them as well. On the other hand, a trust game may be more appealing to those making a decision for someone else, as some level of blame may be given to the participant who was trusted.

The type of task appears to have some no observable impact on the literature's findings. However, lottery choices appear to lean slightly to more risk taken on behalf of others, whereas investment tasks show more mixed results. This could occur due to the framing of the task, namely lotteries being perceived as random, whereas investment tasks may be seen more as having work involved. This could explain differences in the principal agent problem too; with some individuals feeling like outcomes are beyond control (e.g. due to economic conditions), and others crediting their effort for results. One example of this in the literature is Eriksen and Kvaløy (2010), who use an investment management task, whereby the investor is monitored frequently or infrequently, could directly lead to less risk taking on behalf of others, as participants feel less distanced from the decision, even when making it for another.

Whilst each of these tasks could be useful for different contexts, investment tasks or binary choice tasks are likely to be the most applicable outside of the laboratory, as they most closely mimic tasks that could occur in the investor/advisor or principal agent problem. For example, binary choices could be used to indicate two projects or investment strategies, and investment tasks are closer to an investment that an advisor could invest in on behalf of an investor.

On top of the choice of task, the role of incentives within a task is important, and should be carefully considered. Füllbrunn and Luhan (2017) find lower loss aversion on behalf of others with unaligned incentives, but no difference if payoffs are aligned. We note the importance of the payoff structure, and carefully consider the payoff structure for our experiment.

We note that by design, experiments with different tasks will have different payoff structures. However, even within this, the payoff structure between tasks can be vastly different. For example, Montinari and Rancan (2013) use an investment tasks with currency units, whereas Agranov et al. (2014) use an investment task more similar to a lottery. We also stress that the presence of incentives themselves is important to avoid capturing random choice is necessary.

The literature has begun to be extended in conjunction with other decision making concepts, such as loss aversion, with some literature finding less risk taking on behalf of others with losses (Eriksen and Kvaløy, 2010; Eriksen et al., 2020). However, finding less loss aversion on behalf of others is much more common (Polman, 2012; Pahlke et al., 2012; Andersson et al., 2016). Decision making for groups has also been investigated, with less risk taking on behalf of groups often being found (Charness and Jackson, 2009b; Reynolds et al., 2009). Other extensions incorporate self-other differences in ambiguity, but found no differences in ambiguity aversion (König-Kersting and Trautmann, 2016), and structural models of decision making on behalf of others have been created (Vieider et al., 2016).

Existing literature also addresses the relationship between participants. Montinari and Rancan (2013) find that the closer subjects are to the person they are making decision on behalf of, the closer the decision makers are to acting on behalf of themselves.

The table below summarises the differences in tasks and findings of the existing literature.

Study	**Focus**	**Task Type**	**Risk Preference Shift (Others vs. Self)**
Füllbrunn & Luhan (2020)	Delegation and responsibility in risky decisions	Lottery Choices	More risk taken for others
Chakravarty et al. (2011)	Risk preferences over others' money	Lottery Choices	More risk taken for others
Pahlke et al. (2012)	Accountability and risk in surrogate decisions	Lottery Choice	Less risk taken for others (under accountability)
Andersson et al. (2016)	Loss aversion in decisions for others	Lottery Choice	More risk taken for others
Vieider et al. (2016)	Cross-cultural risk preferences in other-regarding decisions	Lottery Choice	Ambiguous — effect direction depends on country; cultural norms influence whether more or less risk is taken for others
Eriksen et al. (2020)	Risk-taking on behalf of others	Investment Task	Less risk taken for others
Polman & Wu (2020)	Meta-analysis of decisions for others	-	More risk taken for others (overall), but effect size small and moderated by factors like framing and reciprocity
Pollmann et al. (2014)	Accountability's role in agent risk-taking	Investment Task	Ambiguous — more risk taken for others without accountability; similar or less risk with accountability mechanisms in place
Montinari & Rancan (2013)	Social distance and risk-taking	Lottery Investment Task	Ambiguous — less risk taken for distant others, more risk taken for close others
Agranov et al. (2014)	Portfolio managers' behavior with others' money	Investment Task	More risk taken for others

Study	Focus	Task Type	Risk Preference Shift (Others vs. Self)
Kvaløy & Luzuriaga (2014)	Trust game using others' money	Trust Game	Ambiguous — sending behavior remains constant, but trustworthiness (returns) drops with third-party funds
Eriksen & Kvaløy (2010)	Myopic investment behavior for others	Investment Task	Less risk taken for others
Humphrey & Renner (2011)	Responsibility in decision-making	Multiple Price List	Less risk taken for others
Polman (2012)	Loss aversion when deciding for others	Riskless Choice, Gambling Tasks, and Social Decisions	More risk taken for others

TABLE 1.1: Existing literature in decision making on behalf of others

Although the existing literature is vast and covers a wide range of other topics in conjunction with decision making for others, the difference in tasks, payment structures and inconsistency within the literature makes it difficult to draw insights from the literature as a whole. Each of these studies is culturally specific. This area is difficult to address without obtaining a number of participants from different cultures.

Existing literature does not currently address the presence of information asymmetry that is often present within the process of decision making on behalf of others. This limits the applicability of results to scenarios whereby all information is known to all parties. Whilst in some cases this may be applicable, for most decisions complete information is not available. For example in the principal agent problem, those that control a company are not able to directly view the efforts of a manager, but instead just the outcome of their decisions (e.g. sales or revenue). This difference in information available limits the applicability of results from decision making on behalf of others to the principle agent problem. Further, the results from tasks may not hold when this information asymmetry has not been accounted for.

This paper aims to address this gap, by conducting an experiment on decision making on behalf of others with partial information. As results often differ by task and experimental design, we choose to use similar tasks to Gurdal et al. (2013), as these tasks enable for simple estimation of the risk aversion parameter, without noise from more difficult to understands tasks. This also allows us to investigate the role of blame when there is partial information.

The task in this paper will be similar to those conducted by Gurdal et al. (2013) and Chakravarty et al. (2011), whereby lottery choices are presented to the participants. However, we will use the payback mechanism from Gurdal et al. (2013) to investigate blame, rather than simply performing the choices. Further, we will introduce partial information into this problem.

We find less risk taking on behalf of others with full information, but no difference in decision making for oneself and others under the Partial Information scenario. The payoff structure outlined in Gurdal et al. (2013) also allows for closer likeness to investment decision making than other task choices. This also allows us to investigate the role of partial information on decision making in investment decision making. In addition to these benefits, the task has the added bonus of allowing us to investigate the level of blame that the investors place on the advisors for their decision and outcomes. With the addition of partial information, we are also able to comment on the interaction between partial information and blame.

1.3 Experimental Design

The experiment was conducted online,[2] using participants from Lancaster University. The study included 84 participants from the University, and the average payment was £4.40, with the experiment usually taking between 20 and 30 minutes. Experiment instructions were displayed to participants, and participants were asked to read these carefully. After this, participants were asked if they have any questions, and encouraged to message the experimenter directly or put their hand up virtually if they require assistance.

The experimental task is similar to the task used by Gurdal et al. (2013), which investigates how decision-making under risk changes when individuals make decisions on behalf of others.

Gurdal et al. (2013) conduct many experiments, but we base our analysis on their main experiment, denoted the *allocate treatment*.

We modify the task, such that the differences between the expected value of the lottery and the sure thing are spread equally between positive and negative. This feature is useful to ensure variety of choice between the sure thing and the lottery, but also to obtain the participant's switching point, and easily determine their risk aversion parameter, in a similar way to Holt and Laury (2002b) task lists. The following table illustrates the tasks given.

[2]The experiment was coded in Python (https://www.python.org/ using Spyder, https://www.spyder-ide.org/), and was based on the oTree software by Chen et al. (2016). This was implemented using Heroku (https://www.heroku.com/).

P1	X1	P2	X2	EV	Y	EV-Y
0.25	80	0.75	0	20	35	-15
0.5	60	0.5	0	30	10	20
0.25	60	0.75	0	15	5	10
0.75	40	0.25	0	30	25	5
0.5	30	0.5	0	15	5	1
0.25	20	0.75	0	5	5	-30
0.25	80	0.75	0	20	45	-25
0.25	20	0.75	0	5	50	-45
0.5	100	0.5	0	50	10	40
0.25	60	0.75	0	15	20	-5

TABLE 1.2: Experimental tasks

Note: P1 illustrating the probability of obtaining X1, and similarly for P2. X1, X2 and Y are payoffs, EV is the Expected Value of the lottery, and EV-Y is the expected value minus the value of Y.

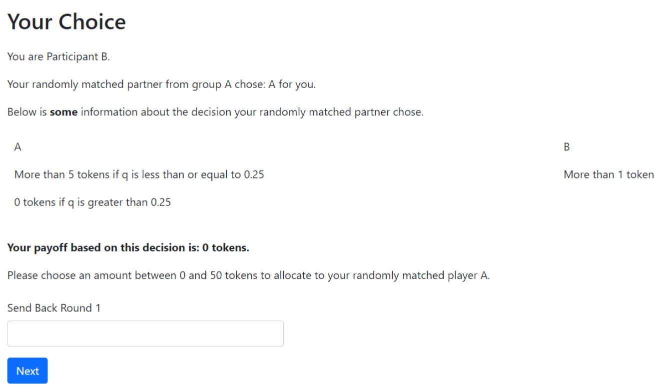
At the beginning of the experiment, participants were randomly split into two groups; Group A and Group B. Group A will be the advisor. Group B will be the investor.

There are three treatments in the experiment; Own treatment, Full Information treatment and Partial Information treatment.

The first treatment (Own treatment) is decision making for oneself. This is the baseline treatment, and forms a basis for comparison for decision making on behalf of others. In this treatment, each participant makes decisions between a lottery and the sure thing. Both investors and advisors make decisions for themselves in this treatment. Payment is according to the outcome of the decision, with the lottery choice being decided by a computer. This is used as a base for comparison to the literature, and existing risk attitudes and loss aversion parameters.

The second treatment (Full Information treatment) is decision making on behalf of another. In this treatment, advisors make the decision between lottery and sure thing for investors. Once this decision has been made, investors are shown the outcome of the decision and task the advisor faced. A computer generates a number between 0 to 1 to decide the outcome of the lottery. The investor then chooses to split a number of tokens between zero and 50 (equivalent of £0 and £5 respectively) between their advisor and a third party (this is another participant, and cannot be the advisor or investor in question). In this treatment, investors receive a payoff according to the outcome of the advisor's decision. Advisors receive their show up fee plus the amount allocated by the investor.

The third treatment is the Partial Information treatment. This works the same as the second treatment, but only partial information is shown to the investor. For this treatment, the advisors make a choice between a lottery and a sure thing on behalf of the investor. The advisor knows all information about the decision. After this decision is made, a computer generates a random number between 0 and 1; and the outcome of the lottery is shown to the investor. The investor also sees some information about the decision made, for example:


Your Choice

You are Participant B.

Your randomly matched partner from group A chose: A for you.

Below is **some** information about the decision your randomly matched partner chose.

A	B
More than 5 tokens if q is less than or equal to 0.25	More than 1 token
0 tokens if q is greater than 0.25	

Your payoff based on this decision is: 0 tokens.

Please choose an amount between 0 and 50 tokens to allocate to your randomly matched player A.

Send Back Round 1

Next

The investor then chooses how to allocate the amount of tokens (0 to 50, £0 to £5) between their advisor and a third party participant. The advisor is aware of the information that will be shared at the time of making their decision, and hence is aware there will be partial information shared. This means there is uncertainty for the agent (the lottery outcome), but ambiguity for the investor (as there is an ambiguous outcome of the lottery).

This task differs from the Gurdal et al. (2013) allocate experiment in two ways. Firstly, the allocation of the pot of tokens is not allocated between the experimenter and participants; as we wished to avoid any experimenter interaction. Instead, the remaining amount was allocated to another participant not in the pair in question. Secondly, the addition of the Partial Information treatment is novel.

As indicated by the screenshot of the task, in this treatment two of the potential outcomes are inequalities. We choose to use two inequalities instead of one, such that if the outcome of the decision reveals part of the information, there is still some unknown information. Again, advisors make the decision for investors, and advisors are paid according to their corresponding investors choice. Advisors will see the information received by investors before making their decision. In the treatments, the advisor always moves first. The investor (if applicable) will move second. There is no feedback between rounds for the advisor, but the investor sees the results from the advisor's choice.

Treatment three aims to see two things: Firstly, whether the lack of information leads to the advisor making more risky decisions to try and obtain the outcome, or whether it will lead to taking less risks because they know that the investor will not be able to fully judge their choice. Secondly, it will tell us whether or not the investor cares about the missing information, or attributes less money to the advisor due to the fact they cannot fully understand the choice of the advisor.

This allows us to investigate the factors that determine the amount the investor gives to the advisor, and whether or not the information impacts their attitude towards rewarding the investor.

We choose a within subject design, whereby each subject participates in all three treatments. This enables us to directly compare the treatments, without difficulties arising from between subject designs. Given the within subject design, we therefore run the Partial Information treatment first, to avoid any potential issues from participants remembering tasks. The other two treatments are randomised, to minimise any impacts of the order of treatments.

It is worth noting that we have many types of partial information that could have been implemented. Each of the following scenarios are examples of partial information, in a lottery with X and Y as outcomes, and Z as a sure thing. Px Represents the probability of X occurring, and $(1 - Px)$ represents the probability of Y:

1. x=1 $Px > 2/3$ and z=2

2. x=1 y=10, $py < 1/8$ and we do not know z

3. $x > 2$ y=10, px= 0.5 and $z > 3$

4. $min(x, y, z)$ and $max(x, y, z)$

5. $x < 2, y > 8$, $px = 0.8$, $py = 0.2$, $z = 8$

We choose to use 3 - whereby two of the payment outcomes are unknown. We choose to do this for two reasons:
Firstly, probabilities are more difficult to understand than outcomes, and hence if we have a scenario whereby probabilities are further modified, participants may find it difficult to interpret the probabilities.
Secondly, it makes it particularly difficult to use weighting functions when calculating risk aversion. It is also worth noting that we choose to have two of the potential outcomes as unknown such that regardless of the outcome, one of the potential outcomes will still be unknown. This means that we avoid the possibility of the lottery becoming full information, for example if the outcome was x, then z would still be unknown. If we chose a lottery such that x=2, y=10, px=0.5 and $z > 3$, if the advisor chooses the sure outcome of z, then the investor will see this, and this task becomes full information.

This option is implemented through wording, with "more than" or "less than" used to avoid confusion of participants.

1.4 Hypothesis

The aim of this section is to state the hypotheses that will be tested.

Hypothesis 1 *Investors will compensate Advisors less for decisions they know less about.*

Given that the only information the investors receive is related to the gamble, if this information is restricted, the investors will attribute successes less to the Advisor. We will test this hypothesis by comparing the amount allocated to the advisor in the Partial Information treatment and the Full Information treatment. We expect that there will be lower allocations to the advisor in the Partial Information treatment than the Full Information treatment.

Hypothesis 2 *Participants take less risks on behalf of others than they would for themselves*

In line with Charness and Jackson, 2009b; Eriksen and Kvaløy, 2010; Pahlke et al., 2012, we expect that participants will take less risks on behalf of others than they would for themselves. This is especially likely given the presence of "accountability" within the experimental design (Agranov et al., 2014), although we note that other papers disagree (Andersson et al., 2020). We expect that advisors will be more cautious with other people's money than their own. This will be investigated through the risk attitudes parameter for each treatment. This is partially due to the fact that advisors do not know the risk attitudes of the investor. We expect that advisors will be more cautious, as the "reckless" choice here of the lottery could lead to a low payment.

Hypothesis 3 *Advisors will take more risks when they know less information will be available to investors*

Considering advisors will also be aware of the information asymmetry, they may change their behaviour in light of this. We hypothesise that advisors will take more risks, to try and impress the investor. We expect to see this for two main reasons- firstly, advisors are encouraged to take more risks, when their decision would be less clear. Secondly, we expect that some advisors will note that they may be given less credit for decisions if the investor knows less. This will lead advisors to take more risks, to obtain a better outcome. We will evaluate this by comparing risk attitudes in the Full Information treatment and Own treatment.

Hypothesis 4 *Investors blame advisors more for decisions they know more about. We assume that more blame will be placed on advisors for decisions that investors know more about.*

Where more information is available, we expect that the investor will blame their advisor. With less available information, we hypothesise that some advisors will not place so much blame on their advisor, but instead blame the lack of information. This will be tested following the approach by Gurdal et al. (2013), outlined in the methodology section below.

1.5 Methodology

We choose to follow Gurdal et al. (2013), and estimate risk aversion parameters using CARA and CRRA utility respectively. The respective utility functions for each are included below.[3]

- **CARA (Constant Absolute Risk Aversion):** $U(x) = -\exp(-rx)$
- **CRRA (Constant Relative Risk Aversion):** $U(x) = \frac{x^{1-r}}{1-r}$ for $r \neq 1$

Where r is the risk aversion parameter to be estimated. In both of these cases, higher values of r indicate greater aversion to risk. Following this approach further, Gurdal et al. (2013) assume that individuals will choose the lottery or sure thing with higher utility, but with some randomness. This is accounted for using the following logit choice model:

$$P(\text{Choose A}) = \frac{\exp\left(\frac{1}{s}EU_A\right)}{\exp\left(\frac{1}{s}EU_A\right)+\exp\left(\frac{1}{s}EU_B\right)}$$

Where EU_A and EU_B are the expected utilities of option A and B respectively, and s is a noise parameter. A lower s means more consistent choices, and less random choices by the participant. This is estimated using a maximum likelihood estimation. This is conducted for each participant separately (individual estimation) and

[3] Estimations were conducted in R, using a general nonlinear augmented Lagrange multiplier optimisation routine that allows for random initialisation of the starting parameters as well as multiple restarts of the solver, to avoid local maxima. The estimation was conducted in the R programming language for statistical computing (The R manuals, version 3.6.4. Available at: http://www.r-project.org/)

for the full dataset together (global estimation). Standard errors are calculated using the curvature of the likelihood function. These estimates are used for a base of comparison against the findings in the literature.

This model has the benefit of estimating noise alongside the risk parameter, which allows us to capture a more accurate risk aversion parameter. The task also allows for varied risk and a switching point within the tasks, enabling individual and group level estimations. This is particularly beneficial for the within-subject design, so we can minimise the likelihood of choices being due to differences between participants. However, the use of CARA and CRRA can also result in unstable estimates, and can be very sensitive. There is also no within-sample validation possible, as the number of tasks is low and further reducing this could lead to inaccurate estimations. Participants may also have more complex preferences, rather than CARA or CRRA preferences. Notwithstanding these drawbacks, these estimates are used by a considerable amount of the academic literature related to decision making on behalf of others. Therefore, we choose to use this methodology and investigate the results carefully to ensure there are no estimation concerns.

Estimations will be compared across treatments, to investigate Hypothesis 1 to 4. We estimate parameters for each participant, as well as each treatment. This allows us to understand the differences in individual decision making, and observe an overall treatment effect for the group.

In addition to estimating risk aversion, the task design also allows us to investigate blame, as the investor sends back money to their randomly matched partner.

Our regression analysis follows that of Gurdal et al. (2013). This regression is a fixed effects model, which estimates the amount of blame which the investor blames on the advisor. This is specified as follows:

$$w_{i,t} = x_{i,t} + \gamma_i + \epsilon_{i,t} \tag{1.1}$$

where $x_{i,t}$ is a column vector consisting of 1 and the independent variables, γ_i representing the latent individual characteristics of subject i, and $\epsilon_{i,t}$ is the error term with the appropriate Guassian distribution assumptions. Unless stated otherwise, the latent effects γ are assumed to be random with mean zero, and uncorrelated with x and ϵ, that is, a random effects model.[4]

There are seven independent variables included: risky, safe, high, low, utility, E(utility premium) and period. The variable risky is a dummy that takes value of 1 if the risky alternative is chosen, and safe being 1 - risky (0). The variable high is a dummy variable equal to one if the value of q means the lottery yields a high payoff, while low = 1 - high (0). The variable E(utility premium) is a control equal to the difference between the expected utility of the choice made and the alternative not chosen. It is a measure of the attitude to the quality of the chosen alternative, and is intended to capture any between-period variation in a possible effect driven by the recognition of the advisor's effort by the investor. The control variable utility is an ex-post measure of the investor's wellbeing, their utility from the payoff, which might be impacted by relief, disappointment, distributional preferences or consequence-based reciprocity. The utilities are calculated using CRRA in this estimation, and estimated at the group level for all principals from the Own treatment.

Whilst this methodology is useful to understand the relationship between the payback mechanism and blame, the experimental design does not allow further investigation into why the payment was given. There are a number of alternative explanations that could explain sending back to another participant, such as reciprocity and fairness. However, controlling for the characteristics of the outcome does indicate whether participants are more likely to give more if the outcome is risky and high.

1.6 Results

This section outlines the results of the estimations outlined in the methodology section above.

[4]Since individuals do not have control over explanatory variables, it is reasonable to assume that any latent individual effects are independent of the explanatory variables. The Hausman test justifies use of the random effects statistic, with a p value of 0.92, failing to reject the null hypothesis of consistency of the random effects model.

1.6.1 CARA & CRRA Estimation

The results of the CARA and CRRA estimation are illustrated in the tables below.

Participant	r (SE)	s (SE)
1	0.111786* (0.010649)	9.839292 (7.184488)
2	0.105080 (0.393391)	4.499877 (5.191157)
3	0.170606** (0.002891)	4.499877 (5.191157)
4	0.028973 (0.406672)	4.499877 (5.405273)
5	0.993057 (0.987878)	0.712125 (3.3927000)
6	0.070474 (0.557235)	7.159694 (11.633587)
7	-0.027250 (0.406672)	4.499877 (5.405273)
8	0.118078 (0.523426)	9.793911 (16.054799)
9	0.994409** (0.038442)	1.203500 (2.3027292)
10	0.116634 (0.947492)	0.123158 (1.0283321)
11	-0.030214 (1.000219)	0.123158 (3.2937482)
12	0.320417 (0.407473)	8.203749 (6.767509)
13	-0.102360 (0.403624)	4.499878 (5.375735)
14	0.643454* (0.275622)	8.802339 (5.957596)
15	0.400120 (0.478554)	4.038112 (5.284045)
16	0.416964 (0.366788)	9.913739* (5.044185)
17	0.521142** (0.112812)	0.039448 (20.375275)
18	0.123582 (0.398423)	4.499877 (5.345201)
19	0.678946** (0.186472)	9.239039** (4.158021)
20	0.098370 (0.398423)	4.499877 (5.345201)
21	0.239288 (0.469426)	9.99273 (12.020077)
22	-0.021225 (0.394656)	4.499878 (5.379013)
23	0.461783* (0.230580)	1.993520 (1.769686)
24	0.209548 (0.710123)	8.393729 (20.657934)
25	0.092061 (0.546593)	9.394728 (12.080856)
26	0.836953 (0.716941)	8.070188 (19.709729)
27	0.118078 (0.470639)	9.207475 (14.105230)
28	0.100127 (0.492819)	0.107442 (0.192373)
29	0.026744 (0.018209)	0.107442* (0.018201)
30	-0.773860 (2.340784)	1.401666 (32.584662)
31	0.111786 (0.272410)	8.2902821 (5.279758)
32	0.947318** (0.135251)	7.103029 (22.367891)
33	0.078169 (0.465325)	9.9347011 (9.287954)
34	0.373746 (0.459124)	8.2394720 (9.796435)
35	0.520888 (0.596827)	0.039275 (21.023056)
36	0.145836 (0.573329)	9.892919 (17.023009)
37	0.429788 (0.255914)	1.731334 (1.714287)
38	-0.130094 (0.397910)	4.499878 (5.310675)
39	-0.041414 (0.383394)	4.499876 (5.179470)
40	0.118078 (0.463904)	8.465019 (14.201924)
41	0.461783* (0.227393)	1.993521 (1.770524)
42	0.813427** (0.257280)	9.999273 (9.661128)

TABLE 1.3: CARA Results for the Partial Information treatment.

Standard errors in parentheses. Estimates are rounded to 6sf. * $p < 0.05$, ** $p < 0.01$

Participant	r (SE)	s (SE)
1	0.000082 (0.000122)	0.001345 (0.000998)
2	0.001569* (0.000773)	0.006839** (0.000989)
3	0.001568* (0.000776)	0.006836** (0.000989)
4	0.001569 (0.000807)	0.006843** (0.000991)
5	0.989751** (0.001000)	0.005846** (0.001000)
6	0.003355** (0.000920)	0.027783** (0.000999)
7	0.001569* (0.000764)	0.006841** (0.000989)
8	0.005351** (0.000967)	0.068097** (0.001000)
9	0.958202** (0.001000)	0.002486** (0.001000)
10	0.021528** (0.001000)	0.001159 (0.001000)
11	0.009684** (0.001000)	0.001294 (0.000995)
12	0.000044 (0.000072)	0.001348 (0.000999)
13	0.001569 (0.000813)	0.006843** (0.000991)
14	0.000006 (0.001000)	0.001347** (0.000028)
15	0.002899** (0.000887)	0.034259** (0.000999)
16	0.000027 (0.000032)	0.001344 (0.001000)
17	0.059420** (0.001000)	0.001012 (0.000997)
18	0.001569 (0.000828)	0.006842** (0.000991)
19	0.000000 (0.001000)	9.247323** (0.001000)
20	0.001570* (0.000771)	0.006843** (0.000989)
21	0.000063 (0.000055)	0.001345 (0.000999)
22	0.001569 (0.000824)	0.006840** (0.000991)
23	0.032745** (0.001000)	0.124879** (0.001000)
24	0.000072 (0.000104)	0.001347 (0.000999)
25	0.000125 (0.000129)	0.001343** (0.000003)
26	0.025821** (0.001000)	0.548123** (0.001000)
27	0.005351** (0.000969)	0.068098** (0.001000)
28	0.009864** (0.000999)	0.001456 (0.001000)
29	0.034961** (0.001000)	0.002945** (0.001000)
30	0.000125 (0.000122)	0.001343 (0.000933)
31	0.000080** (0.000023)	0.001342** (0.000005)
32	1.000000** (0.001000)	0.503473** (0.001000)
33	0.000124 (0.000140)	0.001343 (0.000996)
34	0.000040 (0.000999)	0.001344** (0.000000)
35	0.059290** (0.001000)	0.001011 (0.000996)
36	0.000081 (0.000073)	0.001347 (0.000998)
37	0.020758** (0.000999)	0.088818** (0.001000)
38	0.001569* (0.000790)	0.006841** (0.000990)
39	0.001568* (0.000785)	0.006839** (0.000989)
40	0.005351** (0.000967)	0.068099** (0.001000)
41	0.032745** (0.000999)	0.124879** (0.001000)
42	0.216176** (0.001000)	0.809595** (0.001000)

TABLE 1.4: CARA Results for the Full Information treatment.

Standard errors in parentheses. Estimates are rounded to 6sf. * $p < 0.05$, ** $p < 0.01$

Participant	r	s
1	0.020349 (0.361933)	8.17290 (7.065039)
2	-0.000663 (0.397844)	4.409881 (5.335807)
3	0.009242 (0.512607)	9.912788 (14.191744)
4	0.093097** (0.240997)	5.169507 (16.415554)
5	0.052122** (0.184307)	0.038503 (21.070223)
6	0.025827 (0.508449)	7.9102840 (11.256011)
7	0.071880** (0.276233)	2.830974 (1.822063)
8	0.086011** (0.218872)	9.293749 (9.512892)
9	0.046191 (0.911944)	4.2917393 (18.535340)
10	0.046178* (0.230545)	1.953650 (1.771704)
11	0.001239 (1.000000)	0.107216 (1.000000)
12	0.069051** (0.287205)	3.247483 (6.499603)
13	0.001239 (1.000000)	0.107216 (1.000000)
14	-0.057793 (1.000000)	0.688769 (1.000000)
15	0.011807 (0.427304)	9.394673 (12.593930)
16	0.014614 (0.410648)	6.128774 (7.899869)
17	0.052174** (0.213900)	0.038572 (20.758555)
18	-0.050589 (0.232487)	1.372928 (8.675372)
19	0.011178 (0.417353)	8.379292 (10.697938)
20	0.093447** (0.124199)	6.293343 (15.195021)
21	0.011807 (0.487530)	9.803833 (14.738619)
22	0.023227 (0.439845)	9.8023936 (10.240023)
23	-0.000663 (0.398857)	4.409881 (5.328231)
24	0.046178* (0.228771)	1.953651 (1.766161)
25	-0.000663 (0.398857)	4.409881 (5.328231)
26	0.011807 (0.489950)	9.920380 (14.915192)
27	0.019429 (1.000000)	0.131230 (1.000000)
28	-0.000663 (0.398857)	4.409881 (5.328231)
29	0.046178* (0.226211)	1.953651 (1.764939)
30	0.005195 (1.000000)	0.135634 (1.000000)
31	0.033661 (0.358217)	9.203846 (5.666577)
32	0.086348** (0.154220)	2.088184 (2.472175)
33	0.023835 (0.363993)	9.236389 (6.499486)
34	0.096137 (0.967586)	0.362481 (9.721893)
35	0.088723 (0.923638)	0.834934 (7.9326823)
36	0.062933 (0.923729)	0.014934 (1.000000)
37	0.085235 (0.837280)	0.236947 (1.000000)
38	0.088799 (0.623932)	0.372398 (1.000000)
39	0.023227 (0.439845)	8.268484 (10.240023)
40	0.083862 (0.239037)	0.014934 (1.000000)
41	0.083767* (0.092732)	0.21474 (1.000000)
42	0.023799 (0.8374202)	0.237298 (1.000000)

TABLE 1.5: Own treatment CARA estimates.

Estimates are rounded to 6sf. Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$

Participant	r	s
1	0.724572 (0.591927)	7.871877 (9.959526)
2	0.767478** (0.311222)	7.534659 (7.466583)
3	-0.006634 (0.387609)	4.499878 (5.214977)
4	-0.006634 (0.387609)	4.499878 (5.214977)
5	0.740108** (0.240083)	4.101009 (2.865700)
6	0.281240 (0.436406)	3.962386 (5.158240)
7	0.461783* (0.230614)	1.993521 (1.771417)
8	0.145836 (0.350119)	2.913102 (9.228339)
9	-0.006634 (0.391352)	4.499877 (5.294654)
10	0.936413** (0.075842)	8.250709 (8.782192)
11	0.705504** (0.146361)	1.954257 (1.347649)
12	0.656243** (0.252354)	7.436809 (5.703700)
13	-0.006634 (0.397685)	4.499878 (5.271502)
14	0.451946 (0.357276)	8.813524* (4.920586)
15	0.659997 (0.559158)	9.241797 (8.465982)
16	0.060613 (0.478114)	8.019166 (13.534496)
17	0.705504** (0.146361)	1.954257 (1.347649)
18	-0.035274 (1.000000)	0.128199 (1.000000)
19	0.668934* (0.287404)	9.733016 (6.033767)
20	0.703827** (0.087565)	0.922605 (0.733003)
21	0.755903** (0.155179)	9.063878* (4.967577)
22	-0.080100 (1.141420)	9.742654 (27.012868)
23	0.376689 (0.639698)	5.060922 (13.815329)
24	-0.006634 (0.402088)	4.499878 (5.317588)
25	-0.505897 (0.237047)	6.459697 (9.772256)
26	0.465658 (0.356874)	9.488636 (6.737851)
27	-0.156488 (0.483853)	8.475767 (9.625139)
28	0.011505 (0.575385)	4.243242 (16.613770)
29	0.419354 (0.365128)	3.968538 (4.175150)
30	0.111786 (0.415235)	8.680621 (10.697263)
31	0.500384 (0.448534)	9.019455 (8.205108)
32	0.900219** (0.128568)	5.997660 (8.651228)
33	0.111786 (0.319600)	4.767444 (7.305262)
34	0.802960** (0.186317)	9.075847 (6.365364)
35	0.724572 (0.641972)	6.934677 (9.640385)
36	0.011505 (0.609887)	9.393749 (17.761758)
37	0.281240 (0.444832)	3.962388 (5.224576)
38	0.583627** (0.165274)	2.104973 (1.468532)
39	-0.006634 (0.400496)	4.499878 (5.330743)
40	0.045358 (1.000000)	0.119734 (1.000000)
41	0.930976** (0.237352)	5.274964 (16.247734)
42	0.580339* (0.276208)	8.629965 (5.208312)

TABLE 1.6: CRRA results for the Partial Information treatment.

Standard errors in parentheses. Estimates are rounded to 6sf. * $p < 0.05$, ** $p < 0.01$

Participant	r	s
1	0.111786 (0.319622)	6.629226 (7.174780)
2	0.042383 (0.397731)	4.499878 (5.279090)
3	-0.055073 (0.397731)	4.499878 (5.279090)
4	-0.167161 (0.397731)	4.499878 (5.279090)
5	0.908735 (1.000000)	0.045732 (1.000000)
6	0.027344 (0.557780)	7.159696 (11.637349)
7	0.119950 (0.397807)	4.499878 (5.311366)
8	0.118078 (0.427476)	7.788355 (12.587959)
9	0.999116 (1.000000)	2.459530** (1.000000)
10	0.023653 (1.000000)	0.110207 (1.000000)
11	0.044135 (1.000000)	0.110207 (1.000000)
12	0.320417 (0.408280)	6.250921 (6.788780)
13	-0.097592 (0.384661)	4.499878 (5.227712)
14	0.643454* (0.282522)	6.928730 (5.960965)
15	0.400120 (0.477095)	4.038112 (5.273309)
16	0.416964 (0.374490)	8.981564 (5.448676)
17	0.518029 (7.213404)	0.038781 (22.628014)
18	-0.126959 (0.384661)	4.499878 (5.227712)
19	0.678946** (0.171360)	7.347930** (3.941100)
20	-0.187029 (0.383359)	4.499879 (5.120265)
21	0.239288 (0.549091)	6.071457 (14.742976)
22	-0.228093 (0.383359)	4.499879 (5.120265)
23	0.461783 (0.237395)	1.993521 (1.779391)
24	0.209548 (0.623115)	5.607474 (17.797392)
25	-0.083923 (0.507497)	5.551279 (10.366210)
26	0.836953 (0.716103)	8.070192 (19.949619)
27	0.118078 (0.475492)	6.827709 (14.347216)
28	0.118078 (0.453517)	5.672442 (13.280561)
29	0.367931 (1.000000)	0.030757 (1.000000)
30	-0.793449 (1.613769)	1.517244 (26.419201)
31	0.111786 (0.397362)	5.867597 (10.160269)
32	0.947318** (0.135104)	2.073419 (22.357930)
33	-0.285551 (0.566224)	4.154411 (11.812586)
34	0.373746 (0.428959)	8.011069 (8.225326)
35	0.521857 (0.291142)	0.039325 (20.877663)
36	0.145836 (0.361080)	9.669661 (9.463811)
37	0.429788 (0.243294)	1.731334 (1.674534)
38	-0.172109 (0.404115)	4.499877 (5.390767)
39	0.057933 (0.404115)	4.499877 (5.390767)
40	0.367931 (1.000000)	0.030757 (1.000000)
41	0.461783 (0.238514)	1.993520 (1.778494)
42	0.813427** (0.278688)	8.095236 (10.521985)

TABLE 1.7: CRRA results for the Full Information treatment.

Standard errors in parentheses. Estimates are rounded to 6sf. * $p < 0.05$, ** $p < 0.01$

Participant	r	s
1	0.203492 (0.361933)	6.182637 (7.065039)
2	-0.006634 (0.397844)	4.499878 (5.335807)
3	0.092422 (0.512607)	9.999900 (14.191744)
4	0.930976** (0.240997)	5.275007 (16.415554)
5	0.521225** (0.184307)	0.039288 (21.070223)
6	0.258279 (0.508449)	9.092700 (11.256011)
7	0.718802** (0.276233)	2.888749 (1.822063)
8	0.860112** (0.218872)	8.104927 (9.512892)
9	0.461914 (0.911944)	9.982946 (18.535340)
10	0.461783* (0.230545)	1.993521 (1.771704)
11	0.012397 (1.000000)	0.109404 (1.000000)
12	0.690512** (0.287205)	7.000026 (6.499603)
13	0.012397 (1.000000)	0.109404 (1.000000)
14	-0.577935 (1.000000)	0.702825 (1.000000)
15	0.118078 (0.427304)	8.373832 (12.593930)
16	0.146142 (0.410648)	9.003820 (7.899869)
17	0.521747** (0.213900)	0.039359 (20.758555)
18	-0.505897 (0.232487)	7.0230870 (8.675372)
19	0.111786 (0.417353)	8.0200200 (10.697938)
20	0.934475** (0.124199)	8.0219311 (15.195021)
21	0.118078 (0.487530)	9.090300 (14.738619)
22	0.232272 (0.439845)	8.0392370 (10.240023)
23	-0.006634 (0.398857)	4.499878 (5.328231)
24	0.461783* (0.228771)	1.993521 (1.766161)
25	-0.006634 (0.398857)	4.499878 (5.328231)
26	0.118078 (0.489950)	3.02010 (14.915192)
27	0.194292 (1.000000)	0.133908 (1.000000)
28	-0.006634 (0.398857)	4.499878 (5.328231)
29	0.461783* (0.226211)	1.993521 (1.764939)
30	0.051951 (1.000000)	0.138402 (1.000000)
31	0.336613 (0.358217)	6.000201 (5.666577)
32	0.863480** (0.154220)	2.130800 (2.472175)
33	0.238350 (0.363993)	7.187202 (6.499486)
34	0.961374 (9.967586)	0.369879 (9.721893)
35	0.187945 (1.001200)	0.115239 (1.037947)
36	0.897995 (1.012860)	0.012839 (1.023730)
37	0.987935 (1.023930)	0.015842 (1.247308)
38	0.887995 (1.005730)	0.425239 (1.0374933)
39	0.961374 (9.967586)	0.369879 (1.721893)
40	0.837795 (1.000930)	0.015239 (1.003783)
41	0.300703 (0.378670)	8.0321101 (7.400718)
42	0.556445 (0.346608)	6.130304 (5.841133)

TABLE 1.8: CRRA results for Own treatment. Standard errors in parentheses.

Estimates are rounded to 6sf. * $p < 0.05$, ** $p < 0.01$

In the Full Information treatment, there is a significantly lower level of risk aversion than in the Own treatment. This indicates that individuals take less risk on behalf of others than for themselves. This is similar to existing literature, such as Montinari and Rancan (2013) and Fornasari et al. (2020). This finding does not hold for the Partial Information treatment, where there is no statistical difference between

	Partial Information treatment	Full Information treatment	Own treatment
CARA	0.287309	0.08421653	0.0343060
CRRA	0.3723125	0.2390383	0.3899336

TABLE 1.9: Mean CARA and CRRA results

the Partial Information treatment and the Own treatment. Further, the group level estimate for the Partial Information treatment is statistically different from the group level estimate for the Full Information treatment. This indicates less risk aversion when making decisions are made on behalf of others with Full Information, but when partial information is introduced, participants revert to their own risk aversion preferences.

Summary statistics for the individual estimations are in Table 1.9 above.

We note that for the Own treatment, CARA is lowest but CRRA is highest. This could occur because of noise in the CARA estimates,[5] or extreme individual values of the parameters. Hence, we do not rely on the mean results for inference.

Further, the individual level results having considerably different mean results than the treatment level estimations, indicating that there is a large variation in between individual risk aversion parameters.

Participants are on average less risk averse in the Full Information treatment than the Own treatment for the CARA estimates, but there is no statistical difference between CRRA estimates.[6] This difference is likely attributable to the level of noise in the CRRA estimates.[7] Less risk taking on behalf of others is consistent with existing literature, including (Montinari and Rancan, 2013) and (Fornasari et al., 2020), and consistent with Hypothesis 2.

Similarly to the group level, the CARA results and CRRA results indicate more risk

[5]We note that CARA estimates are significantly noisier than CRRA, as indicated by the value of s

[6]Paired Wilcoxon rank of 0.028 and 0.589 respectively.

[7]Average values of the noise parameter are 0.2022 and 5.80 for CARA and CRRA estimates in the Full Information treatment respectively, and 5.80 and 5.77 in the Own treatment.

aversion on behalf of others when there is partial information. The paired Wilcox rank sum test for both CARA and CRRA illustrates significance at the 5% level for CARA and 5% level for CRRA.[8] The average difference between each participant's individual risk aversion parameters is 0.2031 for CARA estimations, and 0.133 for CRRA.[9] This means when participants know their partner will see only partial information, they are more likely to take risks with other participants' money. Participants in Group A are be taking more risks to impress their counterpart in Group B, as they understand that the full extent of the choice is not evident to their partner. This consistent with Hypothesis 3. However, there is no evidence to suggest that participants become more risky than they would if they were making the decision for themselves. This could suggest participants are more comfortable making decisions when there is less information available to the advisor and being more willing to take risks compared to when all information is known.

Whilst there is a small difference in CARA estimations between the Partial Information treatment and the Full Information treatment, it is not statistically significant.[10] This indicates that when participants face partial information, they return to making decisions for others in the same way they make decisions for themselves.

1.6.2 Regression Analysis

The average number of tokens sent back to the advisor was 17.20 in the Full Information treatment, which is 34.4% of available tokens. This is approximately 3.4 (6.8%) more currency units on average than the Partial Information treatment. This suggests that on average, participants are compensating less for decisions they know less about, and when we conduct the regression in table 1.14, this impact is significant.[11] This finding is in line with Hypothesis 1.

The results of the regression estimation are outlined in the tables below:

[8] Willcox rank sum test p value of 0.028 and 0.034 for CARA and CRRA respectively.

[9] The mean values of our noise parameter in the sample are 0.756 and 0.2022 for CARA estimates in treatment 1 and 2 respectively, and 7.08 and 5.81 for CRRA estimates in the Partial Information treatment and 2 respectively. This indicates that CRRA estimates are noisier than CARA estimates.

[10] Wilcoxon rank sum test p values 0.678 and 0.789 for CARA and CRRA respectively

[11] It is noted that this value is lower than the 3.4 observed difference, indicating that those in the Full Information treatment allocate approximately 2.58 more units.

TABLE 1.10: Random Effects Regression Results the Partial Information treatment

	Dependent variable: Estimated Effects on the Investor's Payment to the Advisor				
	(1)	(2)	(3)	(4)	(5)
Constant	20.627***	21.870***	19.128***	15.675***	17.195***
	(1.562)	(1.929)	(1.589)	(2.431)	(2.717)
Safe x Low	7.494***	7.728***	6.921**	6.141**	5.873**
	(2.809)	(2.817)	(2.770)	(2.793)	(2.799)
Risky x Low	−1.127	−1.232	−1.016	−0.731	−1.318
	(2.209)	(2.211)	(2.175)	(2.174)	(2.223)
Risky x High	1.040	0.829	0.314	0.466	0.408
	(2.449)	(2.455)	(2.418)	(2.412)	(2.411)
Utility		−0.322		0.723*	0.798**
		(0.293)		(0.386)	(0.390)
Period					−0.398
					(0.318)
E[Utility Premium]			0.682***	0.984***	1.053***
			(0.181)	(0.242)	(0.248)
Observations	420	420	420	420	420
R^2	0.023	0.026	0.056	0.064	0.067
Adjusted R^2	0.016	0.017	0.047	0.052	0.054
Residual Std. Error	18.084	18.080	17.805	17.751	17.739
	(df = 416)	(df = 415)	(df = 415)	(df = 414)	(df = 413)
F Statistic	3.332**	2.802**	6.122***	5.628***	4.957***
	(df = 3; 416)	(df = 4; 415)	(df = 4; 415)	(df = 5; 414)	(df = 6; 413)

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The variable risky is a dummy that takes value of 1 if the risky alternative is chosen, and safe being 1− risky (0). The variable high is a dummy variable equal to one if the value of q means the lottery yields a high payoff, while low $= 1-$ high (0). The variable E(utility premium) is a control equal to the difference between the expected utility of the choice made and the alternative not chosen.

TABLE 1.11: Random Effects Regression Results for the Full Information treatment

	Dependent variable: Estimated Effects on the Investor's Payment to the Advisor				
	(1)	(2)	(3)	(4)	(5)
Constant	23.976***	21.110***	24.193***	19.034***	21.573***
	(1.619)	(1.937)	(1.805)	(2.504)	(2.770)
Safe x Low	11.204***	11.328***	11.342***	10.601***	9.912***
	(2.694)	(2.676)	(2.744)	(2.731)	(2.739)
Risky x Low	−3.937*	−4.419*	−3.933*	−4.585**	−3.569**
	(2.267)	(2.258)	(2.269)	(2.259)	(1.901)
Risky x High	3.656***	5.339**	3.815	4.968**	4.393*
	(0.443)	(2.507)	(2.514)	(2.521)	(2.526)
Utility		0.610***		0.795***	0.837***
		(0.230)		(0.270)	(0.270)
Period					−0.701**
					(0.333)
E[Utility Premium]			0.037	−0.207	0.329*
			(0.136)	(0.159)	(0.168)
Observations	420	420	420	420	420
R^2	0.090	0.056	0.050	0.060	0.070
Adjusted R^2	0.093	0.047	0.031	0.048	0.056
Residual Std. Error	18.023	17.894	18.043	17.879	17.805
	(df = 416)	(df = 415)	(df = 415)	(df = 414)	(df = 413)
F Statistic	5.774***	6.149***	4.339***	5.269***	5.166***
	(df = 3; 416)	(df = 4; 415)	(df = 4; 415)	(df = 5; 414)	(df = 6; 413)

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The variable risky is a dummy that takes value of 1 if the risky alternative is chosen, and safe being 1− risky (0). The variable high is a dummy variable equal to one if the value of q means the lottery yields a high payoff, while low = 1− high (0). The variable $E(\text{utility premium})$ is a control equal to the difference between the expected utility of the choice made and the alternative not chosen.

TABLE 1.12: Random Effects Regression Results for the Partial Information treatment (Third Party)

	Dependent variable: Investor's payment to the third party				
	(1)	(2)	(3)	(4)	(5)
Constant	29.373***	28.130***	30.872***	34.325***	32.805***
	(1.562)	(1.929)	(1.589)	(2.431)	(2.717)
Safe × Low	−7.494***	−7.728***	−6.921**	−6.141**	−5.873**
	(2.809)	(2.817)	(2.770)	(2.793)	(2.799)
Risky × Low	1.127	1.232	1.016	0.731	1.318
	(2.209)	(2.211)	(2.175)	(2.174)	(2.223)
Risky × High	−1.040	−0.829	−0.314	−0.466	−0.408
	(2.449)	(2.455)	(2.418)	(2.412)	(2.411)
Utility		0.322		− 0.723*	−0.798**
		(0.293)		(0.386)	(0.390)
Period					0.398
					(0.318)
E[Utility Premium]			−0.682***	−0.984***	−1.053***
			(0.181)	(0.242)	(0.248)
Observations	420	420	420	420	420
R^2	0.023	0.026	0.056	0.064	0.067
Adj. R^2	0.016	0.017	0.047	0.052	0.054
Residual SE	18.084	18.080	17.805	17.751	17.739
(df)	(416)	(415)	(415)	(414)	(413)
F Stat.	3.33**	2.80**	6.12***	5.63***	4.96***
(df)	(3; 416)	(4; 415)	(4; 415)	(5; 414)	(6; 413)

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. *The variable risky is a dummy that takes value of 1 if the risky alternative is chosen, and safe being* 1− *risky (0). The variable high is a dummy variable equal to one if the value of q means the lottery yields a high payoff, while low* = 1− *high (0). The variable* $E(utility\ premium)$ *is a control equal to the difference between the expected utility of the choice made and the alternative not chosen.*

TABLE 1.13: Random Effects Regression Results for the Full Information treatment (Third Party)

	Dependent Variable: Investor's payment to the third party				
	(1)	(2)	(3)	(4)	(5)
Constant	26.024***	28.890***	25.807***	30.966***	28.427***
	(1.619)	(1.937)	(1.805)	(2.504)	(2.770)
Safe × Low	−11.204***	−11.328***	−11.342***	−10.601***	−9.912***
	(2.694)	(2.676)	(2.744)	(2.731)	(2.739)
Risky × Low	−3.937*	−4.419*	−3.933*	−4.585**	−3.569
	(2.267)	(2.258)	(2.269)	(2.259)	(2.301)
Risky × High	3.656	5.339**	3.815	4.968**	4.393*
	(2.443)	(2.507)	(2.514)	(2.521)	(2.526)
Utility		−0.610***		−0.795***	−0.837***
		(0.230)		(0.270)	(0.270)
Period					0.701**
					(0.333)
E[Utility Premium]			−0.037	−0.207	−0.329*
			(0.136)	(0.159)	(0.168)
Observations	420	420	420	420	420
R^2	0.040	0.056	0.040	0.060	0.070
Adj. R^2	0.033	0.047	0.031	0.048	0.056
Residual SE	18.023	17.894	18.043	17.879	17.805
(df)	(416)	(415)	(415)	(414)	(413)
F Stat.	5.77***	6.15***	4.34***	5.27***	5.17***
(df)	(3; 416)	(4; 415)	(4; 415)	(5; 414)	(6; 413)

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. *The variable risky is a dummy that takes value of 1 if the risky alternative is chosen, and safe being 1− risky (0). The variable high is a dummy variable equal to one if the value of q means the lottery yields a high payoff, while low = 1− high (0). The variable E(utility premium) is a control equal to the difference between the expected utility of the choice made and the alternative not chosen.*

TABLE 1.14: Random Effects Regression Results for all treatments.

	Dependent Variable: Investor's payment to the Advisor
	(1)
Constant	20.264***
	(2.011)
Safe Low	−7.972***
	(1.946)
Risky Low	−2.511
	(1.680)
Risky High	−3.283·
	(1.734)
EU Difference	−0.435***
	(0.104)
Utility	0.631***
	(0.179)
Time	−0.544**
	(0.227)
Indicator	−2.580***
	(0.333)
Observations	840
R^2	0.053
Adj. R^2	0.045
Residual SE	17.86
(df)	(832)
F Stat.	6.621***
(df)	(7; 832)

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Indicator is equal to 1 if the observation is in the Full Information treatment, and 0 for the Partial Information treatment.The variable risky is a dummy that takes value of 1 if the risky alternative is chosen, and safe being 1− risky (0). The variable high is a dummy variable equal to one if the value of q means the lottery yields a high payoff, while low = 1− high (0). The variable $E(utility\ premium)$ is a control equal to the difference between the expected utility of the choice made and the alternative not chosen.*

The treatment level estimates are included in the table below.

The Partial Information treatment regression results are indicated in Table 1.10. we focus on specification 5, which includes all of the control variables.

Model	Treatment	r	s	SE(r)	SE(s)		
CARA	Partial Information	0.0157***	0.2698***	0.00094	0.00109	**	**
CRRA	Partial Information	0.3066***	7.7837***	0.00100	0.00201	**	**
CARA	Full Information	0.0075***	0.0976***	0.00067	0.0167	**	**
CRRA	Full Information	0.1508***	9.7900***	0.00100	0.00100	**	**
CARA	Own	0.0194***	0.2286***	0.00094	0.00198	**	**
CRRA	Own	0.3275***	7.4074***	0.00100	0.00205	**	**

TABLE 1.15: Group level estimations for CARA, CRRA for all treatments. Note that all estimates are rounded to 3SF. * $p < 0.05$, ** $p < 0.01$

The coefficient on Safe x Low is 5.87, indicating that participants reward their matched partners with approximately 5.87 (11.74%) more currency units (5.9 pence) if the investor chose the sure thing and the outcome of the lottery was low. This is approximately 60% of the estimation for the Full Information treatment, indicating that participants are crediting the advisor less when they know less about the decision.

The coefficients for Risky x Low and Risky x High are not statistically significant. This indicates that when the advisor chooses the risky option, the investor does not credit the decision for the outcome. This suggests that under uncertainty, investors do not credit advisors for their choices when they choose the risky option. This could also occur due to the uncertainty of choice between the risky and safe outcomes.

Aside from the choice characteristics, investors in the Partial Information treatment credit advisors a relatively small amount for their outcome; approximately 0.798 currency units per additional unit of utility, and 1.05 currency units for each unit of utility over the expected amount. The period is not significant in the Partial Information treatment.

The results in Table 1.10 indicate that when there is partial information, participants struggle to credit or blame their advisor. This is in contrast to Hypothesis 4. It seems that even when advisors do credit their advisor in the Safe x Low scenario, they do so at a lower rate than when they know the full information of the choice. The existence of blame is less prevalent under partial information, and could indicate investors blaming uncertainty for poor outcomes, rather than their advisor. This suggests there is a level of confidence required for investors to place blame on advisors.

The Full Information treatment's regression results are illustrated in Table 1.11. The regression results indicate that participants on average send back approximately 21.6 currency units. This represents approximately 2/5 of the available amount of currency, and approximately 2.16 of currency when converted to value.

The coefficient on Safe x Low is statistically significant at the 1% level, and illustrates that participants reward their matched partner when they take the sure thing and the outcome of the lottery is low with approximately 9.91 currency units (0.99). Similarly, the coefficient on Risky x High is approximately 4.39 (0.44), and statistically significant at the 5% level. This indicates that participants reward their partner with approximately 0.44 when they take the risky option (in this case, the lottery), and it pays off. In all specifications, the magnitude on Risky x High is lower than Safe x Low. This could indicate rewarding the agent for avoiding a disappointing outcome under the low lottery outcome, but attributing the High outcome somewhat to luck when the risky option is taken.

The coefficient on Risky x Low is negative and statistically significant at the 5% level. This indicates that participants may blame participants for a low outcome if they took the risky option, and reduce the amount they send back to their matched participant by approximately 3.57 currency units (0.36).

The source of variation beyond the main effects of choices and outcomes appears to be a combination of utility and expected utility premium. For the each unit of utility gained from the outcome, participants compensate an additional 8.4 pence on average. The coefficient on E[Utility Premium] indicates that participants pay 3 pence for the premium in utility over expected. Since the coefficient on Utility is larger than the coefficient on E[Utility Premium], participants appear to be driven more by their outcome directly than the difference from expected outcome.

The coefficient on period is small, but indicates that for each additional period, participants reduced their average payment sent to the investor by 7 pence.

The Full Information treatment illustrates results similar in direction to those reported by Gurdal et al. (2013), but the results differ in magnitude. This is expected as the value of currency to units is different.

1.7 Conclusion

Decisions made for oneself vary from those made for another. Under full information, individuals make different decisions than for oneself than they do for another and revert to making decisions in the same way in which they do so for themselves; with more aversion to risk. The higher risk aversion is likely to ensure that their matched partner does not blame them for a lower payoff if the risky option gives a low payoff.

We illustrate the existence of blame in the Full Information treatment, but a lack of blame in the Partial Information treatment. This indicates that investors struggle to blame their advisor when they know less about the decision. This is even seen in the average payment to advisors being lower in the Partial Information treatment than the Full Information treatment. This indicates that the third party acts as a buffer, and obtains any share of the currency units only when the investor cannot place credit on their advisor (similar to Gurdal et al. (2013)). This means that with partial information, investors struggle to disentangle between their advisor's decision and luck. Advisors could therefore take advantage of this, and take less care over the decision to blame bad luck rather than their bad decision making skills.

The presence of the payback mechanism appears to not impact the decisions that individuals make when there is partial information, as advisors default back to decisions that they would have made for themselves. This further suggests that advisors may be trying to take advantage of the principal agent problem under full information, but under partial information may not understand how to take advantage of the information. This could be further investigated in future literature by changing the mechanism of partial information, for example by making the probabilities of outcomes unknown, a mixture of probabilities and outcomes or making the risky and sure things unclear to the investor.

The components of blame appear to be consistent with Gurdal et al. (2013), with participants blaming their advisor when there is a low result of the risky option, and crediting advisors for a high result of a risky option or choosing the sure thing when the risky option returns the low result. However, these results are not necessarily causal, and should not be interpreted as such. The experimental design does not investigate the intentions behind the payback mechanism, it is possible that amounts paid back could be driven by social preferences, for example reciprocity and fairness. Whilst the mechanism of allowing allocation to the third party is designed to minimise this, it cannot be ruled out.

Further research could attempt to identify causal impacts of blame and performance by performing a "shock" within the experiment, for example with unexpected income, and performing a difference in difference estimation. Additionally, investigating further whether this payment mechanism is blame or another preference directly, such as fairness or reciprocity, could help to understand the full decision making process of the investor.

Ultimately, the presence of partial information and payback based on outcome increases the risk of an advisor or agent using partial information to their advantage; illustrating findings of moral hazard. This is not surprising, given that partial information brings the game closer to the real life principal agent problem, and is likely to hold in scenarios where there is a separation of ownership and control.

Whilst these results provide some insight into the likely impacts of partial information on blame and decision making on behalf of others and evidence of moral hazard, the form of partial information could have significant impacts on behaviour of advisors. For example, if the final payoff was not known, this is likely to cause a different response from investors. Similarly, if probabilities are unknown this could cause a different reaction from investors. We expect that there would be further evidence of moral hazard, but the extent of this depends on how valuable the investor views the outcome and probabilities to understanding the decision. To further understand the role of partial information, future research could consider changing the type of partial information.

Finally, these results are likely to be culturally specific, given that other cultures have very different approaches to taking risk. Ensuring a diverse participant panel could help to make future studies more easily applicable to other contexts. Additionally, the presence of partial information is limited to one type. Whilst this is a step towards introducing partial information, the reality is that in the principal agent problem, the decision makers themselves face potential uncertainty. For example, a CEO may not be able to accurately define a gamble or the outcomes associated with choices, especially when it comes to external factors like economic growth and household spending. Future research could look to changing the task to be more realistic in this way, and generate a more accurate principal agent scenario.

Additionally, the lottery task itself could encourage riskier behaviour. Further investigation and causal analysis on the interaction between task types and decision making on others could be beneficial to understanding the wider implications of the literature as a whole.

Chapter 2

Risk Preferences, Gender Effects and Bayesian Econometrics

2.1 Introduction

There is no doubt that risk preferences play a central role in every aspect of economic life. Gender differences in risk preferences is a much debated topic and it has often been argued that these differences might provide a possible explanation of the observed differences between the two genders in various aspects of economic life such as financial decision making, hold of front office roles, or entrepreneurship, to name but a few. Nevertheless, there is little agreement on whether there is a universal pattern of differences between the two genders. Early surveys from the economic literature (Eckel and Grossman, 2008; Croson and Gneezy, 2009) provide mostly supporting evidence of women being less willing to accept risks. Recently, Filippin and Crosetto (2016) conducted an extensive meta-analysis on gender differences and risk attitudes, using data from 7000 subjects and 54 replication studies of the Holt and Laury (2002a) risk elicitation task. One of their main findings is that:

> "[..] gender differences appear in less than 10% of the studies and are significant but negligible in magnitude once all the data are pooled."

and they conclude that:

> "[..] the structural model seems to confirm that significant gender differences are detected in the HL task when merging all the observations. The reason is to be found in the sky-rocketing increase of the statistical power of the test, which drives fairly close to zero the likelihood of observing a false negative when data are merged."

The above statement indicates that in order to be in place to detect any potential gender effects, one needs to recruit an extremely large sample of subjects, for the standards of economic experimentation, a task which seems prohibiting given all the time, financial and practical constraints that a researcher may face. In this paper, motivated by the conclusions of Filippin and Crosetto (2016), we investigate how one can increase the extracted information from small sample datasets, and what are the implications of omitting to do so.

One of the most common approaches to explore potential differences between genders is to assume a particular preference functional, pool all the data together, and estimate a representative agent model, using demographic dummy variables to control for heterogeneity (see Harrison and Rutström, 2008, Xie et al. (2017), Vieider et al., 2015, Bouchouicha et al., 2019). Parameters are then obtained by using either Maximum Likelihood Estimation techniques (MLE) or Non-Linear Least Squares estimation methods, and the statistical significance of the dummies defines the existence and the size of potential differences. While the representative approach is attractive, due to its simplicity , it comes with a serious limitation. By ignoring individual heterogeneity, the estimated preferences may not be representative for any of the subjects. Consider an extreme scenario where out of 100 subjects, 50 are male and risk neutral, 25 female and risk seeking with a risk coefficient of -0.50 (assuming a power utility function as in Holt and Laury, 2002a and later in our analysis of the form $x^{1-r}/(1-r)$) and the remaining 25 subjects are females and risk averse, with a coefficient of 0.50. Pooling all the data together and fitting a representative agent model to this dataset, including a control variable to capture potential gender differences, will return an estimated risk aversion very close to zero, implying risk neutrality, and the coefficient of gender effects to be insignificant[1]. The main conclusion that a researcher could draw from a similar analysis is that the observed population has risk neutral preferences and there are no gender effects. Consider now a policy maker who aims to identify the risk seeking women in a population. By conducting a similar analysis, the policy maker will reach the conclusion that no risk seeking women exist in this sample and no action needs to be taken. While this example is extreme and perhaps improbable, it is used to highlight the impact of ignoring potential behavioural heterogeneity in identifying preferences and differences based on demographic criteria.

On the other end of the spectrum, one could estimate preference functionals at the individual subject-level (see Hey and Orme, 1994, Stott, 2006). While this approach takes into consideration the individual characteristics of each subject, a large amount of data points is required in order to obtain robust and reliable estimates. This comes at a high cost for the researcher, as larger number of decision tasks would mean longer sessions that could potentially lead to boredom and eventually to more

[1]We indeed executed a similar simulation exercise where this result was confirmed. Details are available on request.

noisy data.

In the present study we compare the representative agent modelling approach, to two more flexible and informative methods of parameter estimation that allow one to simultaneously make inferences at both the individual subject and the experimental population level. In particular, we compare the frequentist and the Bayesian methods, by analysing the data using Maximum Simulated Likelihood Estimation techniques (MSLE), as well as Hierarchical Bayesian (HB) econometric modelling. We use data from three prominent studies of decision making under risk. First, we use the original data from the Holt and Laury (2002a) experiment, and assuming Expected Utility preferences, we first show how all three inference methods are able to capture gender effects. Then, we extend our analysis to non-Expected Utility preferences, and particularly to Rank Dependent Utility, since our focus in on risky choice in the gains domain. Using the dataset from Baillon et al. (2020), we show that taking into consideration individual heterogeneity, improves the inference, while the MLE representative agent model fails to identify the existence of gender differences. Finally, we focus on the domain of losses, and adopting a Cumulative Prospect Theory framework, we explore the differences between the two genders, across all the components of risk preferences, namely utility curvature, probability weighting and loss aversion. We show how MLE fails to capture gender differences and we also focus on the differences between the MSLE and the HB methods in capturing these differences.

Our results can be summarised as follows. When there is a small number of parameters to estimate, any of the inference methods will be able to detect the presence of gender differences in the key behavioural parameters. As the model complexity increases, and therefore the number of parameters along with their collinearity, more flexible methods that take into consideration individual heterogeneity, provide more robust inference when the focus is on the difference between two populations. We complement our study with an extensive Monte Carlo simulation to compare the three inference methods, and we show that while all MLE, MSLE and HB methods are able to successfully recover the mean values of the simulated parameters, frequentist methods are more prone to ignore statistical significance due to overfitting, compared to Bayesian methods.

The rest of the paper is organised as follows: section 2.2 briefly introduces the idea of Hierarchical Bayesian modelling, section 2.3 focuses on the Holt and Laury (2002a) risk elicitation task and presents, along with the task and the data, the econometric specification for both MLE and HB, assuming Expected Utility preferences (EU), section 2.4 relaxes the hypothesis of EU and introduces Rank Dependent Utility preferences using data that allow the estimation of such preferences, and finally, section 2.5 focuses on the domain of losses, introducing a Cumulative Prospect Theory model and loss aversion. In section 2.6 we report the results of the simulation. We then conclude.

2.2 Frequentist Vs Bayesian Parameter Estimation

The most common approach to estimate structural decision making models is by either pooling all data together and fit a representative agent model, or by assuming complete independence and fit subject-level models, using maximum likelihood estimation techniques (MLE). Fitting a representative agent model ignores much of individual behavioural heterogeneity and generates estimates which potentially, are not representative of any individual subject in the sample. A simple way to introduce heterogeneity to the representative model is to condition the parameters to a set of observable demographics and assume that subjects that belong to the same demographic group share the same behavioural parameters (see for example Harrison and Rutström, 2008, Bouchouicha et al., 2019). An alternative way to introduce heterogeneity, within the frequentist framework, is to use a *random-coefficients* model, a popular method to model unobserved heterogeneity, on top of the observed one (e.g. through demographics). In this kind of modelling, it is assumed that each behavioural parameter in the model is characterised by an underlying distribution across the population. Using MLE techniques and simulation, it is possible to combine estimates of the population distribution (mean and standard deviation) with individual choices, and make inferences at both the population and the subject level (for applications see Gaudecker et al., 2011; Conte et al., 2011; Moffatt, 2016). Nevertheless, it is known that MLE is susceptible to overfitting and may generate noisy and unreliable estimates when there is a lack of a large number of observations (see

Bishop, 2006, pp. 166, Nilsson et al., 2011). An alternative method to introduce heterogeneity and mitigate these drawbacks is to adopt Hierarchical Bayesian estimation techniques (see Balcombe and Fraser, 2015; Ferecatu and Önçüler, 2016 and Baillon et al., 2020 for some recent applications of hierarchical models for choice models under risk and Stahl, 2014 for ambiguity models.). The key aspect of hierarchical modelling is that even though it recognises individual variation, it also assumes that there is a distribution governing this variation (individual parameter estimates originate from a group-level distribution). As Baillon et al. (2020) highlight, Hierarchical Bayesian modelling is a compromise between a representative agent and subject-level type estimation. It estimates the model parameters for each subject separately, but it assumes that subjects share similarities and draw their individual parameters from a common, population level distribution. In that way, individual parameter estimates inform each other and lead to a *shrinkage* towards the group mean that reduces biases in parameter estimates. The latter leads to more efficient and reliable estimates compared to those estimated using frequentist methods. One of the most crucial aspects of Bayesian inference, is the way uncertainty is incorporated in the econometric model in the form of probability distributions. A researcher can use her subjective beliefs or objective knowledge and form a prior distribution which summarises all the available knowledge regarding a particular parameter, before observing any data. In Bayesian inference, the estimation of a parameter of interest corresponds to the calculation of the probability distribution over the parameter, given the observed data and the prior beliefs. Another aspect of the Hierarchical model is that it is applied in an hierarchical form providing both within decision unit analysis (subject level) and across unit analysis (population level). Both the way the Bayesian model incorporates uncertainty and its Hierarchical structure, allows it generate precise estimation of preferences, even when the available data are limited.

Jacquement and L'Haridon (2018, p. 247) provide a comparison between the frequentist and Bayesian methods, highlighting the most important differences, namely the way each method interprets each parameter, the nature of the point estimation, the way intervals for statistical significance are estimated, and; the way hypothesis testing can be done. For the frequentist method the parameter is an unknown constant while for the Bayesian a random variable. Similarly, the point estimation will be the value of the estimator in the former, while a posterior summary in the latter (e.g. the mode of the distribution). For statistical significance, the frequentist method requires the estimation of confidence intervals, compared to the credible intervals in the Bayesian inference. As Huber and Train (2001) point out, in the presence of small samples, the two procedures can provide numerically different results, due to the different way of treating uncertainty in the parameters of the population distribution. In what follows, we compare the three different inference methods (MLE, MSLE and HB) in their capacity to detect gender differences, focusing on three representative examples of decision making under risk.

2.3 Risk Preferences and Expected Utility

Gender differences in risky decision making has been the topic of numerous studies. Eckel and Grossman (2008) and Croson and Gneezy (2009) summarise the literature, finding that female subjects tend to be more risk averse. Charness and Gneezy (2012) and Holt and Laury (2014) discuss how the risk elicitation task affects the inference on differences, while Filippin and Crosetto (2016) challenge the early evidence by finding that the observed effects are negligible in magnitude. In this section we focus on one of perhaps the most common elicitation methods that has been used in the literature, the Holt and Laury (2002a) task.

2.3.1 Decision Task and Data

For the analysis, we use the data from the original Holt and Laury (2002a) study. Each subject is presented with the 10 choice tasks, as shown in Table 2.1. Each task consists of a choice between two paired lotteries A and B. The payoffs for lottery A are fixed to \$2.00 and \$1.6, while for lottery B, the payoffs are 3.85 and \$0.10. Since lottery A is characterized by less variable payoffs, one can label A as the *safe* option and B the *risky* one. In the first choice task, the probability of getting the high payoff is equal to 10% for both lotteries, and it increases as one moves down the table. At the first row, only the extremely risk seeking subjects are expected to choose lottery B. A risk neutral person is expected to choose lottery A for the first 4 tasks (since the expected value of lottery A is greater) and then switches to lottery B for the remaining tasks. Holt and Laury (2002a), assuming a particular form of risky preferences, provide a mapping between then number of safe choices and the value of risk coefficient of a subject (the higher the degree of risk aversion, the higher the number of safe choices).

There are data from 212 subjects (95 females) from 4 treatments, an incentivised low-payoff treatment (LOW_1), with payoffs as those in Table 2.1, a hypothetical treatment (HYP), with the payoffs scaled up by 20, 50 or 90, an incentivised high-payoff treatment ($HIGH$), with payoffs scaled up by 20, and finally, a low-payoff treatment (LOW_2), identical to the first one. For our purposes, we use only the data from the low-payoff treatment (LOW_1).

TABLE 2.1: The 10 Lotteries from Holt and Laury (2002a).

	Option A				Option B			
Task	p_{A1}	x_{A1}	p_{A2}	x_{A2}	p_{B1}	x_{B1}	p_{B2}	x_{B2}
1	0.1	2.00$	0.9	1.60$	0.1	3.85$	0.9	0.10$
2	0.2	2.00$	0.8	1.60$	0.2	3.85$	0.8	0.10$
3	0.3	2.00$	0.7	1.60$	0.3	3.85$	0.7	0.10$
4	0.4	2.00$	0.6	1.60$	0.4	3.85$	0.6	0.10$
5	0.5	2.00$	0.5	1.60$	0.5	3.85$	0.5	0.10$
6	0.6	2.00$	0.4	1.60$	0.6	3.85$	0.4	0.10$
7	0.7	2.00$	0.3	1.60$	0.7	3.85$	0.3	0.10$
8	0.8	2.00$	0.2	1.60$	0.8	3.85$	0.2	0.10$
9	0.9	2.00$	0.1	1.60$	0.9	3.85$	0.1	0.10$
10	1.0	2.00$	0	1.60$	1.0	3.85$	0	0.10$

2.3.2 Theoretical Framework and Econometric Specification

We assume that the agent holds Expected Utility preferences and receives utility from income according to a Constant Relative Risk Aversion (CRRA) utility function of the form:

$$u(x) = \frac{x^{1-r}}{1-r} \tag{2.1}$$

where x is the monetary payoff, and r is the risk coefficient with $r > 0$ indicating a concave utility for gains (risk aversion), $r < 0$ a convex utility (risk seeking) and $r = 0$ a linear utility (risk neutrality). For $r = 1$ the function collapses to the logarithmic function. A lottery is evaluated by the weighted sum of the utilities of the payoffs, therefore, the expected utility of lottery A, for a particular task, is given by

$$EU_A = p_{A1}\frac{{x_{A1}}^{1-r}}{1-r} + (1-p_{A1})\frac{{x_{A2}}^{1-r}}{1-r} \tag{2.2}$$

To account for the stochastic nature in choices, we assume a *logit* link function. Thus, the probability of choosing lottery A is given by:

$$P(A) = \frac{\exp(1/\xi EU_A)}{\exp(1/\xi EU_A) + \exp(1/\xi EU_B)} \tag{2.3}$$

with ξ a precision parameter to be estimated. According to the above assumptions, the log-likelihood function is given by:

$$LL(\theta) = \sum_{n=1}^{N}\sum_{i=1}^{I} y_{ni}\ln(P_{ni}(A_i)) + (1-y_{ni})\ln(1-P_{ni}(A_i)) \tag{2.4}$$

where N is the total number of subjects, I is the number of tasks, $y_{ni} = 1(0)$ is an indicator function denoting the choice of lottery A(B) for subject n in task i, and θ is the vector of behavioural parameters to be estimated. Therefore, there are 2 parameters to estimate, the risk coefficient r and the precision parameter ξ. To introduce gender effects, we introduce a dummy variable y_{FEMALE} which takes the value 1 is the subject is female, otherwise it is equal to 0. For each parameter θ_n in our model, with $\theta_n \in \{r, \xi\}$ we specify

$$\theta_n = \theta_0 + \theta_{FEMALE} \times y_{FEMALE} \tag{2.5}$$

Since we consider a stochastic model which takes into consideration the errors of the decision maker, we include in the analysis the observations of all the subjects (rather than focusing only on subjects without multiple switches). There are in total

4 parameters to estimate, the risk coefficient, the precision parameter and the two parameters that capture gender effects. [2]

For the HB estimation, we follow Rouder and Lu (2005) and Nilsson et al. (2011) set-up. Each subject n made a series of I binary choices in a given dataset and the observed choices vector is denoted by $D_n = (D_{n1} \cdots D_{nI})$. Every subject is characterised by its own parameter vector $\Theta_n = (r_n, \xi_n)$, and we assume that both the utility curvature r_n and the sensitivity parameter ξ_n are normally distributed ($\theta_n \sim N(\mu_\theta, \sigma_\theta)$), while for the hyper-parameters we assume normal priors for the mean μ_θ and uninformative priors (uniform) for σ_θ. We also follow the standard procedure and transform all the parameters to their exponential form to ensure that they lie within the appropriate bounds (see Balcombe and Fraser, 2015). To capture gender differences, we condition the mean of all parameters to a female covariate. For each subject n, each parameter θ_n is assumed to be drawn from a normal distribution of the form: $\theta_n \sim N(\theta + \theta_{FEMALE} \times y_{FEMALE}, \sigma_\theta^2)$, with y_{FEMALE} a female dummy variable. That is, the mean between the two groups differs by θ_{FEMALE}. In what follows, we use either a normal or a log-normal distribution, depending on whether there are constraints for a parameter to be strictly positive.

The likelihood of subject's n choices is given by:

$$P(D_n|\Theta_n) = \prod_{i=1}^{I} P(D_{n,i}|\Theta_n)$$

where $P(D_{n,i}|\Theta_n)$ is given by:

$$LL(\theta) = \sum_{i=1}^{I} y_{ni} \ln(P_{ni}(A_i)) + (1 - y_{ni}) \ln(1 - P_{ni}(A_i)) \tag{2.6}$$

Combining the likelihood of the observed choices and the probability distribution of all the behavioural parameters, the posterior distribution of the parameters is given by:

$$P(\Theta|D) \propto P(D|\Theta) \times P(\Theta)$$

[2]For the estimation we use a general nonlinear augmented Lagrange multiplier optimisation routine that allows for random initialisation of the starting parameters as well as multiple restarts of the solver, to avoid local maxima. The estimation was conducted using the *R* programming language for statistical computing (The *R* Manuals, version 3.6.1. Available at: http://www.r-project.org/).

with $P(D|\Theta)$ being the likelihood of observed choices over all the subjects and $P(\Theta)$ the priors for all parameters in the set Θ.[3]

Monte Carlo Markov Chains (MCMC) were used to estimate all the specifications. The estimation was implemented in JAGS (Plummer, 2017). The posterior distribution of the parameters is based on draws from two independent chains, with 50,000 MCMC draws each. Due to the high level of non-linearity of the models, there was a burn-in period of 25,000 draws, while to reduce autocorrelation on the parameters, the samples were thinned by 10 (every tenth draw was recorded). Convergence of the chains was confirmed by computing the $\hat{R}$ statistic (Gelman and Rubin, 1992).

Finally, for the MSLE we follow Train (2009) and Moffatt (2016) and we estimate the models with the help of simulation. As mentioned before, in this random-coefficient model, the behavioural parameters for a given subject are fixed and they vary across the experimental population according to a distribution (usually assumed Normal).[4] Assuming that a parameter θ is drawn from a distribution with density $g(\theta)$, for a set of I choices, the likelihood of subject's n choices is given by:

$$LL(\theta) = \int \left[\prod_{i=1}^{I} P_{ni}(A_i)^{y_{ni}} \times (1 - P_{ni}(A_i)^{1-y_{ni}}) g(\theta) \right] d\theta \tag{2.7}$$

and the total log-likelihood is given by the sum of the logarithm of (2.7) across all subjects. The parameter θ is distributed over subjects according to the density function $g(\theta)$, and is known as the subject-specific random effect. The variation in θ captures the between-subject heterogeneity. When there are more than one parameters θ, the distribution $g(\theta)$ is a multivariate distribution and the integral is multidimensional. Therefore, the challenge for the estimation method is how to evaluate the integral in (2.7), since there is no analytical solution. In our analysis, we resort to simulation to approximate the integral, using Maximum Simulated Likelihood Estimation techniques. We use 100 Halton draws per subject. Following Conte et al. (2011), we assume the stochastic parameter ξ to be constant. For the Expected Utility model we therefore estimate 5 parameters, the mean and standard deviation of the

[3]These priors were chosen to allow the data to dominate, and as such, the results of these estimations should not be sensitive to changing priors, provided these are not strong priors.

[4]This is standard practice in the literature, and estimations should not be sensitive to this assumption. However, we have not explored this further as it is out of scope of this paper.

risk coefficient r, the precision parameter ξ and the gender effects for both parameters.

2.3.3 Results

Table 2.2 reports the estimates from the three inference methods. The first column reports the results from the MLE, the middle from the MSLE and the last one from the HB model. For each parameter θ, we report the point estimate for the MLE, the mean of the distribution μ_θ for the MSLE, and the mode of the posterior distribution for the HB. The standard errors are reported in the Table, with the exception of the HB model where the standard deviation of the posterior distribution of each parameter is reported instead. The statistical significance is based on the respective confidence intervals (credible intervals for the HB).

TABLE 2.2: Estimates using the Holt and Laury (2002a) data.

	MLE	MSLE	HB
r	0.289^{***}	0.265^{***}	0.292^{***}
s.e.	0.022	0.021	0.033
r_{FEMALE}	0.103^{***}	0.103^{***}	0.103^{**}
s.e.	0.034	0.021	0.050
σ_r	-	0.192	-
s.e.	-	0.000	-
ξ	0.253^{***}	0.184^{***}	0.086^{***}
s.e.	0.014	0.016	0.014
ξ_{FEMALE}	0.050^{**}	0.136	0.041
s.e.	0.023	0.193	0.333

The Table reports estimates from all three inference methods: Maximum Likelihood Estimation (MLE), Maximum Simulated Likelihood Estimation (MSLE) and Hierarchical Bayesian (HB). For each parameter θ, the Table reports the point estimate for the MLE, the mean of the distribution μ_θ for the MSLE, and the mode of the posterior distribution for the HB. Standard errors are reported (standard deviation for the HB). *p<0.1; **p<0.05; ***p<0.01

In all cases, the risk coefficient is positive and statistically significant, indicating risk averse preferences for all subjects. The coefficient of risk aversion ranges between 0.265 and 0.292 between the three inference methods, what Holt and Laury (2002a) characterise as "slightly risk averse". Focusing on the gender effects parameters, the coefficient is positive and statistically significant in all three cases, and remarkably at the same magnitude of 0.103. Finally, focusing on the precision parameter ξ, the effect of introducing more flexible inference methods to its magnitude, is apparent. The estimate of ξ using MLE is equal to 0.253 which is quite large compared to the other two methods. Since there is an inverse relationship between the size of ξ and the estimated noise (the lower the ξ the higher the precision) a larger estimate of ξ indicates issues with overfitting. As the inference methods become more flexible, the estimate of ξ becomes smaller, indicating more precise and less noisy estimates. The main conclusion from this analysis, is that by ignoring the between-subject heterogeneity, and estimating a model assuming a basic level of heterogeneity, as in the case of the MLE estimation, it is possible to detect the existence of gender differences, regardless of which estimation method is adopted. In what follows, we explore whether this result can be generalised when the complexity of the model increases. Filippin and Crosetto (2016) extend their analysis and investigate whether relaxing the expected utility assumption, has an effect to the inferred gender differences. By introducing a probability weighting function and non-expected utility preferences, they estimate a structural specification, using MLE, and show that the gender differences in the risk coefficient disappear, and they appear in the probability weighting parameter. As the original Holt and Laury (2002a) task was not developed with non-expected utility preferences in mind, in the next section we repeat the same analysis as above, using data from an experiment which was particularly developed to identify risk preferences, stemming from both the curvature of the utility function and the shape of the probability weighting function.

2.4 Risk Preferences and Rank Dependent Utility

Motivated by the Allais paradox, a vast theoretical and experimental literature emerged, challenging the assumption of expected utility preferences (see Starmer, 2000 for a review of non-EU theories; Camerer, 1995 for an early discussion of the experimental work; and Hey, 2014 for a more recent review). In this section, we focus on one of the most influential alternatives to EU, the Quiggin (1982) Rank Dependent Utility model (RDU) which later led to the modification of the Original Prospect Theory model and the development of the Tversky and Kahneman (1992) Cumulative Prospect Theory model (which we explore in the next section). In the RDU model, attitudes towards risk are characterised by both the curvature of the utility function, and the shape of the probability weighting function, while there is evidence that the two components are not strongly correlated (Qui and Steiger 2011; Toubia et al. 2013). Therefore, given the extensive empirical evidence of the existence of non-EU preferences, it is crucial to take both components into consideration, when one investigates the existence of gender differences in risk preferences. We do so by using the data from Baillon et al. (2020).

2.4.1 Decision Task and Data

Objective of this experiment was to identify the reference point that subjects are using when they make choices under risk. Each experimental task involved a choice between two paired lotteries again, A and B. An optimal design was employed to construct the questions of the experiment in a way that they would satisfy the following 5 criteria: (1)the questions must be diverse in terms of number of outcomes and magnitudes of probabilities involved, (2)the questions within each choice must have nonmatching maximal or minimal outcomes, (3) the questions must be diverse in terms of relative positioning in the outcome space, (4) they must have similar expected value to avoid trivial or statistically noninformative choice situations, and; (5) they must be "orthogonal" in some sense to maximise statistical efficiency. The number of the outcomes within each lottery varied between tasks, from 2 to 4 outcomes, all in the gains domain (strictly positive). An example of a task is provided below:

$$A = \begin{cases} 135, & \text{with probability } 0.55 \\ 290, & \text{with probability } 0.35 \\ 329, & \text{with probability } 0.10 \end{cases} \quad B = \begin{cases} 159, & \text{with probability } 0.05 \\ 259, & \text{with probability } 0.55 \\ 359, & \text{with probability } 0.10 \\ 409, & \text{with probability } 0.30 \end{cases}$$

The order of the tasks was randomised, and there was a total of 70 tasks per subject, with varying payoff and probability levels, generating a rich dataset for structural estimations. There are in total data from 139 subjects (49 females).[5] The experimental population consisted of students in Moldova, and the payoffs were expressed in the local currency. To incentivise the experiment, each subject had a one-third chance to be selected among all the subjects, to play out one of their choices for real. The experiment involved high stakes with payoffs up to a week's salary.

2.4.2 Theoretical Framework and Econometric Specification

As mentioned before, the RDU model consists of two components, the utility function and the probability weighting function, which transform every objective probability p to the decision weight $w(p)$ in the interval $[0, 1]$. We again assume a CRRA utility function, while for the probability weighting function, we assume the widely used Tversky and Kahneman (1992) function of the form:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{2.8}$$

where γ is the probability weighting parameter. The form of the function is inverse-S shaped for $\gamma < 1$, indicating overweighting of low probabilities and underweighting of moderate and high probabilities. To evaluate the RDU of a lottery, we first need to rank the outcomes of the lottery from the best to the worst, such that $x_1 \geq x_2, \cdots, \geq$

[5]For our analysis we use the data from 136 subjects as there were missing data on the gender of 2 subjects, and 1 subject had missing data.

x_n. The decision weight associated with each outcome is given by:

$$\begin{aligned}\pi(x_1) &= w(p_1)\\ \pi(x_2) &= w(p_1+p_2)-w(p_1)\\ &\dots\\ \pi(x_n) &= 1-w(p_1+p_2+\cdots+p_n)\end{aligned}$$

The RDU of lottery A is then given by

$$RDU(A)=\sum_{n=1}^{N}\pi(x_n)\frac{{x_n}^{1-r}}{1-r} \tag{2.9}$$

We assume the same stochastic function as in Equation 2.3, by replacing the expected utility with the corresponding Rank Dependent Utility, and we form the log-likelihood function as in Equation 2.4 for the MLE estimation. As there are no multiple treatments, we control only for gender differences by introducing a gender dummy for all the parameters (r, γ, ξ, giving in total 6 parameters to estimate).

For the HB model, on top of the specifications for r and ξ, which are exactly the same as in the EU case, we need an additional specification for the γ parameter. This parameter must be positive, with a lower bound equal to 0.279 to ensure the monotonicity of the function. For the MSLE estimation, we need to estimate the parameters of the two distributions for r and γ, namely the means μ_r and μ_γ and their standard deviations σ_μ and σ_γ[6].

2.4.3 Results

Table 2.3 reports the results from all the three inference methods. The results are quite similar to what is usually observed in this literature. The estimated risk coefficient r is between 0.360 and 0.480, indicating moderate risk averse preferences, while the estimate for the probability weighting function is equal to 0.586 and 0.621, indicating an inverse-S shape of the function. While these results are quite uniform

[6] In the framework of MSLE, the coefficient vector θ is assumed to be normally distributed, across the population, with mean equal to a vector b and covariance matrix W. To maintain a manageable number of parameters, we assume that the off-diagonal elements of W are equal to zero and estimate the variance of each distribution. Allowing for correlation between the parameters led to worse performance of the model.

and the estimates look quite close in terms of magnitude and statistical significance, there are contradictory results regarding the presence of gender effects. Assuming heterogeneity only at the gender level (MLE) fails to capture any kind of effects for any of the parameters, while a same pattern is observed when MSLE is used to estimate the model. Nevertheless, when HB is used, one can infer that there is a significant difference between males and females in the way objective probabilities are transformed. With an estimate of γ equal to 0.705 (compared to 0.621 for men), it seems that women tend to exhibit lower probability distortion. Again, the effect of the different estimation methods on the precision parameter ξ is similar as in the Expected Utility case (the noise in the estimates decreases when more flexible inference methods are introduced).

TABLE 2.3: Estimates using the Baillon et al. (2020) data.

	MLE	MSLE	HB
r	0.479***	0.424***	0.360**
s.e.	0.032	0.038	0.029
r_{FEMALE}	-0.025	-0.030	-0.054
s.e.	0.111	0.203	0.099
σ_r	-	0.291**	-
s.e.	-	0.122	-
γ	0.598***	0.586**	0.621***
s.e.	0.016	0.251	0.029
γ_{FEMALE}	0.050	−0.060	0.084*
s.e.	0.032	0.244	0.049
σ_γ	-	0.943	-
s.e.	-	0.821	-
ξ	0.121***	0.076***	0.083***
s.e.	0.001	0.007	0.010
ξ_{FEMALE}	0.041	-0.003	0.026***
s.e.	0.044	0.013	0.005

The Table reports estimates from all three inference methods: Maximum Likelihood Estimation (MLE), Maximum Simulated Likelihood Estimation (MSLE) and Hierarchical Bayesian (HB). For each parameter θ, the Table reports the point estimate for the MLE, the mean of the distribution μ_θ for the MSLE, and the mode of the posterior distribution for the HB. Standard errors are reported (standard deviation for the HB). *p<0.1; **p<0.05; ***p<0.01

The analysis above provides an example of the implications of ignoring heterogeneity between (as well as within) participants. While a basic MLE estimation provides no evidence of any kind of gender differences, allowing for a more informative approach reveals the existence of such differences. In the next section, we extend our analysis to one of the most important domains of decision theory under risk, that of loss aversion.

2.5 Risk Preferences and Loss Aversion

In this section we focus on the three components that characterise risk preferences in the losses domain, as these are articulated in the Tversky and Kahneman (1992) Cumulative Prospect Theory (CPT) model. The CPT model adopts a similar approach to the RDU model, on the way it handles monetary payoffs and probabilities, with the additional feature of *loss aversion*, the concept that "losses loom larger than gains". The results from the literature are mixed. Some studies find that women are more less loss averse (see Schmidt and Traub, 2002, Brooks and Zank, 2005), others that males are more loss averse (Booij et al., 2009), others that there is no difference (Harrison and Rutström, 2008), and others with a mixed result (Bouchouicha et al., 2019). As Bouchouicha et al. (2019) argue, currently, there is no consensus of what is the appropriate definition of loss aversion in the literature[7]. Nevertheless, for the sake of the example, we will focus on the CPT definition of loss aversion, while our approach can be extended to alternative definitions.

2.5.1 Decision Task and Data

To estimate a CPT specification when losses are present, we use the data from Bouchouicha et al. (2019) which is a subset from the data used in Vieider et al. (2015). There are in total observations of almost 3000 subjects, from 30 countries, on decision making under risk and ambiguity, in both the gains and losses domain. As our focus is on small samples, we use only the USA data. This set includes the choices of 95 subjects (47 females) in 12 choice tasks (6 in the gains domain, 5 in the losses domain, and 1 in the mixed domain to identify the loss aversion parameter). While there are available data on a larger set of risky tasks (28 tasks), we follow Bouchouicha et al. (2019) and use only the smaller subset for two reasons: (1) this set of tasks includes only 50:50 gambles, which allows the estimation of a functional-free probability weighting function, and; (2) estimating a structural model from a small set of observations per participant is one of the strengths of the Hierarchical approach, and this dataset allows to test the limits of this approach.

[7] See Schmidt and Zank (2005) for the various definitions of loss aversion.

All tasks are in the form (x, y), representing the prospect of getting the monetary payoff x with probability 50% or y with the residual probability, with x and y being positive, negative or zero, depending on the task (Table 2.4 lists the 12 tasks). The subject had to express her *certainty equivalent* for each of the tasks. For the mixed domain prospect, the amount l was elicited, that would make the subject indifferent between a 50:50 gamble of (20,l) and the status quo of zero. The experiment was incentivised and an endowment equal to the largest possible loss was provides to the subject, to cover for potential losses.

TABLE 2.4: The tasks from Bouchouicha et al. (2019)

Gains	Losses	Mixed
(5,0)	(-5,0)	(20,-l)
(10,0)	(-10,0)	
(20,0)	(-20,0)	
(30,0)	(-20,-5)	
(30,10)	(-20,-10)	
(30,20)		

2.5.2 Theoretical Framework and Econometric Specification

We model preferences assuming a CPT decision maker. We employ a power utility function as before, of the form:

$$u(x) = \begin{cases} \frac{x^{1-r}}{1-r}, & \text{if } x \geq 0 \\ -\lambda \frac{(-x)^{1-r}}{1-r}, & \text{if } x < 0 \end{cases}$$

with r the risk coefficient, and λ the parameter of loss aversion. The status quo of zero is assumed as a reference point. We assume a common parameter for r for gains and losses, for two reasons: (1) there is extensive empirical evidence of no difference between the two domains (see Fox and Poldrack 2009), and; (2) to avoid any potential identification issues of the loss aversion parameter (see Wakker, 2010). As mentioned before, since only 50:50 gambles are used in the analysis, there is no need to specify a functional form for the probability weighting function. Therefore, we introduce two parameters to estimate, w_g and w_l, which represent the probability weighting for gains and losses respectively. Summarising, a prospect $L = (x, y)$ can be evaluated as:

$$U(L) = w_s u(x) + (1 - w_s)u(y)$$

with $s \in \{g, l\}$, while for the mixed prospect $L = (x, l)$, the prospect is evaluated as:

$$U(L) = w_g u(x) + w_l u(l)$$

The certainty equivalent $\hat{ce}$ for a prospect L is then given by:

$$\hat{ce} = u^{-1}[w_s u(x) + (1 - w_s)u(y)]$$

To form the likelihood function we need a different approach to the one used in the previous sections. In particular we assume that a decision maker states her certainty equivalent with some noise. The observed certainty equivalent of a subject in a task i is equal to $ce_i = \hat{ce} + \varepsilon_i$, where $\hat{ce}$ is the theoretical optimal certainty equivalent, for a set of behavioural parameters, and $\varepsilon \sim \mathcal{N}(0, \xi^2)$ with ξ being the standard deviation of the Fechner error (see Hey and Orme, 1994). We assume that this error is domain-specific (for mixed gambles we use the error for losses) and we also take into consideration a *contextual* error Wilcox 2011 by making the parameter ξ to be dependent on the difference between the best and the worst outcome of each prospect. That is, $\xi_i = \xi|x_i - y_i|$. The loglikelihood function for N subjects and I tasks is then given by:

$$LL(\theta) = \sum_{n=1}^{N} \sum_{i=1}^{I} \ln[\psi(\theta_n, L_i)] \tag{2.10}$$

with θ a vector of behavioural parameters to be estimated, L_i a task i and ψ the contribution to the likelihood function given by:

$$\psi(\theta_n, L_i) = \phi\left(\frac{\hat{ce}_{ni} - ce_{ni}}{\xi_{nis}}\right)$$

where ϕ is the standard normal density function. For the MLE estimation, we follow Bouchouicha et al. (2019) and we assume heterogeneity of the parameters at the gender level, and the domain level for the decision weights and the precision parameters. We need to estimate 12 parameters in total ($r, \lambda, w_g, w_l, \xi_g, \xi_l$ along with the controls for gender).

For the HB model, the specification of the likelihood function remains the same as in the MLE case. We specify distributions for the six parameters as above, with the decision weights constrained to the interval $[0, 1]$ and the loss aversion parameter to the interval $[0, 10]$, while for the MSLE, we estimate the parameters of the distributions for the risk attitude, the loss aversion and the probability weighting for gains and losses.

2.5.3 Results

Table 2.5 reports the estimates from the three inference methods. Three points are worth to mention: (1) there is significant loss aversion in this sample with a λ parameter statistically significant ranging between 1.596 and 1.672, (2) the risk coefficient is not statistically different than zero for the MLE and the HB cases, indicating a linear utility function, (3) the probabilities in the gains domain are distorted more that the probabilities in the losses domain (for instance, the decision weight of 0.5 is estimated to be 0.426 for gains and 0.478 for losses in the MLE case), and; (4) the control coefficient for gender differences is insignificant for all the major parameters of interest, in the MLE case with the exception of the noise parameter. Once again, using MLE techniques, one can conclude that there are no gender differences in the way females and males perceive monetary outcomes, transform probabilities to decision weights or perceive losses. Focusing on the more flexible methods of MSLE and HB, two points are interesting. First, the estimates of the mean, for all the parameters, are remarkably close between the two methods reinforcing the result of Huber and Train (2001). Nevertheless, when gender effects are considered, while both methods find differences in the loss aversion parameters between the two groups, the MSLE methods fails to detect any gender effects in the key parameter or risk attitude. A potential explanation for this result could be the larger estimate of the precision parameter (a lower value indicates more precise estimates).

TABLE 2.5: Estimates using the Bouchouicha et al. (2019) data.

	MLE	MSLE	HB
λ	1.596***	1.672***	1.615***
s.e.	0.099	0.112	0.113
λ_{FEMALE}	0.321	0.430***	0.398***
s.e.	0.211	0.141	0.136
σ_λ	-	0.467***	-
s.e.	-	0.044	-
r	-0.133	0.146**	-0.026
s.e.	0.07	0.061	0.019
r_{FEMALE}	-0.011	-0.061	−0.082*
s.e.	0.115	0.083	0.041
σ_r	-	0.000	-
s.e.	-	0.053	-
w_g	0.426***	0.433***	0.444***
s.e.	0.019	0.023	0.017
$w_{g FEMALE}$	-0.009	-0.032	-0.037
s.e.	0.031	0.031	0.056
σ_{w_g}	-	0.38***	-
s.e.	-	0.051	-
w_l	0.478***	0.467***	0.501***
s.e.	0.019	0.022	0.010
$w_{l FEMALE}$	0.007	-0.010	-0.011
s.e.	0.031	0.031	0.037
σ_{w_l}	-	0.436***	-
s.e.	-	0.048	-
ξ	0.173***	0.178***	0.097***
s.e.	0.007	0.012	0.012
ξ_{FEMALE}	0.028***	0.000	0.051
s.e.	0.012	0.000	0.261
ξ_l	0.150***	0.112	0.063***
s.e.	0.006	0.008	0.007
$\xi_{l FEMALE}$	0.033***	0.000	0.055
s.e.	0.011	0.000	0.311

The Table reports estimates from all three inference methods: Maximum Likelihood Estimation (MLE), Maximum Simulated Likelihood Estimation (MSLE) and Hierarchical Bayesian (HB). For each parameter θ, the Table reports the point estimate for the MLE, the mean of the distribution μ_θ for the MSLE, and the mode of the posterior distribution for the HB. Standard errors are reported (standard deviation for the HB). *p<0.1; **p<0.05; ***p<0.01

In this additional example, we provide further evidence that as the model complexity increases, by ignoring heterogeneity at the subject level, it may lead to incorrect inference regarding the difference between different demographic groups. Both methods that allow for this kind of heterogeneity (MSLE and HB) managed to detect the existence of such effects. Nevertheless, the results are not uniform. To identify which method is the most appropriate to use, in the next section we report the results of an extensive simulation exercise were we compare the performance of each of the methods.

2.6 Exploring the Advantages of HB Modelling

In the previous sections, we have shown that the identification on gender effects largely depends on the adopted inference method. We have provided a rigorous comparison of the representative agent model against two alternative methods that allow for extensive behavioural heterogeneity, even when the available sample size is small. Given that all three methods result is quantitatively different estimates, it raises the question of which method should one adopt. In this section we aim to provide an answer to this question, by means of an extensive Monte Carlo simulation exercise. Several studies have focused on the comparison between classical and Bayesian estimates, providing support on the latter (see for example Nilsson et al., 2011 or Gao et al., 2020). Here we repeat a similar exercise, suitably adapted to our objective of identifying gender differences in the elicited behaviour.

The main goal of this simulation study is two-fold. First, we want to confirm whether all estimation procedures are able to accurately recover the true parameter values from simulated data. Secondly, we test whether the inference methods under consideration, are equally efficient in detecting gender effects. To make the simulation as general as possible, we focus on the Bouchouicha et al. (2019) design and the CPT model, which satisfies the conditions for which researchers usually resort to pool their data (a relatively large number of parameters to estimate using a relatively low number of data points per subject). For our exercise, we simulate data of 100 subjects which we then estimate using each of the three inference methods: MLE, MSLE and HB.

We assume that gender differences exist only in two of the model's parameters, the coefficient of loss aversion and the risk coefficient[8]. The parameters used in the simulation, are normally distributed across the experimental population with mean θ_n and standard deviation σ_θ. In the simulation we set the gender difference in the risk coefficient to be small but significant (mean of 0.500 for males and 0.600 for females) with a standard deviation equal to 0.05[9]. The loss aversion is set to 1.648 for males and 2.013 for females[10] with a standard deviation of 0.100. The probability weighting coefficient for gains w_g is set equal to 0.540 while the probability coefficient for losses w_l is set equal to 0.510. We assume no heterogeneity by setting the standard deviation equal to 0 for the weighting parameters and we also assume a common Fechnerian error for gains and losses. We conducted the simulation for three different levels of noise by setting the value of the error term equal to 0.130 (low noise), 0.150 (medium noise) and 0.200 (high noise). We report the results of the medium noise specification as they are the most representative[11]. For each simulation, we generate the data of the 100 artificial subjects by drawing parameters from the relevant distributions that were described above. This dataset was then estimated using each of the methods. Table 2.6 reports the results of 100 simulations. In particular, we report the mean and the standard deviation of the point estimates, in the case of MLE, the mean of the distribution means in the case of MSLE, and the mean of the posterior means of the distributions in the case of HB.

[8]We make this assumption in order to keep the simulation as simple as possible. Of course this analysis can be extended to any of the parameters of the model (i.e. probability weighting function, noise coefficient) since empirically, gender effects are observed in all components of preferences.

[9]We confirmed that the statistical significance of the two distributions is indeed significant based on a two-sided t-test ($p<0.000$).

[10]Since we transform the parameters to be drawn from a log-normal distribution, the values of loss aversion correspond to $\exp(0.500)$ for men, and $\exp(0.700)$ for women.

[11]Bouchouicha et al., 2019 using this dataset, estimate the noise parameter to be equal to 0.170. For our simulations, we are using a noise parameter of 0.150, which is in the middle of the interval between the low noise parameter (0.130) and the empirically observed parameter (0.170). We delegate the estimates from the low and high noise simulations to the online Appendix (see Tables B.1 and B.2).

TABLE 2.6: Mean and standard deviations of the parameters.

Parameter	True value	MLE	MSLE	HB
λ	1.648	1.575	1.637	1.657
s.e.	-	0.056	0.062	0.067
λ_{FEMALE}	0.365	0.698	0.374	0.389
s.e.	-	0.132	0.118	0.129
σ_λ	0.100	-	0.081	-
s.e.	-	-	0.046	-
r	0.500	0.538	0.500	0.493
s.e.	-	0.019	0.031	0.025
r_{FEMALE}	0.100	0.064	0.105	0.108
s.e.	-	0.034	0.031	0.032
σ_r	0.050	-	0.046	-
s.e.	-	-	0.014	-
w_g	0.540	0.559	0.543	0.536
s.e.	-	0.013	0.014	0.014
w_l	0.510	0.528	0.510	0.510
s.e.	-	0.014	0.013	0.014
ξ	0.150	0.153	0.150	0.148
s.e.		0.007	0.005	0.006

The Table reports estimates from the simulation exercise on the three inference methods : Maximum Likelihood Estimation (MLE), Maximum Simulated Likelihood Estimation (MSLE) and Hierarchical Bayesian (HB), for the medium level of noise. For each parameter θ, the Table reports the mean of the point estimates, in the case of MLE, the mean of the distributions in the case of MSLE, and of the posterior mean of the distributions in the case of HB. Standard deviations in parentheses.

We first focus on the parameter recovery performance of each of the methods. The first column of the Table reports the true values of the coefficients that were used in the simulation. Compared to the true value, it is apparent that the MLE estimates have the worst performance in terms of precision. First, most of the parameters deviate significantly from the true value, compared to the other two methods. Then, in terms of gender effects, there is significant overestimation of the difference in loss aversion where the parameter is estimated to be almost twice the true value (0.698 compared to the true value of 0.365) while there is underestimation of the difference in the risk coefficient (0.064 compared to the true value of 0.100). As far as the MLE and MSLE estimates are concerned, both are remarkably close to each other and both have recovered the true parameters with quite high precision. The first conclusion from this simulation exercise is that if one is interested in the mean values of the parameters of different groups, then both MSLE and HB are equally good in recovering unbiased parameter values compared to the MLE.

We now turn to the identification of gender effects. For each of the simulations, we generate the 95% confidence interval (credible interval in the case of HB) to test the statistical significance of the estimate. When we focus on the gender effect for the risk coefficient, the MLE estimate is statistically significant for 55% of the simulations, the MSLE for 66% while the HB for 96%. Similarly, when we focus on the loss aversion parameter the MLE estimate is statistically significant for 53% of the simulations, the MSLE for 67% while the HB for 89%. Table 2.7reports the frequency with which statistically significant gender effects were detected, for each of the three inference methods, and for each of the three levels of noise (low, medium and high). The Table confirms the pattern that higher levels of noise lead to lower detection levels of gender effects, with MLE having the worst performance, HB the best, and MSLE in the between.

TABLE 2.7: Identification of gender effects.

		r_{FEMALE}	
	$\xi = 0.130$	$\xi = 0.150$	$\xi = 0.200$
MLE	62%	55%	36%
MSLE	82%	66%	59%
HB	98%	96%	76%
		λ_{FEMALE}	
	$\xi = 0.130$	$\xi = 0.150$	$\xi = 0.200$
MLE	63%	53%	37%
MSLE	81%	67%	50%
HB	97%	89%	66%

The Table reports the rate of success of each inference method to identify gender effects for each of the three levels of noise, for the gender specific parameter for risk attitude (r_{FEMALE}) and loss aversion (λ_{FEMALE}) are statistically significant, at the 5% level.

Our results mirror the conclusions of Huber and Train (2001). In this study the authors compare classical and Bayesian estimates by providing a comparison between MSLE and HB. They show that both methods result in virtually equivalent conditional estimates of the parameters. Then, they provide a list of differences between the two methods including (1) the difficulty of MSLE to locate the maximum of the likelihood function; (2) the computational burden that the variance-covariance matrix poses to the estimation of the MSLE parameters, and; (3) the identification issues that the classical approach faces compared to the Bayesian estimation. Our simulation shows that when the identification of differences between different populations is the objective, then HB is the clear winner as the most appropriate inference method. This result can be attributed to the way each of the methods handles uncertainty in the estimates and the fact that the estimate of the unobserved heterogeneity in the MSLE estimates is much noisier (larger standard errors) compared to the HB ones.

To investigate the role of the sample size in the detection of gender effects, we ran some additional simulations for the MSLE methods, varying the sample size.

Assuming a fixed level of noise ($\xi = 0.150$), we repeated the simulation exercise for $N = 200$ and $N = 500$ and again we report the rate of success to identify gender effects for the risk attitude (r_{FEMALE}) and the loss aversion (λ_{FEMALE}) gender specific parameters[12]. When the sample size is equal to 200, the risk (loss aversion) coefficient is significant for 89% (90%) of the simulations, while when the sample size increases to 500, the risk (loss aversion) coefficient is significant for 93% (94%) of the simulations. This analysis further highlights the advantages of the HB modelling since this inference method needs only half of the sample that MSLE needs in order to achieve the same detection rate of success in the case of loss aversion, while it needs only one fifth of the sample that MSLE needs, to reach the same success rate, in the case of the risk coefficient.

2.7 Concluding remarks

In this study, we focus on gender differences and compare the inference made by three econometric methods, Maximum Likelihood Estimation, Maximum Simulated Likelihood Estimation and Hierarchical Bayesian modelling, on three representative domains of risk preferences. We show that when all the data are assumed to come from a representative agent, and assume heterogeneity (gender differences or any other demographic differences) at a very basic level (e.g. all black females have the same level of loss aversion), valuable information might be ignored, and therefore, distorted conclusions may be drawn. Nevertheless, opting for a more flexible approach, and taking into consideration both the individual variation and the population-level characteristics, the inference about individual risk preferences is massively improved, and significant differences are captured.

In particular, we compare the representative agent modelling approach, to two more flexible and informative methods of parameter estimation that allow one to simultaneously make inferences at both the individual subject and the experimental population level. We compare the frequentist and the Bayesian methods, by analysing the data using Maximum Simulated Likelihood Estimation techniques (MSLE), as well as Hierarchical Bayesian (HB) econometric modelling. We use data

[12]The estimates are delegated to the online Appendix (see Table B3). There, it can be seen that as the sample size increases, the standard errors decrease, which allows for better identification of the effects.

from three representative studies on decision making under risk and we study Expected Utility preferences, for a simple analysis of risk attitudes, Rank Dependent Utility preferences, to incorporate probability weighting, and Cumulative Prospect Theory, to investigate loss averse behaviour. We show that by ignoring heterogeneity at the subject level, it may lead to incorrect inference regarding the difference between distinct demographic groups.

Recent research on Hierarchical Bayesian modelling has shown that MLE estimates are both susceptible to overfitting and dominated by outliers (Nilsson et al., 2011, Murphy and Brincke, 2018), while Bayesian modelling improves the robustness of the estimation, by shrinking the parameters towards the group's mean. This method allows the robust estimation of preferences, and it is particularly useful, especially when one has a limited number of data points from each subject, as is often the case with field studies, or when additional tasks are used, along with the main experimental design, to control for particular preferences. With the aid of an extensive simulation exercise, we show that Bayesian methods are better placed to capture differences between groups, and this result can be attributed to the way that each of the methods handles uncertainty in the estimates.

In this study, we do not argue in favour of any particular preference functional or model, nor we claim that there is a uniform pattern of gender differences. In our analysis, we opted for the models and the preference functionals that are often assumed in this literature. These models acted as "vehicles" to illustrate the machinery behind both estimation techniques, and this approach could be extended to any alternative model. Our main objective is to warn researchers on the dangers of small sample datasets and ignoring heterogeneity of the subjects. Of course this method could be extended to other important fields of decision making such as ambiguity preferences, time preferences or social preferences. Even more, as Gao et al. (2020) highlight, HB methods are particularly useful when one is interested in joint estimation of perhaps non-correlated preferences (e.g. joint estimation of risk and time preferences) where the need of robust estimates is important at the individual level.

Chapter 3

Wishful Thinking in Strategic Games

3.1 Introduction

Many biases are difficult to quantify, and often require complicated statistical techniques, or specifically designed experiments to quantify, e.g. overconfidence and hindsight bias. A similar bias, wishful thinking, describes a cognitive bias which occurs when an individual believes something to be true, regardless of the likelihood of the event occurring. It is worth noting that whilst wishful thinking is similar to optimism bias, wishful thinking occurs due to hope, e.g. I know I will not get into an accident drunk driving; whereas optimism bias is driven by distorted relative assessment, e.g. I know accidents happen when drunk driving, but it won't happen to me. Overconfidence also behaves similarly, but hinges on overestimating one's abilities, knowledge, accuracy, or control. For example, I will not crash while drunk driving because I am a better driver than anyone else. In practice, overconfidence will happen in situations where individuals perceive that skill, knowledge, or control is required, leading them to overestimate their own abilities, whereas wishful thinking and optimism bias can occur regardless of these factors being absent. The game below illustrates an example of when wishful thinking and optimism give different outcomes.

	B: Swerve	B: Straight
A: Swerve	(0, 0)	(-1, 1)
A: Straight	(1, -1)	(-10, -10)

In this game, optimism bias and wishful thinking could lead to different strategies due to the distinct psychological mechanisms underlying each strategy. A player that exhibits wishful thinking may choose to swerve, prioritizing the mutually desirable outcome of avoiding the worst-case scenario (i.e. both players crashing). This decision is driven by the hope that the opponent will also swerve. In contrast, a player that exhibits optimism bias may be more likely to choose “Straight,” based on an inflated belief that the opponent will swerve to avoid a crash. While both players seek a positive outcome, the wishful thinker acts on desire to avoid disaster, whereas the optimism-biased player makes a riskier choice based on a skewed belief of having a personal advantage. Thus, the behaviour of participants can be different within a game.

This paper focuses on wishful thinking in beliefs. Wishful thinking has yet to be investigated quantitatively. This is important for understanding how beliefs are formed, and whether this has an impact on belief learning. We aim to investigate whether wishful thinking should be considered in belief models, and to what extent it should be included in belief models.

Whilst biased beliefs don’t necessarily lead to "bad" decisions, these can be detrimental to policy makers. Wishful thinking has a tendency to push one towards believing outcomes that are favourable to you or simply fit your ideals. When making policy decisions, this can in fact lead to a bad decision. This combined with group decision making amplifying such biases through impacts such as the "in-group" (Paetzel and Sausgruber, 2018) can lead to individuals consistently making decisions that align with their wishful thoughts. Whilst this may not be troublesome for a while, there may come a time where wishful thinking causes the policy maker to make a detrimental decision. For example, for most policies there are pros and cons of implementing the policy. However, for some policies, there may be a negative net present value (NPV), and it would be difficult to justify implementation of the policy. If a group of policy makers is wishful about the policy, this could lead to them implementing a policy that would not be implemented in the absence of wishful thinking. It is hence important for researchers to understand if wishful thinking does occur in beliefs and then to further investigate steps that can be taken to reduce it, or factors that leave individuals at particular risk to wishful thinking.

Additionally, we see that groups can be particularly susceptible or tend to high performing group members, the so called "in-group" bias (Paetzel and Sausgruber, 2018). If wishful thinking leads to "better" performance for those that experience it compared to those who do not, the bias would be amplified in groups where high performing decision makers are present (Paetzel and Sausgruber, 2018) .

Biases are also amplified when beliefs are shared, so in a situation whereby decision makers discuss this with their colleagues when making a decision, we would expect the group to move in favour of a bias (Opera and Sausgruber 2022). If further individuals experience wishful thinking, this could lead to a vastly warped decision making process. Hence, it is important to understand whether or not this bias is experienced in beliefs, such that it can be further tested if it impacts groups. As policy decisions are made by groups, this could lead to policy choices being highly influenced by wishful thinking.

Belief learning literature has recently begun to investigate the mechanism behind belief learning, and how individuals adjust their beliefs over time. Whilst we acknowledge that wishful thinking is a bias, we also acknowledge that individuals may become aware of their potential bias or their previous mis-estimations of their opponents behaviour, and modify this. Hence, we will also observe the value of our wishful thinking parameter across two different time periods. We expect that, from the definition of wishful thinking, individuals will be consistent with their wishful thinking if they exhibit it. If this bias is heavily impacting an individual's ability to make decisions, and an alternative approach to prevent wishful thinking should be considered.

Consider an example of where these might differ when thinking about the weather. Wishful thinking would tell you that regardless of the weather forecast, it is going to be sunny today. Throughout the day, even when it becomes cloudy and rains, then the wishful individual would still believe that it would be sunny for the remainder of the day. At the beginning of the day, the optimistic individual would believe that it is going to be sunny, but given that it begins to cloud over and then rain, they would adjust their belief when presented with sufficient information. For this reason, optimism can be seen as more rational than wishful thinking.

It is important that we make progress towards quantifying biases or heuristics such as wishful thinking, such that we can establish if these effects directly impact an individual's ability to be rational. As wishful thinking is a concept that defines our beliefs, in this paper we modify a belief learning approach from Nyarko and Schotter (2002) to include wishful thinking. We provide analysis for two data sets, from Nyarko and Schotter (2002) and Attanasi et al. (2018), and find evidence to support the presence of wishful thinking. Additionally, we investigate how wishful thinking occurs with time, and find evidence of wishful thinking in short games, but with longer periods participants learn and reduce incidents of wishful thinking. We find that some participants are never wishful, and some participants continue to be wishful throughout the 60 period game.

3.2 Literature Review

There are three main subsections of the literature that are relevant for quantifying wishful thinking: Theoretical literature, belief learning literature and other bias literature. We discuss all of these in the following sections.

3.2.1 Wishful Thinking and Optimism

Individuals often confuse wishful thinking with optimism bias, and hence it is important to distinguish between the two to ensure our analysis is clear. Wishful thinking is an unrealistic belief that an outcome is obtainable. Optimism bias is a belief that is obtainable, but may be optimistic. This bias individuals are willing to learn from, and on reflection they will be able to evaluate given new information that this event is not likely to occur.

Whilst optimism bias was first introduced to the literature in the 1980s Weinstein (1980), wishful thinking has it's origins in psychology and a modification of cognitive dissonance theory by Bruner and Goodman (1947). Wishful thinking embodies an unrealistic optimist, whom believes that their desired outcome will prevail regardless of the information they receive on the way to the outcome. Optimism bias is hence the more rational version of wishful thinking. That said, wishful thinking itself need not be irrational. Heller and Winter (2020) show theoretically, that a strategy profile that exists that is consistent with wishful thinking and rational. This gives rise to the idea that wishful thinking does not need to be unrealistic, it can be focused on an outcome that can occur but still be unlikely. This rational wishful thinking forms the basis for our model, with participants being partially wishful.

Wishful thinking has been investigated in the context of sports Babad (1987), however our aim is to bring a more general modification that can be used when considering beliefs in the future. Additionally, the lack of financial incentives used by Babad (1987) leave questions as to whether the results of their experiment are simply due to the lack of motivation.

We aim to generalise a model of wishful thinking in beliefs, using the Nyarko and Schotter (2002) experiment whereby beliefs were incentivised. We do this by incorporating Bayesian belief updating. This allows for beliefs to be updated within each period, but with allowance for existing biases, for example wishful thinking, to prevail through beliefs.

3.2.2 Belief Learning

Wishful thinking has previously been theoretically modelled by Yildiz (2007). This model outlines strategy profiles that are consistent with wishful thinking. We give a brief outline of the model below.

> "wishful thinker if she hopes to enjoy the highest pay-off that is consistent with her information about the others' strategies." -Yildiz (2007).

In practice, the model states that wishful thinking occurs if a participant chooses their strategy and belief in a way that is consistent with the best possible outcome for themselves.

The model requires expected utility maximisation, and defines a wishful thinker as someone who expects to enjoy the highest possible payoff given their beliefs of other players strategies. For example, consider the following game.

	l	r
t	2,1	0,0
b	0,0	1,2

TABLE 3.1: Battle of the Sexes Game

Yildiz (2007) suggests that the strategies (t,l) and (b,r) are consistent with wishful thinking given that a player could expect to enjoy the highest payoff possible given information about other player's strategies.

Whilst this model does give guidance on how to identify wishful thinking consistent strategies, it offers no way to identify wishful thinking itself, other than an indication of if the individual is following the wishful thinking consistent strategy profile. Additionally, the presence of wishful thinking depends on the action of the opponent in this model, which is somewhat unrealistic. This makes it easy to identify wishful thinking strategies, but difficult to use for quantifying wishful thinking. Additionally, the characteristics of this model (i.e. requiring utility maximisation and obtaining the highest payoffs given beliefs), often are similar to the characteristics of nash equilibrium.

Existing belief learning literature focuses on participants learning from their beliefs across time, e.g. Ioannou and Romero (2014), Nyarko and Schotter (2002). Various data driven belief models have been formed, such as fictitious play and cournot beliefs, which are based on measuring beliefs based on unobservable data. We modify one existing approach developed by Nyarko and Schotter (2002), whereby the participants were asked about their beliefs using a "proper scoring rule". This model is as simplistic as possible, capturing the widely accepted cournout and fictitious play beliefs, whilst allowing for flexibility between the two models. However, this model does not allow for biases in stated beliefs. Whilst individuals may be aware of some biases, their stated beliefs will also incorporate their biases. Additionally, the Nyarko and Schotter (2002) belief learning model is the most proficient at predicting individual participants actions.

Other belief learning models have been studied extensively including in the context of oligopoly (Offerman et al., 2002) and repeated games (Ioannou and Romero, 2014).

We choose to use the Nyarko and Schotter (2002) model, as it is a general belief model for 2x2 games. Additionally, it performs well compared to traditional belief models, yet is not as complicated as other models such as Ioannou and Romero (2014). Additionally, the Nyarko and Schotter (2002) model is easy to adjust for wishful thinking, in that the individuals form their beliefs in a simplistic way. Whilst we acknowledge that other belief learning models exist, we aim to show that belief learning models should be adapted for wishful thinking, and our simple methodology modification can be applied to similar models.

3.2.3 Bias Literature

Similar biases that also impact beliefs such as overconfidence Hoelzl and Rustichini (2005) have also been evaluated. Hoelzl and Rustichini (2005) designed an experiment to investigate overconfidence, and find that individuals are overconfident in easy tasks. This overconfidence leads individuals to make a decision to participate in a task that relies on their innate ability. This choice of a different strategy in a game illustrates the possibility of a bias based on beliefs impacting the choices of individuals in any decision making process. Whilst overconfidence relies on individuals putting too much faith in their own ability, wishful thinking differs in that it puts too much faith in the universe.

Other biases, such as hindsight bias, have been investigated by psychology literature (Bernstein et al., 2011).

Our paper takes an innovative approach to belief formation, such that we can capture wishful thinking when individuals form their beliefs.

3.2.4 Experimental Literature

Existing literature has investigated the prevalence of wishful thinking in decision making in the context of farming and baking (Mayraz, 2011) and been expanded to investigate the role of wishful thinking in group decisions.

More recent literature (since 2023) has investigated the relationship between overconfidence and wishful thinking (Gneezy et al., 2023). (Engelmann et al., 2024) investigate wishful thinking to alleviate anxiety about adverse future outcomes, and find that wishful thinking disappears in the domain of monetary gains, indicating that negative emotions could be an important driver of wishful thinking consistent behaviour.

3.3 Hypotheses

In order to test whether wishful thinking exists, and if so what characteristics exhibits, we form two main hypotheses. The details of these follow.

Hypothesis 5 *Wishful Thinking in Beliefs Exists*

We aim to investigate whether wishful thinking is present in beliefs, and if so to what extent. We modify stated beliefs as follows:

$$P_t = \theta * P_{t-1} + (1 - \theta) \tag{3.1}$$

Where θ is a weight given to all previously available information and (1-θ) is the weight given to wishful thinking. This is a modified version of Bayesian updating of beliefs. For this hypothesis to hold, our estimation of θ must be statistically greater than 0. Additionally, to further verify that wishful thinking exists, we recreate the results from Nyarko and Schotter (2002), and compare our model to the results of this exercise. We compare using two main calculations - the Mean cross-subject mean squared deviation (MCSMSD) estimator from page 989 of Nyarko and Schotter (2002), and the αs our model and the original model by Nyarko and Schotter (2002) are nested, we will use a likelihood ratio test to determine which model "fits" the data better.

Wishful thinking is measured by our parameter θ and can be measured for each participant individually, to see variations in wishful thinking across participants. We hypothesise that allowing for wishful thinking in beliefs will better fit the Nyarko and Schotter (2002) data.

We further test this with data from Attanasi et al. (2018), as in a trust game, this belief modification has less alternative explanations, and so if our model performs better in this context we find the argument for wishful thinking to be more compelling.

We expect to find wishful thinking in the Nyarko and Schotter (2002) dataset, as there are profiles that are consistent with wishful thinking over time. If player 1 continues to play Green despite their player 2 partner playing red, then are being wishful.

However, pessimism in this game is not punished significantly, with the lowest payoff for player 1 being half that of the highest payoff, and for player 2 being 2/5 of the highest payoff.

In the Attanasi et al. (2018) data, we expect to find wishful thinking because player A deciding to continue the game is considered wishful. This action could encourage the wishful action of player B, that is share. This could create a self-fulfilling prediction of continued wishful thinking.

Whilst we expect to find evidence of wishful thinking, we do not expect all of the subjects to exhibit this behaviour. Optimism is a mindset that is not experienced by all, but it is proposed that there will be some individuals do in fact experience wishful thinking.

Hypothesis 6 *Wishful thinking Bias persists over time*

As we have many time periods, and two different data sets with different time periods, we can see how wishful thinking estimates are different if the time horizon is larger. The Nyarko and Schotter (2002) data set has 60 time periods, compared to the 4 time periods from Attanasi et al. (2018). If we experience true wishful thinking, we would expect wishful thinking to be present in both data sets.

If θ estimates are lower for the Nyarko and Schotter (2002) data set, this would indicate learning from wishful thinking, or adjusting beliefs over time, as the individual puts more weight on the previous beliefs than their wishful thinking. As the game is repeated, the longer the game goes on, the more likely that optimism bias is captured within the estimate of θ. This could be more indicative of optimism bias, and should follow in line with existing belief learning literature, that finds that individuals slowly adjust their beliefs, e.g. Erev and Roth (1998), Camerer and Ho (1997) and Nyarko and Schotter (2002).

Additionally, we compare the first 30 rounds of the Nyarko and Schotter (2002) data to the last 30 observations, in order to identify belief learning if the learning takes place across a longer time frame. If there is still evidence of wishful thinking after the initial periods, then what we are capturing is wishful thinking, and not simply optimism bias.

3.4 Data

We draw data from two existing studies, by Nyarko and Schotter (2002) and Attanasi et al. (2018). These two data sets combined help us to investigate the previous hypotheses.

3.4.1 Nyarko & Schotter (2002)

We choose to use the data from the original Nyarko and Schotter (2002) experiment, such that we can directly compare our model to the estimations of the original belief model.

We use the data from the experiment in Nyarko and Schotter (2002) to estimate the structural model. This involves an experiment that has 112 participants. The participants take part in a 2x2 game that is repeated 60 times, with the following payoffs:

	Player 2		
		Green	Red
Player 1	Green	6,2	3,5
	Red	3,5	5,3

This game was specifically chosen to have an equilibrium that was easily calculable or able to be learned through play, and was a mixed strategy equilibrium. However, the authors did not want equilibrium beliefs to be degenerate. These features were conveniently provided by a constant sum 2x2 game, and this game is as simple of a game where such equilibria are supported Nyarko and Schotter (2002).

The participants play this game over 60 rounds, and before each round, they are asked for their beliefs of the probability that their opponent will play green next round. This obviously also gives the belief of the probability their opponent plays red next turn as well.

These beliefs are incentivised, such that the participants are encouraged to provide their true beliefs. They are paid according to the following rule, on top of the outcome of their decisions:

$$\pi_i = 0.10 - 1/20[(1 - r_i)^2 + (r_j)^2] \tag{3.2}$$

where π_i represents the probability of playing strategy i, and r_i and r_j are binary variables for the opponent's decision in the period. This means that participants are rewarded for how close their prediction was to the actual action of their opponent, and incentivising individuals to reveal their best guess of their opponent's likely action for the next round.

We focus on the data whereby the individuals are asked for their beliefs of their opponent's strategy before each round of the game, as the model we are estimating is formed based on elicited beliefs. This leads us to exclude half of the original subject pool, as half of the treatments do not elicit beliefs prior to the game taking place. The reasons for this are twofold; because we do not wish to use beliefs elicited from the data, as they are constructed from a model- and this construction could lead to excessive noise. Identification of beliefs and wishful thinking parameters would lead to identification difficulties. Additionally, we cannot identify wishful thinking with static observations. We require at least 3 periods of data to identify our wishful thinking parameter and the existing model parameters simultaneously.
Nyarko and Schotter (2002) also separated the experiment into two treatments: fixed and random. In the fixed treatment, subjects faced the same individual for each of the 60 rounds. In the random treatment, subjects faced a different opponent in each of the rounds. We estimate the parameters for the matched treatment firstly, so that belief learning can occur, which enables us to use the Bayesian updating rule as outlined in the Methodology in Section 5. We also provide estimations for the unmatched treatment. It is expected that this will provide less informative estimates than the matched experiment.
These restrictions leave us with 56 subjects, and for each of these subjects we have 60 decisions, and 60 elicited beliefs.

Nyarko and Schotter (2002) use these tasks to develop logistic belief models, and compare this against the previous existing belief models; Cornout beliefs, fictitious play beliefs. Nyarko and Schotter (2002) propose a belief learning model, in the following form, based on the notation of Cheung and Friedman (1997), player i's γ-weighted empirical beliefs (or, for simplicity, empirical beliefs) are the sequence defined by:

$$b_{it+1}^{j} = \frac{1_t(a^j + (\Sigma_{u=1}^{t-1}\gamma_i^u 1_{t-u}(a^j)}{1 + (\Sigma)_{u=1}^{t-1}\gamma_i^u} \tag{3.3}$$

where b_{it+1}^{j} is player i's belief about the likelihood that the opponent will choose action a^j in period t+1, (a^j) is an indicator function equal to 1 if a^j was chosen in period t and 0 otherwise, and γ_i^u is the weight given to the observation of action a^j in period $t-u$. It should be noted that fictitious play beliefs are a special case of this function, where $\gamma = 1$ and Cournot beliefs are for which $\gamma = 0$.

3.4.2 Attanasi et al. (2018)

We choose to add an additional data set, from Attanasi et al. (2018)[1]. This data set was specifically chosen due to the repeated nature of the game, and the context the game was presented in.

A trust game requires that individuals are at the mercy of their partner in the game. This means that there is the opportunity for an individual to believe that their partner is trustworthy, even when there is no reason to believe this when the game is repeated. This is more suitable for our estimations, as the nature of the game gives the opportunity for true wishful thinking, and less alternative explanations for wishful thinking than a 2x2 game. In the trust game, any money given back is purely goodwill, and so belief of anything being returned is wishful thinking. Additionally, there are less alternative explanations for this game than the Nyarko and Schotter (2002) game. For example, it is not possible to use a mixed strategy with the trust game for one player, as they can simply "Dissolve" the game.

Attanasi et al. (2018) use a 2x2 game that has the following form:

	Take	Share
Dissolve	1,1	1,1
Continue	0,4	2,2

This game is played as a simultaneous-move game. The game was designed to have a "cooperative" action (continue), and allow for intention-based reciprocity

[1] We are thankful to Attanasi et al. (2018) for providing us with this data set.

when repeated.

The game was repeated 4 times, with the same opponent. Participants were first and second-year undergraduate students in Economics at Bocconi University of Milan, and earned an average of €8.86. There were 16 sessions, with 20 participants per session, 320 subjects in total. Beliefs are elicited before each round, which is done by asking participants to predict the probability of an opponent playing a strategy, which differs from the Nyarko and Schotter (2002) data in that these beliefs are taken as percentages, rather than probabilities. The percentages are in intervals of 10, i.e. 0%,10%, 20% and so on up to to 100%. We modify our methodology slightly for this, and convert the percentages to probabilities. This allows us to use the same methodology for both of the data sets.

3.5 Methodology

We start with the idea that wishful thinking occurs when beliefs are formed. This means that when an individual forms their beliefs, they combine all of the information they have in the current period to obtain an estimation of their opponents' behaviour in the next round.

There have been many different methods for belief formation offered by the literature, including the aforementioned Cournot and Fictitious Play beliefs. However, as individuals are asked for their beliefs, we believe that they will base this on information they have available to them. This leads us to a similar estimation method offered by Bayesian updating.

We adapt the existing Bayesian updating method to include the possibility of biases, in this case that of wishful thinking. This means that we have a convex combination of previous beliefs, and a bias. We use θ to denote the weight which the players give to the Bayesian updating and (1-θ) becomes the remainder, i.e. wishful thinking. As θ is a weight, we restrict the values to be between 0 and 1. The Bayesian updating rule becomes:

$$P_t = \theta P_{t-1} + (1 - \theta) \tag{3.4}$$

where $0 < \theta < 1$. If we view the highest value of expected payoff to be a "prize", we can interpret the bias that is denoted by θ as wishful thinking. This means $(1 - \theta)$ represents the amount of wishful thinking the individual does in this particular decision. If θ is 0, this means participants are completely ignoring previous data, and relying on the information they create through bias, which is their wishful thinking. If θ is 1, then the participant is putting all of their weighting on previous data, and not exhibiting wishful thinking.

Focusing on the data and belief model by outlined by Nyarko and Schotter (2002), we then use this estimation to obtain an estimate for the probability of a given player playing green, similarly to the Nyarko and Schotter (2002). This is formed by first calculating the expected utilities for each strategy based on the modified stated beliefs, for example:

$$EU_i = p_i S_i + (1 - p_i) S_j \tag{3.5}$$

where EU_i indicates the expected utility of strategy i, p_i indicates probability of strategy i (based on stated beliefs), S_i indicates strategy i, and S_j indicates strategy j.

Given the expected utility, this is used to calculated the probability of a player's opponent choosing a given strategy.

We must note that this differs from the previous probability estimation, as P_t represents an opponent's strategy, where as the following equation refers to own strategy:

$$\pi_{it} = \frac{exp(exp(\beta_1 * EU_i(G))}{exp(\beta_1 * EUi(R)) + exp(\beta_1 * EUi(G))} \tag{3.6}$$

Where $EU_i(G)$ and $EU_i(R)$ are the expected values of oneself playing the Green and Red strategies, which are calculated using the stated beliefs for the original (Nyarko and Schotter, 2002) paper, modified by equation (3.4) for our model. This differs slightly from the structural model used by Nyarko and Schotter (2002), in that we remove the parameter β_0. The original model used by Nyarko and Schotter (2002) is as follows:

$$\pi_{it} = \frac{exp(\beta_0 + exp(\beta_1 * EU_i(G))}{exp(\beta_0 + \beta_1 * EUi(R)) + exp(\beta_0 + \beta_1 * EUi(G))} \tag{3.7}$$

In the original model, β_0 represents the intercept of the regression model, however we choose to drop it. We do this because for some values of β_0 the estimates would behave strangely, and it would lead the estimates to violate monotonicity. For example, if we assume EU_A>EU_B, then a β_0 value of 0.5 gives a predicted probability of 0.52, when it should always be below 0.5. Even with constraining the parameter, this hits the upper bound.

It is worth noting here that while we drop β_0 from our estimations, we additionally conducted the analysis with β_0 included, to ensure that this does not impact the interpretation of our results. The results are not statistically different from those conducted with β_0 included.

We then use maximum likelihood estimation techniques to find λ and θ, using a logistic function. The function takes the following form:

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} \text{if previous action = green}, ln(\pi_{it}), \text{if previous action = red}, ln(1 - \pi_{it}) \tag{3.8}$$

that is, if the player previously plays strategy green, then their likelihood of choosing green is p, else it is $1 - p$.

We use the same methodology to estimate the Nyarko & Schotter (2002) and our modified belief model for the Attanasi et al. (2018) data set.

3.6 Results

The results of our estimations are in the following two sectors, separated by data set.

3.6.1 Results: Hypothesis I

Attanasi et al. (2018)

The results of our $(1-\theta)$ estimations are shown below. The average estimation is 0.651 (3 s.f), which indicates that individuals do experience wishful thinking. In total, just over half (163 of the 320) participants have a value of $(1-\theta)$ that is statistically greater than zero. Full results of the estimations are available in Appendix C. [2]

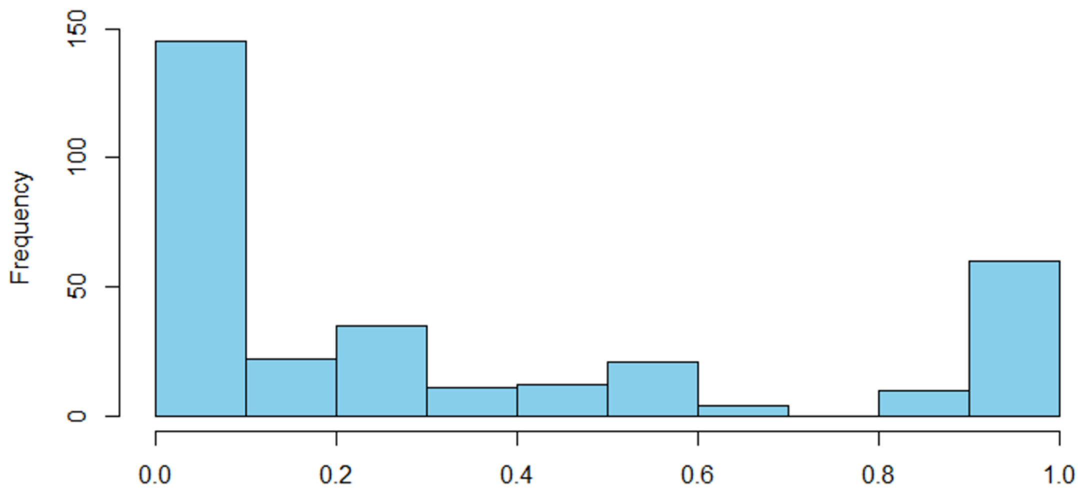


FIGURE 3.1: $(1-\theta)$ estimates for Attanasi et al. (2018) data
Note the x axis is indicative of the value of $(1-\theta)$

We see many estimates that are non zero, with approximately 30% of subjects having a wishful thinking parameter of over 0.5. This is strong evidence to suggest wishful thinking for these individuals, and compelling evidence for the inclusion of wishful thinking in belief learning or belief formation models. Whilst we acknowledge that there are some participants that have low estimates for the $(1-\theta)$ estimates, with 45.3% of participants having an estimate of less than 0.1, the estimation defaults to the Nyarko and Schotter (2002) model when the value of $(1-\theta)$ is close

[2] It should be noted that results in Appendix C are rounded, and whilst it appears that results hit the bounds, there are no results which are equal to the bounds.

to zero. This means that there is no detriment to including the wishful thinking parameter. Additionally, there are 95 subjects that wishful thinking does impact, for which including the wishful thinking parameter allows for better estimation of the individual's belief[3]. Whilst this is not the case for every participant, wishful thinking is something that is experienced on an individual level.

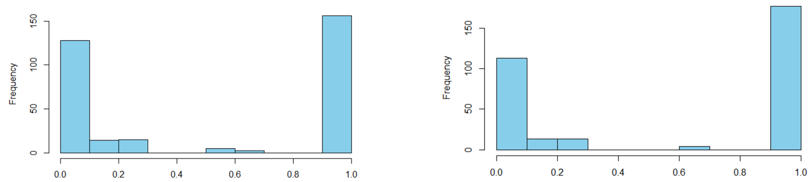
We also note that the game itself does give rise to differences in wishful thinking and parameters. As the game is a binary choice and hence there are only four outcomes, and the decision is repeated only four times, there are limited estimated parameters that can come from this game.

The trust game design leads to less alternative explanations for wishful thinking. This is because any behaviour that relies on trust, hinges on the belief that one's opponent will be kind in their response or behaviour. In this game, if one acts based on wishful thinking and is rewarded, in future rounds the best response of their partner could be wishful thinking, and then further wishful thinking will be encouraged. This could lead to a self-fulfilling prophecy of wishful thinking. These differences could be attributable to the differences between the games. The Nyarko and Schotter (2002) 2x2 game has two wishful thinking consistent strategies; Green for player 1 and Red for player 2. In this game, pessimism is punished comparatively more than optimism. Hence we expect to see more wishful thinking consistent behaviour in this game.

The distribution of wishful thinking $(1-\theta)$ estimates is approximately bimodal, with peaks at 0 and 1. We see some participants that do not exhibit wishful thinking consistent behaviour, and some participants that exhibit wishful thinking consistent behaviour. The data indicates that participants are either wishful thinking, or do not exhibit wishful thinking at all.

Whilst we acknowledge that the previous estimations using the Nyarko and Schotter (2002) original model work well for some participants in this sample, we argue

[3]This will be discussed further later, in the Robustness Section



(A) Nyarko & Schotter (2002) (B) Wishful Thinking

FIGURE 3.2: β Estimates for Attanasi et al. (2018) data
Note the x axis is indicative of the value of β

that some are better served by the addition of our wishful thinking parameter, as indicated by the results of the likelihood ratio test (reported in tables 5, 6 and 7 in Appendix C.) At the 1% level, 76 of the 320 participants reject the null hypothesis, suggesting that the addition of the wishful thinking parameter provides a sufficient improvement in fit to warrant the inclusion of an additional parameter.

We turn to the estimations of β and γ, to indicate if there are large differences in the other estimations.

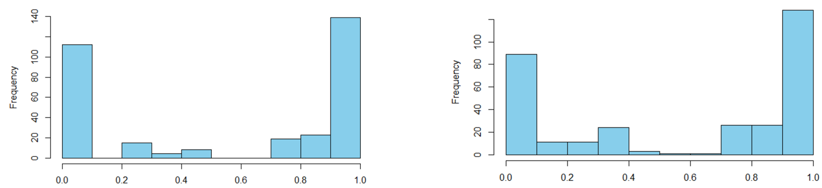
The table below illustrates the mean and standard deviation of the γ and β parameters for the Attanasi et al. (2018) data set. As seen in the table, the estimations for β are similar, a mean of 0.52 compared to 0.58, and there are few subjects for which the estimated parameter changes. This indicates that that the addition of the θ parameter does not change the estimations of the existing parameter β by much.

TABLE 3.2: Mean and Standard Deviation of Gamma and Beta Estimates for Attanasi et al. 2018 Data

		Nyarko & Schotter (2002) Model	Wishful Thinking Model
Gamma	Mean	0.56	0.52
	Standard Deviation	0.42	0.42
Beta	Mean	0.52	0.58
	Standard Deviation	0.44	0.48

TABLE 3.3: MCSMSD Estimations

The estimations for the γ parameter are shown below.



(A) Nyarko & Schotter (2002) (B) Wishful Thinking

FIGURE 3.3: γ estimates for both datasets

Note the x axis is indicative of the value of γ

We see that our estimations are on average lower for the Attanasi et al. (2018) data set. We also see, similar to Attanasi et al. (2018), that many of the participants follow $fictitious\ play$ closely, as the estimation for gamma is γ approximately 1. We note that while there are small changes in the estimated γ parameter, there is a similar pattern to the estimations of Nyarko and Schotter (2002).

Additionally, the inclusion of $1 - \theta$ parameter could lead to a better fit of belief model. We investigate this in the robustness section.

The following two sections present the results of the estimations for the Nyarko and Schotter (2002) data, firstly for the matched elicitation treatment, and then for the randomly matched treatment.

Nyarko & Schotter (2002) matched elicitation data

We next turn to the wishful thinking parameter estimations for the Nyarko and Schotter (2002) matched elicitation data. The results are indicated below, with the full results reported in Appendix C.[4]

[4]It should be noted that results in Appendix C are rounded, and whilst it appears that results hit the bounds, there are no results which are equal to the bounds.

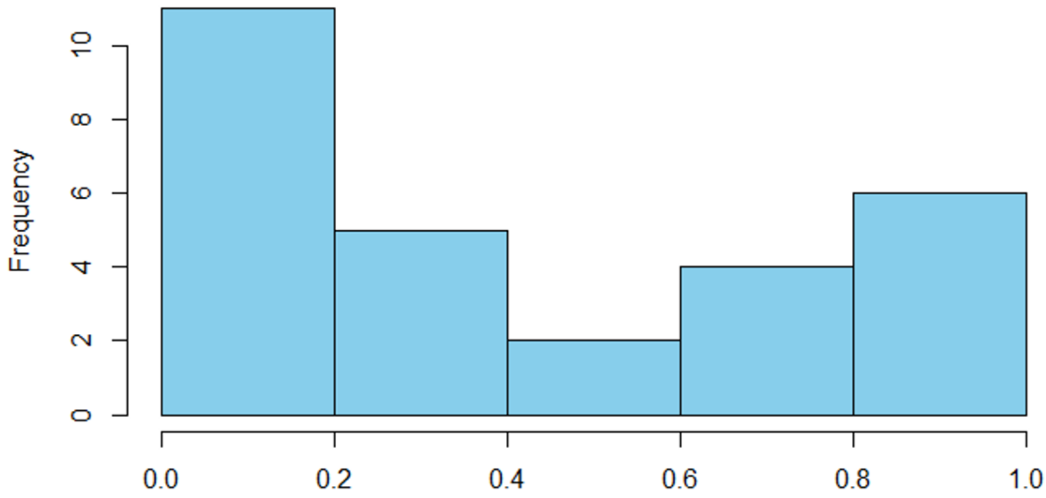
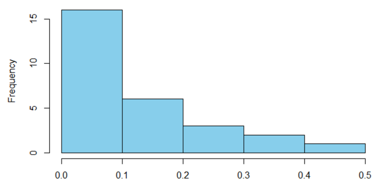


FIGURE 3.4: $(1-\theta)$ Estimates for Nyarko and Schotter (2002) Data

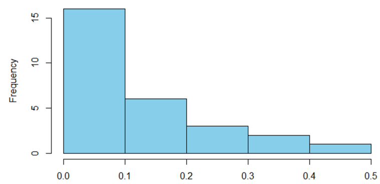
Note the x axis is indicative of the value of $(1-\theta)$

We see an average estimation of 0.412, and a standard deviation of 0.37(3 s.f), which is considerably lower than the estimations for the Attanasi et al. (2018) sample. We see that the estimations for our wishful thinking parameter are more frequent than the Attanasi et al. (2018) data results. The distribution of wishful thinking parameter estimates is once again bimodal, with more participants having an estimated wishful thinking parameter close to zero or one. Eleven participants show almost no indication of wishful thinking consistent behaviour and beliefs. Six participants show behaviour and beliefs consistent with wishful thinking.

There are a few plausible explanations for these results. This game does not have a pure strategy Nash equilibrium, and hence the incidence of wishful thinking as theorised by Yildiz (2007) will be reduced. Secondly, the game design is such that player 1 wants to co-ordinate, whilst player 2 wants to be subversive to co-ordination. The final explanation is that some individuals simply do not exhibit wishful thinking consistent behaviour or beliefs. We see similar estimations in the Attanasi et al. (2018) data set with 46% of participants showing no indications of wishful thinking consistent behaviour or beliefs; for this reason, our preferred explanation is that some individuals exhibit no wishful thinking.



(A) Nyarko & Schotter (2002)



(B) Wishful Thinking

FIGURE 3.5: β Estimates for Nyarko and Schotter (2002) Data
Note the x axis is indicative of the value of β

There are six participants who show evidence of wishful thinking consistent behaviour that is statistically different from zero.

Furthermore, there is also evidence of individuals not exhibiting wishful thinking in both the Attanasi et al. (2018) and Nyarko and Schotter (2002) data sets. Whilst at first this may seem to contradict our theory, we posit that although individuals can be wishful in their thinking, there is also the possibility that some people are simply not wishful. For example, an individual whom is highly pessimistic is hardly likely to exhibit wishful thinking in this type of game. Our intentions were not to show that all the participants were wishful, but that some participants exhibit behaviour that is consistent with this parameter.

We next turn to the estimations for β and γ. The estimations for β are shown below.

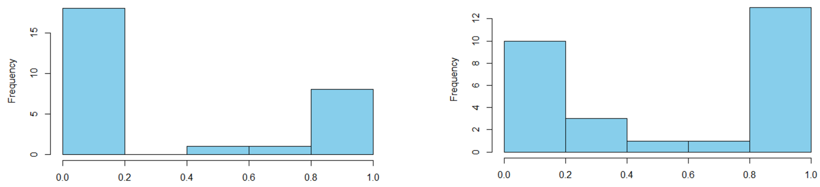
We see remarkably similar estimates for both β estimations for the original Nyarko and Schotter (2002) model and our modification. This mirrors our findings from the Attanasi et al. (2018) data set, whereby the inclusion of θ has little impact on β estimations. The table below illustrates the mean and standard deviation estimates for the Nyarko and Schotter (2002) data set.

Next, we turn to the estimations of γ.

We see that the estimates for γ are slightly different, with higher estimates for our estimations. Once again, this is similar to the Attanasi et al. (2018) results. We see

		Nyarko & Schotter (2002) Model	Wishful Thinking Model
Gamma	Mean	0.34	0.51
	Standard Deviation	0.43	0.43
Beta	Mean	0.11	0.13
	Standard Deviation	0.13	0.43

TABLE 3.4: MCSMSD Estimations



(A) Nyarko & Schotter (2002) (B) Wishful Thinking

FIGURE 3.6: γ Estimates for Nyarko and Schotter (2002) Data
Note the x axis is indicative of the value of γ

changes in the potential belief model that we use.

Nyarko & Schotter (2002) randomly matched elicitation data

The estimates for $(1-\theta)$ are illustrated below.

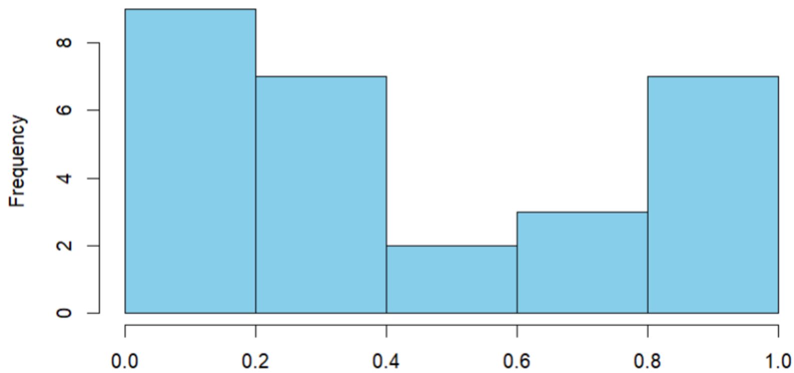


FIGURE 3.7: $(1-\theta)$ Estimates for Nyarko and Schotter (2002) Data

Note the x axis is indicative of the value of $(1-\theta)$

The Nyarko and Schotter (2002) results for γ are illustrated below.

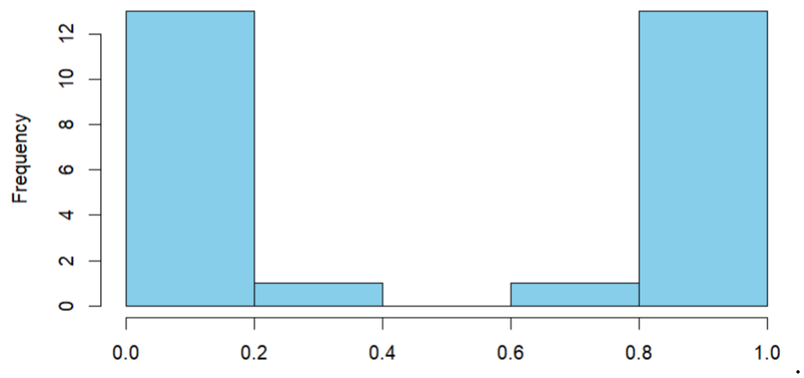


FIGURE 3.8: γ Estimates for Nyarko and Schotter (2002) Data

Note the x axis is indicative of the value of γ

The results are approximately similar to the results in the matched elicitation experiment. Turning to β, the results are similar to the matched elicitation, with a higher frequency of β values of 1. This indicates that participants in the random matched treatment prefer the Green strategy.

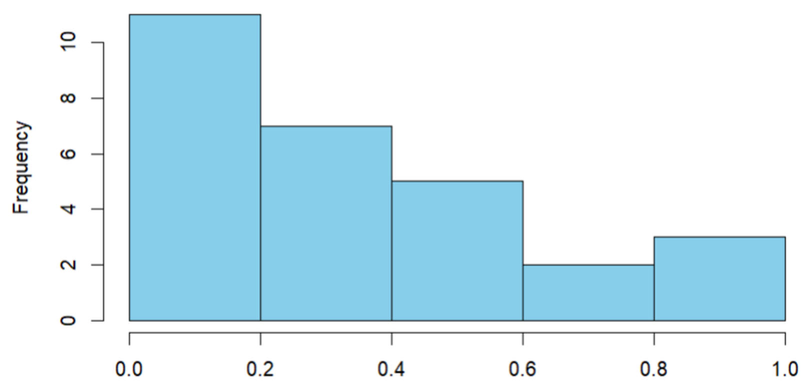


FIGURE 3.9: β Estimates for Nyarko and Schotter (2002) Data

Note the x axis is indicative of the value of β

3.6.2 Results: Hypotheses II

The estimations from the Attanasi et al. (2018) data set indicate that some individuals do exhibit wishful thinking in the four period game. Additionally, we see evidence of wishful thinking in the Nyarko and Schotter (2002) data set. This provides compelling evidence to suggest that we have captured the wishful thinking for some subjects.

This estimation of wishful thinking is seen at different levels for different subjects. Whilst we see in Figure I that there are only a few possible values of wishful thinking, due to the limited choices in the Attanasi et al. (2018) data set. When compared to Nyarko and Schotter (2002), whereby subjects had 60 choices, we see more accurate estimations across the distribution potential wishful thinking parameters. This shows that our estimations allow for participants to change their level of wishful thinking throughout the evolution of the game, for example if a participant had high wishful thinking at the beginning of the game, but reduced it throughout, the participant would more likely show between 0.2 and 0.8 rather than the binary nature of wishful thinking indicated by the Attanasi et al. (2018) data. This is also evidence that there may be some optimism bias in initial rounds. Whilst the weighting of prior knowledge will capture some optimism bias after the first round, it could be the case that individuals are slow to learn from the prior information. The longer rounds go on, the more that optimism bias is removed as a viable alternative explanation.

To investigate this theory, we reduce the number of rounds in the Nyarko and Schotter (2002) data set, by restricting the rounds to round one to four. [5] As seen in Figure

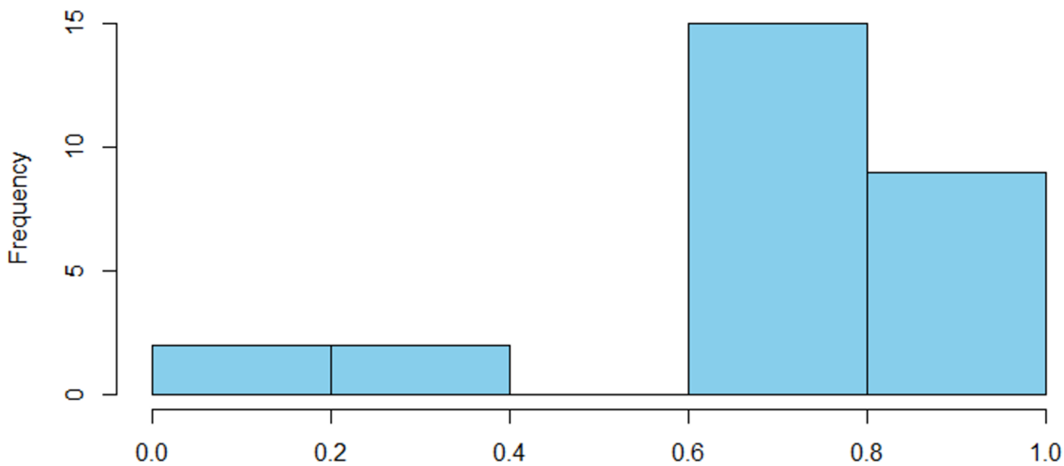


FIGURE 3.10: $(1-\theta)$ Estimates for Nyarko and Schotter (2002) Data, restricted to first four periods
Note the x axis is indicative of the value of $(1-\theta)$

[5]We note that estimations for γ and β are broadly similar for the reduced number of periods as previously seen in both data sets.

7, the estimates for wishful thinking are considerably higher than those for the unrestricted sample. This indicates that wishful thinking is present for more participants early in the 60 period game that is presented in Nyarko and Schotter (2002). This confirms the potential for optimism bias to be present in games with fewer periods. This leads us to believe that the driver of the differences in apparent wishful thinking estimates is optimism bias, rather than the difference in the type of game between the Attanasi et al. (2018) and Nyarko and Schotter (2002) data set. To further investigate this, we run an estimation at 30 periods using our model and the Nyarko and Schotter (2002) model. We run two seperate estimations; one for the first 30 periods and a second for the last 30 periods. The Likelihood ratio test between them suggests that the model fits the last 30 data

When we restrict to the first 30 periods, we see that there is some evidence to support wishful thinking, with some participants exhibiting no wishful thinking behaviour. This has a noticable decrease from the first four periods, indicating that the wishful thinking model is picking up some of the optimism bias present. By 30 periods, lots of this optimism has been eliminated.

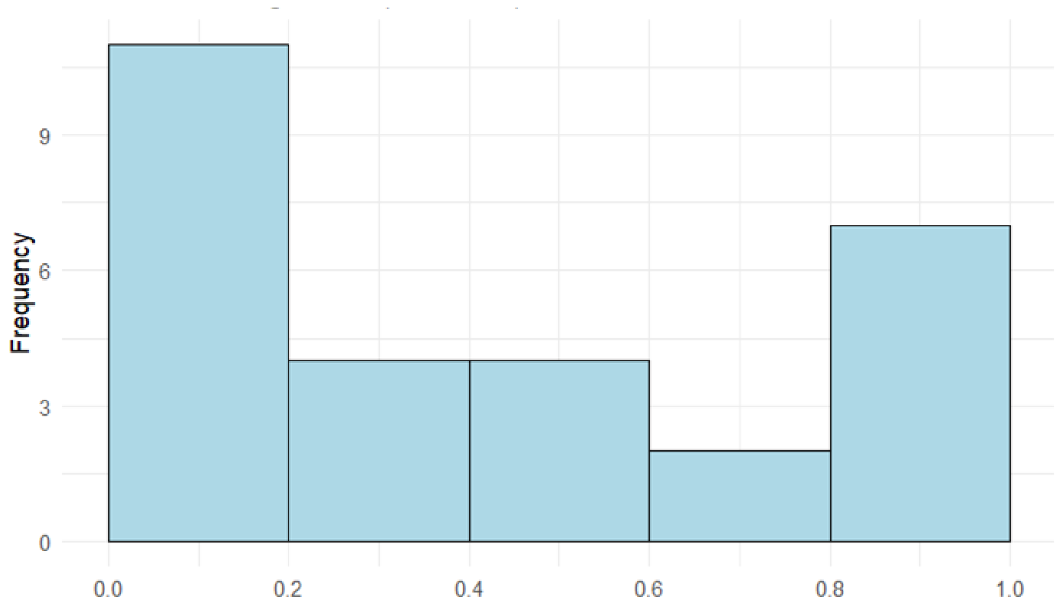


FIGURE 3.11: $(1-\theta)$ estimates for Nyarko and Schotter (2002) Data, restricted to the first 30 periods
Note the x axis is indicative of the value of $(1-\theta)$

When we turn to restricting to the last 30 periods, there are some noticable differences. The results for $(1-\theta)$ estimation are summarised in the figure below. This

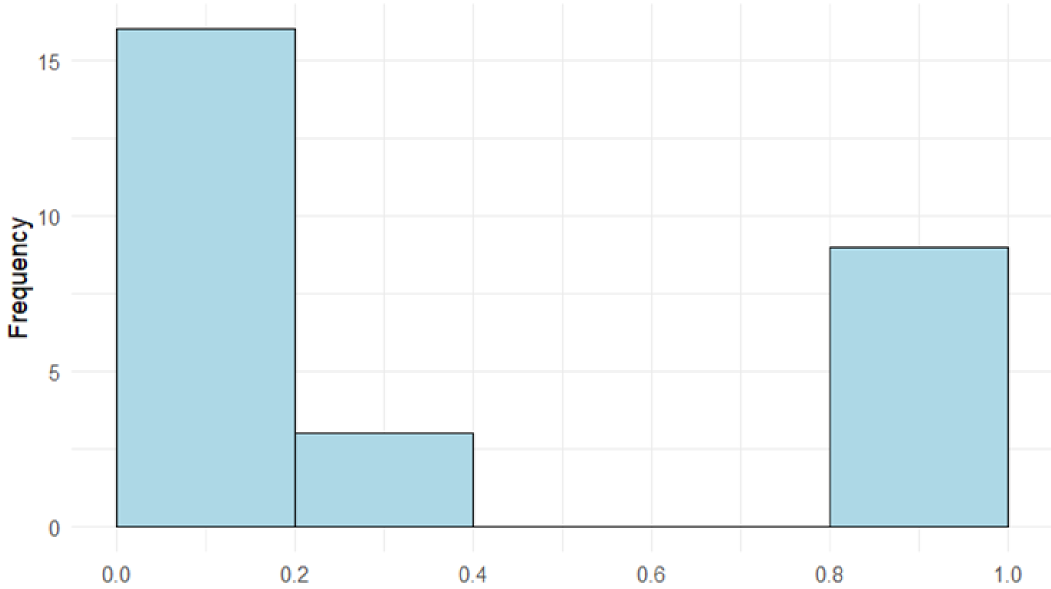


FIGURE 3.12: $(1-\theta)$ estimates for Nyarko and Schotter (2002) Data, restricted to the last 30 periods
Note the x axis is indicative of the value of $(1-\theta)$

figure indicates differences in wishful thinking compared to the 60 period estimates; with some individuals remaining with a high value of $(1-\theta)$ and hence showing evidence of wishful thinking, and others not showing any wishful thinking at all. This indicates that some of the participants are not wishful, but are simply optimistic; and their optimism disappears in the face of the historical bad outcomes. However, some participants remain wishful regardless of prior outcomes.

3.7 Robustness

Our model differs from that used by Nyarko and Schotter (2002) and Attanasi et al. (2018)- we include a parameter for bias and exclude the constant β_0. We therefore use two measures to evaluate how our model performs relative to the model used originally by Nyarko and Schotter (2002). We will explain each of these methods, and then evaluate the meaning behind these after.

We use the so called Mean Cross-Subject Mean Squared Deviation (MCSMSD) estimator from page 989 of Nyarko and Schotter (2002). This estimator represents how far away the predictions of the model are from the actual behaviour. Essentially, we compare the predictions of the probability of a player playing the strategy against the strategy they played. This is then summed up and averaged across the time period. We use the maximum likelihood estimates to compute fitted values for the probability, and then we then calculate the statistic by using the following formula:

$$MCSMSD_t = 1/k \sum_{i=1}^{K} (p_t^i - a_t^i)^2 \tag{3.9}$$

where k is the number of observations, p_t^i represents the probability of playing a strategy predicted by our model, and a_i^t is the binary action of the player, 1 for if the player plays the given strategy, and 0 otherwise. This is able to be calculated for both data sets, and gives an estimation of how closely the model predicts the actual behaviour of the participant.

As our model and the original model by Nyarko & Schotter (2002) are nested, we will use a likelihood ratio test to determine which model "fits" the data better. The test statistic is calculated as follows:

$$LRT = -2 * log_e(\frac{L_s\hat{\theta}}{L_g\hat{\theta}}) \tag{3.10}$$

where L_s is the model with less parameters, and L_g the model with more parameters included. We estimate these at a subject level, and then evaluate this compared to the critical statistic.

3.7.1 Robustness Results

The MCSMSD estimations are shown in the table below:

	Nyarko Estimate	Wishful Thinking Model
Nyarko & Schotter (2002) (fixed)	3.281469	3.283496
Nyarko & Schotter (2002) (random matched)	3.341705	3.343617
Attanasi et al. 2018	1.653929	0.7683095

TABLE 3.5: MCSMSD Estimations

This indicates that the wishful thinking model performs significantly better than the Nyarko and Schotter (2002) model for the Attanasi et al. (2018) data set, with the inclusion of $(1-\theta)$ increasing the fit (0.768 vs. 1.653).

This supports Hypothesis I in Section 3.3, indicating that on average, our model fits the data better than the Nyarko and Schotter (2002) model. This is likely due to the fact that 95 subjects exhibit wishful thinking throughout the game, their beliefs would be better outlined by our model, as opposed to the Nyarko and Schotter (2002) model.

Turning to the Nyarko and Schotter (2002) data set, we see that the original Nyarko and Schotter (2002) model outperforms the inclusion of $(1-\theta)$ by a small amount for both the fixed matched and random matched elicitation data (3.283 vs 3.281). There are a few potential explanations for these findings. The parameter estimation may be context specific, and as such in a trust game, where there is more opportunity for wishful consistent strategies (as defined by Yildiz, 2007), it is more likely that wishful thinking will be present and therefore detected. Alternatively, the effect of observed within the Attanasi et al. (2018) data could be initially optimism bias, which could be reduced over time, and hence we are seeing that individuals are learning from this over time, and adjusting their belief. However, this is contraindicated by the results of the separated estimations in the results section; with the presence of wishful thinking being found in the last 30 periods. What is clear is that the Nyarko and Schotter (2002) data set spans 60 periods, whereas the Attanasi et al. (2018) data set spans 4 repeated periods. This could lead to our model picking up optimism bias in addition to wishful thinking. Hence, the difference between these results could be due to the large difference in the number of repetitions in the game. On the other hand, the individuals in the Nyarko and Schotter (2002) experiment could simply not be exhibiting wishful thinking, which could limit the applicability of the model to all participants.

Another explanation for why the individuals for the Nyarko and Schotter (2002) experiment do not appear to exhibit wishful thinking could be cognitive ability

(Oechssler et al., 2009). If those who participated in the Nyarko and Schotter (2002) experiment have higher cognitive ability, then we would see lower levels of bias in this sample. Unfortunately, we are unable to test for this given the data that is available for both data sets. We do note that the Nyarko and Schotter (2002) participants were students, from Economics courses at the University of New york, whereas the Attanasi et al. (2018) participants were Economics students from the Bocconi University of Milan.

Secondly, the participants could simply not be wishful. Wishful thinking is widely understood as an unconscious belief phenomenon, and some individuals could be pessimistic, or simply not exhibit wishful beliefs. In this case, and due to the increase in fit given by the Nyarko and Schotter (2002) model in this scenario, we suggest thinking of wishful thinking somewhat similar to fitting a preference function, in that each individual has a different preference function, each individual's belief model should be considered for wishful thinking.

Finally, the differences in the context of the game should be considered. A trust game exhibits more opportunities for wishful thinking, in that the participant is able to believe that their partner will share more than they would otherwise. In the context of a trust game, there may be more anxiety about a future outcome than a simple 2x2 game. This may cause participants to engage in more wishful thinking behaviour, in line with Engelmann et al. (2024).

Further investigations into these causes is beyond the scope of this paper.

Next, we turn to the likelihood ratio test. The results of the likelihood ratio test are shown in Figure 7 in Appendix C.

These results show that there is not a significant difference between our model and the Nyarko and Schotter (2002) model. This suggests that the addition of the wishful thinking parameter does not make a significant improvement to the model. This result holds even with the randomly matched elicitation data when compared to the fixed match data. This suggests that individuals do learn over time to adjust their wishful thinking behaviour. This could be an area for future research to investigate further.

3.8 Conclusion

Whilst belief formation is continuing to be investigated, we posit that wishful thinking should be considered when modelling beliefs and belief learning. It may be the case that wishful thinking does not impact some decisions, but it should be considered. This is evident by the "U shape" found when estimating our model, with some individuals displaying wishful thinking consistent behaviour, with others displaying no such evidence. Many belief models, such as Nyarko and Schotter (2002), and Ioannou and Romero (2014) do not include the possibility of wishful thinking, and this leaves them susceptible to not only failing to capture the bias, but not fitting the data to the best of the model's ability. We see this when we estimate the four period Attanasi et al. (2018) data set, whereby a significant number of the participants experience wishful thinking. This highlights that is particularly important to consider the context of the game when deciding if wishful thinking should be modeled. For example, in a trust game, the desire for a higher outcome is much more likely to encourage players to place trust in their fellow participants than in a simple 2x2 game. This makes the presence of wishful thinking likely to be context specific.

The modification of belief learning models can be as simple as we have illustrated, a simple modification of the model to allow for the possibility of belief learning. This should be conducted in belief learning models where wishful thinking could be present, although we recognise that in other contexts, whereby luck or trust play a little part, modifying for wishful thinking may result in estimations of null or close to null of $(1-\theta)$. This would indicate that there is no wishful thinking for this participant. We suggest an approach whereby wishful thinking is considered, but not mandatory, for each participant.

We suggest an approach similar to preference function fitting, where the best belief model for the individual participant is fitted. This allows for a better estimation of the individual's beliefs. The methods outlined in this paper also allow for further investigations for wishful thinking. As such, Heller and Winter (2020) has begun to investigate that wishful thinking can be rational, and the next step is to determine if wishful thinking is problematic for decision makers. Whilst this paper discusses the existence of wishful thinking, we make no attempt discuss the impact of exhibiting wishful thinking on the outcome of the decision. However, the impact of wishful thinking on policy decisions could be large given that policy decisions are often one-shot games, and we see that wishful think is more prominently defined in data with less repeated games.

Whilst these data sets give us a good base to form our estimations of wishful thinking, they were not designed for such calculations. Whilst the Nyarko and Schotter (2002) data set has many observations, there are a number of explanations that could explain the evidence of wishful thinking. For example, choosing randomly or mixed strategies cannot be excluded. Additionally, the elicitation method of beliefs encourages a sort of hedge against holding beliefs of either 0 or 1. This is because the beliefs are elicited using a payoff system to try and identify true beliefs, but it does leave room for individuals to put 0.9 instead of 1 for a belief, due to the possibility of them being wrong. This would lead to beliefs converging towards the middle, and less "extreme" belief values. This therefore would have an impact on our θ values as well, as these are constructed from beliefs.

On the other hand, whilst the data from Attanasi et al. (2018) has the benefit of being a trust game, it does not have as many time periods as the Nyarko and Schotter (2002) data set includes. This means that our estimates are somewhat noisier than those with the Nyarko and Schotter (2002) data set. However, we still find evidence of wishful thinking.

Ideally, we would have a data set from an experiment designed to elicit wishful thinking, for example by directly asking what individuals hope will happen, and what they think will happen, and rating their confidence in their desired outcome. This repeated multiple times with the same candidate would create the best possible data set for our estimations, allowing for a specific estimation of wishful thinking. This approach would allow for studying different scenarios where wishful thinking is greatest. This would further allow for differentiation from optimism bias within the estimations. Alternatively, drawing from specific games, for example those mentioned in Yildiz (2007), could be key to identifying when wishful thinking is occurring in regular games. These games could lead to a further understanding of wishful thinking, and the factors that could influence wishful thinking.

Ultimately, we find evidence in favour of the existence of wishful thinking in the Attanasi et al. (2018) data set, and limited evidence for wishful thinking in the Nyarko and Schotter (2002) set. We find that individuals do behave in a way consistent with wishful thinking, which further backs up the theories of Yildiz (2007). These impacts are particularly prominant in trust games, and are still present when there are a large number of periods. Additionally, our paper is the first to address quantifying wishful thinking in repeated games.

Appendix A

Appendix for Chapter 1

A.1 Experiment Instructions

The following instructions were provided to participants at each stage. Group A and Group B are shown different instructions for partial information treatment and the full information treatment, but the same instructions for the own treatment.

The following instructions were given to Group A in the partial information treatment:

Welcome to the experiment.

Instructions

During this experiment, you will be allocated to either group A or group B. You are in group A, and you have been allocated a random partner from group B.

You will remain in Group A throughout this experiment.

You will participate in 30 decision periods. In each period you will be asked to make a decision. Some of the decisions depend on q. q is a number randomly generated by the computer between 0 and 1. At the end of an experiment, one of the periods will be randomly selected by the computer to be the period that you are paid for. Since any period could be chosen, please choose carefully in each period. Regardless of the outcome of your choices, providing you finish this experiment, you will receive an experiment completion fee of £3. The outcome of the chosen period can yield between 0 and 100 tokens. These tokens will be converted to pounds with the following exchange rate:

$$Value = (number of tokens)/10. \tag{A.1}$$

For the experiment you will receive the £ Value of your tokens. For example, if you receive 75 tokens your £ Value would be £7.50. The £ Value is then added to the show-up fee of £3, so your total payment would be £7.50+ £3 = £10.50.

This experiment will be in 3 stages. The instructions for Stage 2 and Stage 3 will follow after Stage 1 is complete.

Stage 1: This Stage will consist of 10 periods, after which the instructions for Stage 2 will appear. The outcome of these options depends on q, a number randomly generated by the computer between 0 and 1.

You will be asked to choose between two options, option A and option B.

In this Stage, the decision will appear on the screen as follows[1]: Please Choose.

A

Participant B will receive 26 tokens if q is less than or equal to 0.5

Participant B will receive 5 tokens if q is greater than 0.5

or

B

Participant B will receive 5 tokens if q is greater than 0.5

[1] Please note that in the experiment, these were displayed side by side

You will be asked to choose either A or B. Below the choice, you will be able to see what your randomly matched partner in group B will see. This will be differ from what you see, for example:

Your randomly matched partner in group B will see your choice, and the information presented below:

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

or

B

25 tokens

Your partner from group B will wait on you to make a decision.

After you have chosen between A and B, your partner in group B will see the information that was displayed to you, alongside your decision and their outcome for the period. An example is included below: Your matched player A chose: A.

Your payoff for this round is: 26 tokens.

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

or

B

25 tokens

Player B will then decide how many tokens from a pot of 50 tokens to allocate to you and a random participant. This does not impact Player B's payoff for the experiment, but is a separate pot of 50 tokens to allocate between you and a random other participant. For example: Please choose an amount between 0 and 50 tokens to allocate to your randomly matched player A. The amount you choose will be allocated to your matched player A.

Once this decision has been made, the round is over.

Please ensure you click "Next" once you are finished making your decision.

Please contact the experimenter if you have any questions or experience a problem during the experiment.

The following instructions were given to Group B in the partial information treatment:

Instructions: Stage 1 Group B

Welcome to the experiment.

Instructions

During this experiment, you will be allocated to either group A or group B. You are in group B, and you have been allocated a random partner from group A.

You will remain as Group B throughout this experiment.

You will participate in 30 decision periods. In each period you will be asked to make a decision.

Some of the decisions depend on q. q is a number randomly generated by the computer between 0 and 1.

At the end of an experiment, one of the periods will be randomly selected by the computer to be the period that you are paid for. Since any period could be chosen, please choose carefully in each period. Regardless of the outcome of your choices, providing you finish this experiment, you will receive an experiment completion fee of £3. The outcome of the chosen period can yield between 0 and 100 tokens. These tokens will be converted to pounds with the following exchange rate:

$$Value = (number of tokens)/10. \tag{A.2}$$

For the experiment you will receive the £ Value of your tokens. For example, if you receive 75 tokens your £ Value would be £7.50. The £ Value is then added to the show-up fee of £3, so your total payment would be £7.50+ £3 = £10.50.

This experiment will be in 3 stages. The instructions for Stage 2 and Stage 3 will follow after Stage 1 is complete.

Stage 1:

This stage will consist of 10 periods, after which the instructions for stage 2 will appear.

The outcome of these options depends on q, a number randomly generated by the computer between 0 and 1.

Your partner in group A will be asked to choose between two options, option A and option B. An example is shown below:

In this stage, the decision will appear on the screen as follows:

Please Choose.

A

Participant B will receive 26 tokens if q is less than or equal to 0.5

Participant B will receive 5 tokens if q is greater than 0.5

or

B

Participant B will receive 5 tokens if q is greater than 0.5

Player A will be asked to choose either A or B.

Below the choice, players in group A will be able to see what their randomly matched partner in group B will see. This will be different from what group A sees, for example:

Your randomly matched partner in group B will see your choice, and the information presented below:

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

or

B

25 tokens

At this stage your page will read "Waiting for the other participant."

After Player A chooses between A and B, you will see the information displayed to player A, alongside player A's decision and their outcome for the period. An example is included below:

Your matched player A chose: A.

Your payoff for this round is: 26 tokens.

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

or

B

25 tokens

You will then decide how many tokens from a pot of 50 tokens to allocate to your matched player A and a random participant. This does not impact your payoff for the experiment, but is a separate pot of 50 tokens to allocate between your matched player A and a random other participant. For example:

Please choose an amount between 0 and 50 tokens to allocate to your randomly matched player A. The amount you choose will be allocated to your matched player

A.

Once this decision has been made, the round is over.

Please ensure you click "Next" once you are finished making your decision.

Please contact the experimenter if you have any questions or experience a problem during the experiment.

Please provide your student ID number and university email below. These are required for payment.

The instructions for Stage 2 for Group A are indicated below:

Instructions: Stage 2 Group A

Stage 2

Instructions

This stage will consist of 10 periods.

In this stage, your decision will appear on the screen as follows:

Please Choose.

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

B

25 tokens

You will be asked to choose either A or B on behalf of your randomly matched participant B.

After you choose between A and B, your randomly matched partner in group B will see the same information that was displayed to you, alongside your decision and the outcome for the period. An example is included below:

Your matched player A chose: A.

Your payoff for this round is: 26 tokens.

Please Choose.

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

B

25 tokens

Player B will then decide how many tokens from a pot of 50 tokens to allocate to you and a random participant. This does not impact Player B's payoff for the experiment, but is a separate pot of 50 tokens to allocate between you and a random other participant. For example:

Please choose an amount between 0 and 50 tokens to allocate to your randomly matched player A. The amount you choose will be allocated to your matched player A.

During this time, player A will see a screen reading "Waiting for the other participant."

Once this decision has been made, the round is over.

Players will then be matched to another random participant in the opposite group, and this process will be repeated until all ten rounds are complete.

To ensure the experiment runs smoothly, please ensure you click "Next" when finished making a decision.

Please contact the experimenter if you have any questions or experience a problem during the experiment.

Once you have read the instructions above and are satisfied, please click "Next" below.

The following text is displayed to Group B for The full information treatment

Instructions: Stage 2 Group B

Stage 2

Instructions

This stage will consist of 10 periods.

In this stage, the decision for participant A will appear on the screen as follows:

Please Choose.

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

B

25 tokens

Player A will be asked to choose either A or B on your behalf.

At this stage your page will read "Waiting for the other participant."

After Player A chooses between A and B, you will see the same information displayed to player A, alongside player A's decision and the outcome for the period. An example is included below:

Your matched player A chose: A.

Your payoff for this round is: 26 tokens.

Please Choose.

A

More then 25 tokens if q is less than or equal to 0.5

Less than 22 tokens if q is greater than 0.5

B

25 tokens

You will then decide how many tokens from a pot of 50 tokens to allocate to your matched player A and a random participant. This does not impact your payoff for the experiment, but is a separate pot of 50 tokens to allocate between your matched player A and a random other participant. For example:

Please choose an amount between 0 and 60 tokens to allocate to your randomly matched player A. The amount you choose will be allocated to your matched player A.

During this time, player A will see a screen reading "Waiting for the other participant."

Once this decision has been made, the round is over.

Players will then be matched to another random participant in the opposite group, and this process will be repeated until all ten rounds are complete.

To ensure the experiment runs smoothly, please ensure you click "Next" when finished making a decision.

Please contact the experimenter if you have any questions or experience a problem during the experiment.

Once you have read the instructions above and are satisfied, please click "Next" below.

The final set of instructions is displayed to participants in stage 3 is as follows:

In this stage you will be asked to choose either option A or B. The task will appear on screen as follows.

Please Choose.

A

26 tokens if q is less than or equal to 0.5

5 tokens if q is greater than 0.5

B

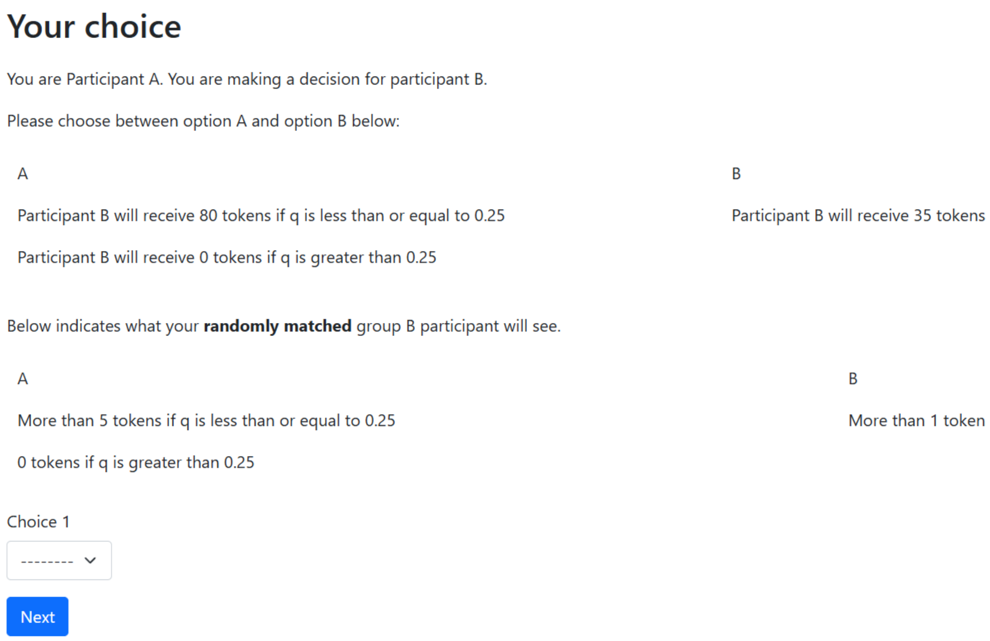
25 tokens

If one of these rounds is chosen to be the round that you will be paid for, then the lottery in A will then play out (the value of q is randomly decided between 0 and 1 by the computer) and the number of tokens you receive will then depend on either the outcome of the lottery if you chose A (in this either 26 or 5 case tokens) , or the number of tokens related to B (in this case 25 tokens). The outcome can be between 0 and 100 tokens.

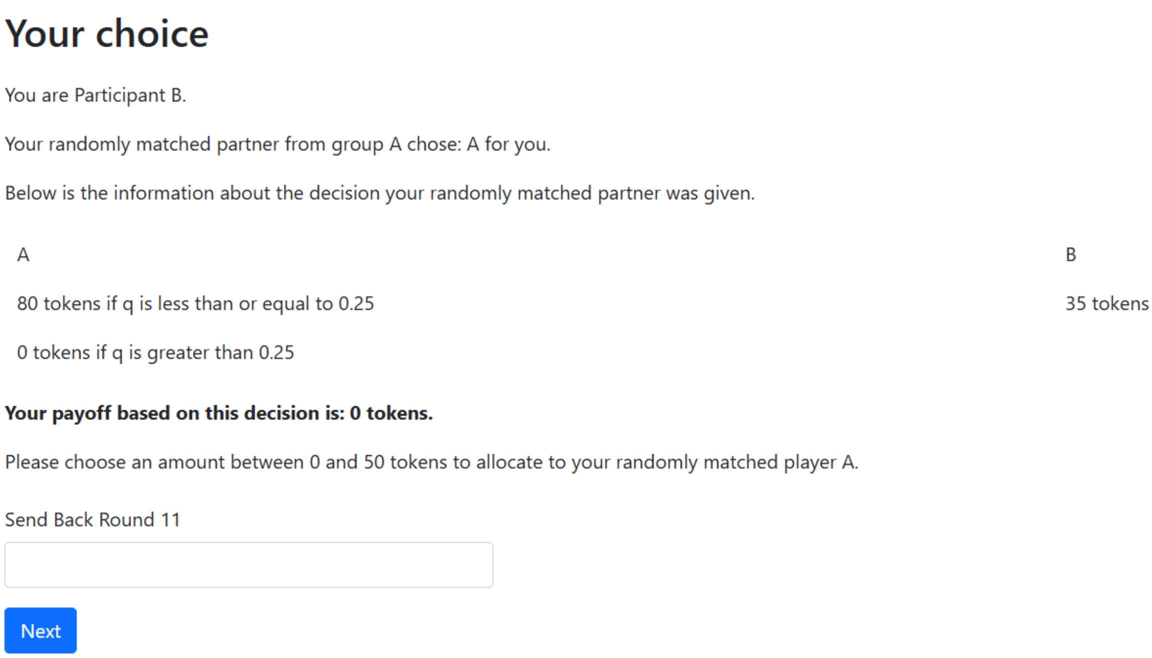
Please contact the experimenter if you have any questions or experience a problem during the experiment.

The decisions were displayed as follows.

For Group A, in the partial information treatment, the image shows the decision that the player is asked to make, with the drop down displaying A and B in a random order. Similarly, the lottery and sure thing switch sides randomly throughout the experiment.



For Group B, the decision is shown on screen as follows.



The participant is required to input a value between 0 and 50.

Group A's decision for the full information treatment is displayed as follows, and requires a choice of A or B as in the partial information treatment. Similarly, A and B are randomised in their location.

Your Choice

You are Participant A. You are making a decision for participant B.

Please choose between option A and option B below:

A	B
Participant B will receive 80 tokens if q is less than or equal to 0.25	35 tokens
Participant B will receive 0 tokens if q is greater than 0.25	

Your randomly matched player B participant will see this information, alongside your choice and their potential payoff for the round.

Choice 11

Next

Group B's decision is displayed as follows, with the same conditions as the partial information treatment, whereby the participant must input between 0 and 50.

Your choice

You are Participant B.

Your randomly matched partner from group A chose: A for you.

Below is the information about the decision your randomly matched partner was given.

A	B
80 tokens if q is less than or equal to 0.25	35 tokens
0 tokens if q is greater than 0.25	

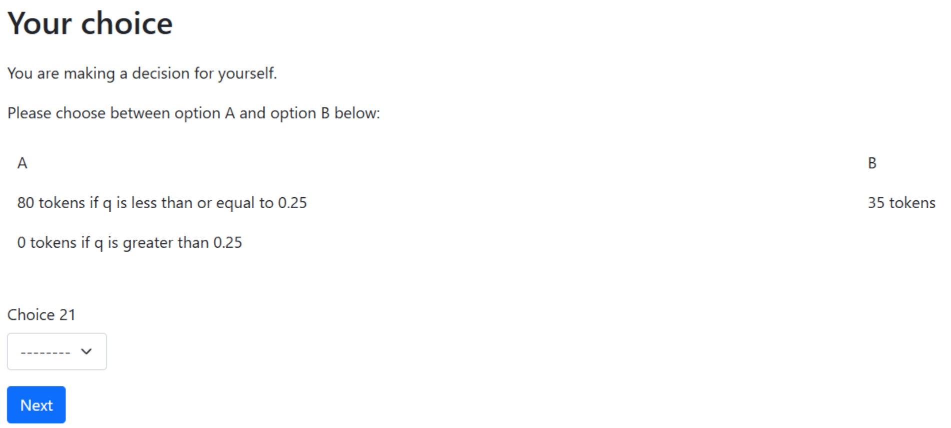
Your payoff based on this decision is: 0 tokens.

Please choose an amount between 0 and 50 tokens to allocate to your randomly matched player A.

Send Back Round 11

Next

For the own treatment, Group A and B are asked to make the same decision. Hence, the tasks are displayed in the same way, as indicated below. Similarly to the partial information treatment and 2, the drop down menu requires a choice between A and B, and the choice and placement of A and B on the page are randomised.



Appendix B

Appendix For Chapter 2

B.1 Monte Carlo Simulation

This Appendix presents further results of the simulation in Section 6 when the noise is low ($\xi = 0.130$) in Table B.1, and when it is high ($\xi = 0.200$) in Table B.2.

TABLE B.1: Mean and standard deviations of the parameters.

Parameter	True value	MLE	MSLE	HB
λ	1.648	1.576	1.640	1.648
s.e.	-	0.050	0.035	0.035
λ_{FEMALE}	0.365	0.696	0.368	0.379
s.e.	-	0.117	0.057	0.055
σ_λ	0.100	-	0.088	-
s.e.	-	-	0.036	-
r	0.500	0.537	0.500	0.497
s.e.	-	0.017	0.023	0.021
r_{FEMALE}	0.100	0.065	0.105	0.107
s.e.	-	0.030	0.028	0.027
σ_r	0.050	-	0.047	-
s.e.	-	-	0.012	-
w_g	0.540	0.559	0.540	0.536
s.e.	-	0.012	0.050	0.022
w_l	0.510	0.527	0.509	0.506
s.e.	-	0.012	0.048	0.023
ξ	0.130	0.134	0.130	0.128
s.e.	-	0.005	0.005	0.043

The Table reports estimates from the simulation exercise on the three inference methods : Maximum Likelihood Estimation (MLE), Maximum Simulated Likelihood Estimation (MSLE) and Hierarchical Bayesian (HB), for the low level of noise (0.13). For each parameter θ, the Table reports the mean of the point estimates, in the case of MLE, the mean of the distributions in the case of MSLE, and of the posterior mean of the distributions in the case of HB. Standard deviations in parentheses.

TABLE B.2: Mean and standard deviations of the parameters.

Parameter	True value	MLE	MSLE	HB
λ	1.648	1.572	1.640	1.670
s.e.	-	0.072	0.051	0.054
λ_{FEMALE}	0.365	0.705	0.371	0.398
s.e.	-	0.173	0.081	0.089
σ_λ	0.100	-	0.072	-
s.e.	-	-	0.057	-
r	0.500	0.539	0.500	0.489
s.e.	-	0.024	0.031	0.032
r_{FEMALE}	0.100	0.062	0.106	0.111
s.e.	-	0.045	0.040	0.041
σ_r	0.050	-	0.043	-
s.e.	-	-	0.022	-
w_g	0.540	0.560	0.540	0.534
s.e.	-	0.018	0.077	0.036
w_l	0.510	0.528	0.509	0.502
s.e.	-	0.018	0.075	0.038
ξ	0.200	0.202	0.199	0.198
s.e.		0.008	0.006	0.042

The Table reports estimates from the simulation exercise on the three inference methods : Maximum Likelihood Estimation (MLE), Maximum Simulated Likelihood Estimation (MSLE) and Hierarchical Bayesian (HB), for the high level of noise (0.20). For each parameter θ, the Table reports the mean of the point estimates, in the case of MLE, the mean of the distributions in the case of MSLE, and of the posterior mean of the distributions in the case of HB. Standard deviations in parentheses.

B.2 Sample size

Table B3 reports the results of the simulation exercise when the size sample of 100 increases by a factor of 2 (N=200) and 5 (N=200). All the parameter values

TABLE B3: Mean and standard deviations of the parameters.

	True value	N=100	N=200	N=500
λ	1.648	1.637	1.640	1.648
s.e.	-	0.062	0.029	0.021
λ_{FEMALE}	0.365	0.374	0.370	0.368
s.e.	-	0.118	0.046	0.031
σ_{λ}	0.100	0.081	0.088	0.093
s.e.	-	0.046	0.029	0.016
r	0.500	0.500	0.503	0.504
s.e.	-	0.031	0.017	0.015
r_{FEMALE}	0.100	0.105	0.099	0.096
s.e.	-	0.031	0.022	0.017
σ_r	0.050	0.046	0.048	0.050
s.e.	-	0.014	0.008	0.005
w_g	0.540	0.543	0.542	0.542
s.e.	-	0.014	0.041	0.029
w_l	0.510	0.510	0.511	0.511
s.e.	-	0.013	0.043	0.029
ξ	0.150	0.150	0.150	0.150
s.e.	-	0.005	0.003	0.002

The Table reports estimates from the simulation exercise using Maximum Simulated Likelihood Estimation (MSLE) for three levels of sample size (N) namely 100, 200 and 500. Standard deviations in parentheses.

Appendix C

Appendix For Chapter 3

C.1 Appendix

Note that estimates are rounded to 2 significant figures. It is noted that for Gamma estimates, no standard errors are reported. The reasons for this are twofold, firstly as the Nyarko and Schotter, 2002 model provides some explanation of this, as the model does not provide statistical significance, but provides the best explanation of beliefs possible at present, and secondly as the parameter represents a weighting between two models.

TABLE B1: Likelihood Ratio Tests for Nyarko & Schotter 2002 fixed matched elicitation Data

Participant	**LRT**
1	7.795641e-06
2	8.559455e-04
3	2.790775e-04
4	1.239120e-06
5	2.035178e-05
6	1.002677e-03
7	1.212960e-09
8	1.180611e-08
9	5.070130e-04
10	1.135067e-03
11	4.540431e-03
12	1.489720e-02
13	2.748978e-04
14	2.176754e-04
15	6.404516e-06
16	1.759431e-05
17	2.373717e-04
18	3.190443e-03
19	3.477806e-05
20	1.764579e-05
21	8.987695e-06
22	5.414643e-05
23	1.163008e-07
24	9.248624e-04
25	2.582428e-10
26	2.828257e-05
27	8.021069e-05
28	1.487373e-02

TABLE B2: Likelihood Ratio Tests for Nyarko & Schotter 2002 randomly matched elicitation Data

participant	LRT
1	0.00
2	0.07
3	0.00
4	0.00
5	0.00
6	0.00
7	0.15
8	0.00
9	0.01
10	0.01
11	0.00
12	0.03
13	0.00
14	0.00
15	0.38
16	0.00
17	0.00
18	0.00
19	0.06
20	0.01
21	0.00
22	0.00
23	0.00
24	0.00
25	0.00
26	0.00
27	0.03
28	0.00

	Nyarko and Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1-Theta
1	0.003	0.36*	0.004	0.358*	0.599
		(0.114)		(0.103)	(2.439)
2	0.958	0.000	0.842	0.000	0.733
		(0.12)		(0.07)	(151.044)
3	0.967	0.000	0.965	0.000	0.688
		(0.136)		(0.111)	(363.759)
4	0.060	0.000	0.060	0.000	0.313
		(0.074)		(0.085)	(1.55)
5	0.007	0.034	0.010	0.034	0.245
		(0.074)		(0.079)	(5.252)
6	0.060	0.000	0.935	0.000	0.877
		(0.057)		(0.137)	(84.304)
7	0.002	0.149	0.002	0.149	0.839
		(0.089)		(0.089)	(1.)
8	0.009	0.011	0.009	0.011	0.003
		(0.087)		(0.087)	(1.)
9	0.835	0.152	0.835	0.152	1.000
		(0.187)		(2.951)	(50.459)
10	0.060	0.000	0.850	0.000	0.839
		(0.051)		(0.083)	(292.71)
11	0.993	0.205	0.993	0.205	1.000
		(0.184)		(6.659)	(34.743)
12	0.990	0.434	0.990	0.434	0.99*
		(0.519)		(0.606)	(0.199)
13	0.060	0.000	0.885	0.000	0.669
		(0.07)		(0.098)	(442.388)
14	0.000	0.000	0.966	0.277	0.270
		(0.092)		(0.526)	(1.948)
15	0.568	0.013	0.568	0.013	1.000
		(0.108)		(1.363)	(676.702)
16	0.004	0.057	0.006	0.057	0.171
		(0.087)		(0.087)	(1.)
17	0.060	0.000	0.820	0.000	0.605
		(0.059)		(0.089)	(497.496)
18	1.000	0.178	1.000	0.178	1.000
		(16.066)		(5.765)	(5.985)
19	0.028	0.102	0.201	0.102	0.000
		(0.079)		(0.073)	(1.999)
20	0.004	0.057	0.006	0.057	0.171
		(0.087)		(0.087)	(1.)
21	0.708	0.016	0.708	0.016	1.000
		(0.179)		(0.73)	(132.017)

FIGURE C.1: Beta, Gamma, and Theta Values for Nyarko & Schotter 2002 fixed matched elicitation Data (Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko and Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1-Theta
22	0.003	0.248*	0.273	0.248	0.000
		(0.113)		(0.219)	(0.347)
23	0.060	0.000	0.060	0.000	0.313
		(0.087)		(0.087)	(1.509)
24	0.938	0.052	0.938	0.052	1.000
		(0.197)		(1.27)	(9.32)
25	0.004	0.057	0.004	0.057	0.486
		(0.087)		(0.087)	(1.)
26	0.012	0.103	0.224	0.102	0.000
		(0.088)		(0.076)	(1.913)
27	0.000	0.198*	0.002	0.197*	0.649
		(0.091)		(0.091)	(1.)
28	1.000	0.378	1.000	0.378	1.000
		(0.209)		(1.677)	(3.052)

FIGURE C.2: Beta, Gamma, and Theta Values for Nyarko & Schotter 2002 fixed matched elicitation Data

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1-Theta
1	0.01	0.2 *	0.01	0.2 *	0.67
	(1.00)	(0.09)	(1.00)	(0.09)	(1.00)
2	1.00	1.00	1.00	0.91	0.77 *
	(0.19)	(0.90)	0.36)	(1.14)	(0.31)
3	0.04	0.10	0.16	0.10	0.00
	(14.98)	(0.29)	(38.20)	(0.07)	(2.32)
4	1.00	0.11	1.00	0.11	1.00
	(35.68)	(99.19)	(7.15)	(22.68)	(17.48)
5	0.00	0.13	1.00	0.15	0.19
	(6.62)	(0.08)	(2.38)	(4.24)	(24.48)
6	0.01	0.33 *	1.00	0.45	0.44
	(1.00)	(0.10)	(2.05)	(12.66)	(27.23)
7	1.00	1.00	1.00	1 *	1.00
	(0.25)	(0.95)	(0.42)	(0.32)	(0.77)
8	0.03	0.08	0.04	0.08	0.96
	(20.45)	(0.07)	(1.00)	(0.09)	(1.00)
9	1.00	0.41 *	1.00	0.41	1.00
	(0.11)	(0.17)	(0.14)	(0.25)	(1.37)
10	1.00	0.41 *	1.00	0.41	1.00
	(0.11)	(0.17)	(0.29)	(0.52)	(4.15)
11	0.00	0.13	0.01	0.13	0.67
	(6.62)	(0.08)	(7.13)	(0.08)	(1.73)
12	1.00	0.90	1.00	0.72	0.76 *
	(0.20)	(0.97)	(0.20)	(0.54)	(0.32)
13	0.74	0.24	0.74	0.24	1.00
	(0.79)	(0.24)	(1.11)	(0.40)	(11.88)
14	0.01	0.01	0.01	0.01	0.00
	(1.00)	(0.09)	(1.00)	(0.09)	(1.00)
15	1.00	1 *	1.00	1.00	1.00
	(0.03)	(0.39)	(0.06)	(11.79)	(3.37)
16	0.01	0.57 *	0.05	0.57 *	0.08
	(1.00)	(0.12)	(1.00)	(0.12)	(1.00)
17	0.00	0.22 *	1.00	0.29	0.51
	(1.00)	(0.09)	(2.77)	(7.58)	(39.20)
18	0.01	0.57 *	0.05	0.57 *	0.08
	(1.00)	(0.12)	(1.00)	(0.12)	(1.00)
19	1.00	0.28	1.00	0.28	1.00
	(4.40)	(13.65)	(1.86)	(11.38)	(6.35)
20	1.00	0.26	1.00	0.26	1.00
	(1.00)	(0.99)	(1.63)	(1.49)	(5.21)

FIGURE C.3: Beta, Gamma, and Theta Values for Nyarko & Schotter 2002 randomly matched elicitation Data

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Participant	Gamma	Beta
21	0.00	0.22 *	0.02	0.22 *	0.71
	(1.00	(0.09)	(1.00)	(0.09)	(1.00)
22	0.01	0.01	0.01	0.01	0.00
	(1.00)	(0.09)	(1.00)	(0.09)	(1.00)
23	0.01	0.03	0.01	0.03	0.25
	(10.70)	(0.07)	(16.86)	(0.08)	(5.24)
24	0.33	0.00	0.06	0.00	0.31
	(78.63)	(0.08)	(1.00)	(0.09)	(1.00)
25	0.92	0.00	0.06	0.00	0.31
	(95.82)	(0.10)	(1.24)	(0.08)	(1.59)
26	0.00	0.25 *	0.28	0.25	0.00
	(5.03)	(0.11)	(86.45)	(0.24)	(0.30)
27	1.00	0.91 *	1.00	0.74	0.77
	(0.05)	(0.25)	(3.00)	(17.75)	(21.59)
28	0.00	0.22 *	0.00	0.22 *	0.65
	(1.00)	(0.09)	(1.00)	(0.09)	(1.00)

FIGURE C.4: Beta, Gamma, and Theta Values for Nyarko & Schotter 2002 randomly matched elicitation Data
(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

TABLE B3: Likelihood Ratio Test Results for Attanasi 2018 Data (Participants 1-150.)

Participant	Likelihood Ratio	Participant	Likelihood Ratio	Participant	Likelihood Ratio
1	8.47	101	1.04	151	1.02
2	1.08	102	1.02	152	0.77
3	8.47	103	3.92	153	3.97
4	0.01	104	0.00	154	7.63
5	7.63	105	1.02	155	7.63
6	8.47	106	1.04	156	0.00
7	0.00	107	4.55	157	7.63
8	7.63	108	0.02	158	1.04
9	0.61	109	0.97	159	0.84
10	1.04	110	1.04	160	0.07
11	0.57	111	1.08	161	1.04
12	0.00	112	0.82	162	0.00
13	0.84	113	0.83	163	1.02
14	8.47	114	7.63	164	7.63
15	0.61	115	7.63	165	0.00
16	1.08	116	1.08	166	0.00
17	0.82	117	1.04	167	0.00
18	8.47	118	0.82	168	1.01
19	0.00	119	0.01	169	0.82
20	0.00	120	0.00	170	1.04
21	0.02	121	1.04	171	1.02
22	1.01	122	0.01	172	0.84
23	0.06	123	1.04	173	0.00
24	1.04	124	3.91	174	0.00
25	7.63	125	1.04	175	0.00
26	0.83	126	0.83	176	0.00
27	1.41	127	0.00	177	0.00
28	1.45	128	1.04	178	1.01
29	1.45	129	1.04	179	7.63
30	7.63	130	1.05	180	1.04
31	3.97	131	0.77	181	0.84
32	1.04	132	0.84	182	0.00
33	1.08	133	7.63	183	7.63
34	0.03	134	0.00	184	0.00
35	1.45	135	1.04	185	0.00
36	1.41	136	0.00	186	1.08
37	0.62	137	1.04	187	1.08
38	7.63	138	1.04	188	7.63
39	3.97	139	0.84	189	0.84
40	1.04	140	1.04	190	1.04
41	1.45	141	0.00	191	0.00
42	1.08	142	1.04	192	1.02
43	8.47	143	0.84	193	1.04
44	0.02	144	7.63	194	7.63
45	7.63	145	0.82	195	0.82
46	0.82	146	0.00	196	0.00
47	7.63	147	0.06	197	1.04
48	1.04	148	1.04	198	1.04
49	1.08	149	1.02	199	0.00
50	1.01	150	1.04	200	7.63

TABLE B4: Likelihood Ratio Test Results for Attanasi 2018 data (Participants 151-250)

Participant	Likelihood Ratio	Participant	Likelihood Ratio
151	1.02	201	1.04
152	0.77	202	0.82
153	3.97	203	0.83
154	7.63	204	7.63
155	7.63	205	0.00
156	0.00	206	1.01
157	7.63	207	1.04
158	1.04	208	1.08
159	0.84	209	0.82
160	0.07	210	0.83
161	1.04	211	7.63
162	0.00	212	7.63
163	1.02	213	1.04
164	7.63	214	0.00
165	0.00	215	1.01
166	0.00	216	1.45
167	0.00	217	1.04
168	1.01	218	0.77
169	0.82	219	1.45
170	1.04	220	0.00
171	1.02	221	1.45
172	0.84	222	1.04
173	0.00	223	0.82
174	0.00	224	0.00
175	0.00	225	0.84
176	0.00	226	1.04
177	0.00	227	1.04
178	1.01	228	0.82
179	7.63	229	0.77
180	1.04	230	1.04
181	0.84	231	1.08
182	0.00	232	0.84
183	7.63	233	0.00
184	0.00	234	1.01
185	0.00	235	0.00
186	1.08	236	1.04
187	1.08	237	1.04
188	7.63	238	0.00
189	0.84	239	1.04
190	1.04	240	1.08
191	0.00	241	1.04
192	1.02	242	0.00
193	1.04	243	1.04
194	7.63	244	1.04
195	0.82	245	1.08
196	0.00	246	1.02
197	1.04	247	0.00
198	1.04	248	1.04
199	0.00	249	1.04
200	7.63	250	0.77

TABLE B5: Likelihood Ratio Test Results for Attanasi 2018 Data (Participants 251-320)

Participant	Likelihood Ratio	Participant	Likelihood Ratio
251	1.01	301	0.00
252	0.82	302	1.04
253	0.84	303	1.08
254	0.77	304	0.01
255	1.04	305	0.62
256	0.84	306	7.63
257	7.63	307	3.97
258	0.77	308	1.04
259	1.04	309	1.08
260	0.84	310	0.00
261	1.04	311	1.04
262	0.84	312	0.77
263	0.77	313	0.84
264	1.04	314	0.00
265	0.84	315	0.01
266	1.02	316	0.84
267	1.04	317	1.04
268	1.04	318	1.04
269	0.00	319	7.63
270	1.08	320	1.02
271	1.01		
272	1.04		
273	0.77		
274	1.04		
275	1.08		
276	0.77		
277	0.00		
278	0.84		
279	0.00		
280	1.02		
281	0.00		
282	0.77		
283	0.84		
284	0.84		
285	1.04		
286	0.84		
287	0.00		
288	0.77		
289	1.08		
290	1.02		
291	1.04		
292	1.04		
293	0.77		
294	1.04		
295	0.77		
296	1.02		
297	1.04		
298	0.77		
299	1.02		
300	1.08		

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
1	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
2	0.817	0	0.756	0	0.192
		(0.004)		(0.02)	(0.556)
3	0.901	0	0.85	0	0.087
		(0.006)		(0.02)	(0.459)
4	0.901	0	0.85	0	0.087
		(0.006)		(0.02)	(0.459)
5	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
6	0.901	0	0.85	0	0.087
		(0.006)		(0.02)	(0.459)
7	0.009	0	0.013	0	0.263 *
		(0.02)		(0.001)	(0.013)
8	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
9	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
10	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
11	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
12	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
13	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
14	0.901	0	0.85	0	0.087
		(0.006)		(0.02)	(0.459)
15	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
16	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)
17	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
18	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
19	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
20	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)
21	1	0.166	1	0	0
		(0.202)		(0.299)	(0.225)
22	1	0.239	1	0.24*	0
		(0.159)		(0.013)	(0.093)

FIGURE C.5: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
23	1	0.166	1	0	0
		(0.202)		(0.299)	(0.225)
24	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
25	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
26	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
27	0.009	0	0.013	0	0.263 *
		(0.02)		(0.001)	(0.013)
28	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
29	0.009	0	0.639	0	0.559 *
		(0.02)		(0.001)	(0.013)
30	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
31	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
32	0	1 *	0.767	1*	1 *
		(0.02)		(0.001)	(0.018)
33	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
34	1	0.654	1	0.65*	0
		(0.347)		(0.007)	(0.057)
35	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
36	0.396	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.012)
37	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
38	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
39	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
40	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
41	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
42	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
43	0.009	0	0.013	0	0.263 *
		(0.02)		(0.001)	(0.013)
44	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)

FIGURE C.6: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
45	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
46	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
47	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
48	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
49	1	0.166	1	0	0
		(0.202)		(0.299)	(0.225)
50	1	0.239	1	0.24*	0
		(0.159)		(0.013)	(0.093)
51	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
52	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
53	0.001	1 *	0.002	1*	0.914 *
		(0.019)		(0.001)	(0.015)
54	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
55	0.001	1 *	0.002	1*	0.914 *
		(0.019)		(0.001)	(0.015)
56	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
57	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
58	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
59	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)
60	0.799	0	0.326	0	0.259
		(0.003)		(0)	(0.215)
61	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
62	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
63	0.311	0	0.013	0	0.263 *
		(0.003)		(0.001)	(0.012)
64	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)
65	1	0.654	1	0.65*	0
		(0.347)		(0.007)	(0.057)

FIGURE C.7: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
66	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
67	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
68	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
69	1	0.239	1	0.24*	0
		(0.159)		(0.013)	(0.093)
70	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
71	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
72	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
73	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
74	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
75	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.012)
76	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
77	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
78	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
79	0	1 *	0.767	1*	1 *
		(0.02)		(0.001)	(0.018)
80	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
81	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.011)
82	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
83	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
84	1	0.239	1	0.24*	0
		(0.159)		(0.013)	(0.093)
85	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.011)
86	0.009	0	0.013	0	0.263 *
		(0.02)		(0.001)	(0.013)
87	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)

FIGURE C.8: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
88	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
89	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
90	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
91	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
92	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.011)
93	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.011)
94	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
95	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
96	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
97	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
98	0	1 *	0.767	1*	1 *
		(0.02)		(0.001)	(0.018)
99	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
100	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
101	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
102	0.009	0	0.013	0	0.263 *
		(0.02)		(0.001)	(0.013)
103	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
104	0.817	0	0.756	0	0.192
		0.004		0	0.556
105	0.021	1 *	0.009	1*	0.036 *
		(0.02)		(0.001)	(0.013)
106	0.001	1 *	0.002	1*	0.914 *
		(0.019)		(0.001)	(0.015)
107	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
108	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)

FIGURE C.9: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
109	0	1 *	0.742	1*	1 *
		(0.02)		(0.002)	(0.014)
110	0.942	1 *	0.378	1*	0.52 *
		(0.023)		(0.001)	(0.013)
111	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
112	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
113	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
114	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
115	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
116	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
117	0.009	0	0.013	0	0.263 *
		(0.02)		(0.001)	(0.013)
118	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.011)
119	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
120	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
121	0.009	0	0.09	0	0.429 *
		(0.02)		(0.001)	(0.013)
122	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
123	0.021	1 *	0.009	1*	0.036 *
		(0.02)		(0.001)	(0.013)
124	0.799	0	0.326	0	0.259
		(0.003)		(0)	(0.215)
125	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
126	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
127	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
128	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
129	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
130	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)

FIGURE C.10: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
131	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
132	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
133	0	1 *	0.767	1*	1 *
		(0.02)		(0.001)	(0.018)
134	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
135	0.396	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.012)
136	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
137	0.799	0	0.326	0	0.259
		(0.003)		(0)	(0.215)
138	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
139	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)
140	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
141	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
142	0.404	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.013)
143	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
144	0.799	0	0.326	0	0.259
		(0.003)		(0)	(0.215)
145	0.799	0	0.326	0	0.259
		(0.003)		(0)	(0.215)
146	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
147	0	1 *	0	1*	0.959 *
		(0.03)		(0.001)	(0.01)
148	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
149	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
150	0.901	0	0.85	0	0.087
		(0.006)		(0)	(0.459)
151	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
152	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)

FIGURE C.11: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
153	0.942	1 *	0.391	1*	0.885 *
		(0.021)		(0.001)	(0.013)
154	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
155	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
156	0.379	0	0.013	0	0.263 *
		(0.002)		(0.001)	(0.012)
157	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
158	0	1 *	0.733	1*	1 *
		(0.02)		(0.003)	(0.011)
159	1	1 *	1	1*	0
		(0.059)		(0.001)	(0.013)
160	0.817	0	0.756	0	0.192
		(0.004)		(0)	(0.556)
161	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
162	1	1 *	0.837	0	0.067
		(0.075)		(0)	(0.076)
163	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
164	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
165	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
166	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
167	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
168	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
169	0.967	1 *	1	1*	0
		(0.234)		(0.01)	(0.028)
170	0.295	0	0.022	1*	0.166 *
		(0.01)		(0.001)	(0.039)
171	1	1 *	0.954	0	0.126
		(0.037)		(0)	(3.496)
172	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
173	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
174	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)

FIGURE C.12: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
175	0.967	1 *	1	0.17	0
		(0.234)		(0.01)	(0.028)
176	1	1 *	0.837	0	0.067
		(0.075)		(0)	(0.076)
177	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
178	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
179	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
180	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
181	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
182	0.009	0.549 *	0.736	0	0.07 *
		(0.008)		(0)	(0.029)
183	1	0.277 *	0.042	0	0.936 *
		(0.065)		(0.001)	(0.013)
184	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
185	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
186	1	0.051	0.042	0	0.936 *
		(0.136)		(0.001)	(0.013)
187	1	1 *	0.954	0	0.126
		(0.037)		(0)	(3.496)
188	0.967	1 *	1	0.17	0
		(0.234)		(0.01)	(0.028)
189	0.967	1 *	1	0.17	0
		(0.234)		(0.01)	(0.028)
190	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
191	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
192	0.295	0	0.022	1*	0.166 *
		(0.01)		(0.001)	(0.039)
193	1	1 *	0.837	0	0.067
		(0.075)		(0.23)	(0.076)
194	1	0.016	0.469	0	0.684 *
		(0.037)		(0.001)	(0.013)
195	0.967	1 *	1	0.17	0
		(0.234)		(0.01)	(0.028)
196	1	1 *	0.954	0	0.126
		(0.037)		(0)	(3.496)

FIGURE C.13: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
197	0.793	1 *	1	0.65	0
		(0.04)		(0.015)	(0.062)
198	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
199	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
200	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
201	0.967	1 *	1	0.17	0
		(0.234)		(0.01)	(0.028)
202	1	1 *	0.837	0	0.067
		(0.075)		(0)	(0.076)
203	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
204	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
205	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
206	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
207	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
208	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
209	1	1 *	0.837	0	0.067
		(0.075)		(0)	(0.076)
210	0.009	0.549 *	0.736	0	0.07 *
		(0.008)		(0)	(0.029)
211	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
212	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
213	0.869	0	0	1*	0.034 *
		(0.005)		(0.001)	(0.013)
214	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
215	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
216	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
217	0.967	1 *	1	0.17*	0
		(0.234)		(0.01)	(0.028)
218	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)

FIGURE C.14: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
219	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
220	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
221	0.002	1 *	0.542	0	0.664
		(0.014)		(0)	(4.662)
222	1	1 *	0.954	0	0.126
		(0.037)		(0)	(3.496)
223	1	1 *	0.837	0	0.067
		(0.075)		(0)	(0.076)
224	1	1 *	0.837	0	0.067
		(0.075)		(0)	(0.076)
225	0.009	0.549 *	0.736	0	0.07 *
		(0.008)		(0)	(0.029)
226	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
227	0.009	0	1	0.24	0
		(0.02)		(0.002)	(0.043)
228	1	0.016	0.469	0	0.684 *
		(0.037)		(0.001)	(0.013)
229	0.295	0	0.022	1*	0.166 *
		(0.01)		(0.001)	(0.039)
230	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
231	0.869	0	0	1*	0.034 *
		(0.005)		(0.001)	(0.013)
232	1	0.205 *	0.042	0	0.936 *
		(0.08)		(0.001)	(0.013)
233	0.92	0.3 *	0.042	0	0.936 *
		(0.025)		(0.001)	(0.013)
234	0.92	0.3 *	0.042	0	0.936 *
		(0.025)		(0.001)	(0.013)
235	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
236	0.967	1 *	1	0.17*	0
		(0.234)		(0.01)	(0.028)
237	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
238	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
239	0.869	0	0	1*	0.034 *
		(0.005)		(0.001)	(0.013)
240	0.009	0.549 *	0.736	0	0.07 *
		(0.008)		(0.001)	(0.029)

FIGURE C.15: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
241	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
242	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
243	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
244	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
245	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
246	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
247	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
248	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
249	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
250	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
251	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
252	1	1 *	0.954	0	0.126
		(0.037)		(0.001)	(3.496)
253	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
254	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
255	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
256	0.92	0.3 *	0.042	0	0.936 *
		(0.025)		(0.001)	(0.013)
257	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
258	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
259	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
260	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
261	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
262	1	1 *	0.954	0	0.126
		(0.037)		(0.001)	(3.496)

FIGURE C.16: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
263	1	0.205 *	0.042	0	0.936 *
		(0.08)		(0.001)	(0.013)
264	0.877	0	0	1*	0.034 *
		(0.003)		(0.001)	(0.013)
265	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
266	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
267	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
268	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
269	1	0.277 *	0.042	0	0.936 *
		(0.065)		(0.001)	(0.013)
270	0.801	0	0	1*	0.917 *
		(0.003)		(0.001)	(0.034)
271	0.967	1 *	1	0.17*	0
		(0.234)		(0.01)	(0.028)
272	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
273	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
274	0.92	0.3 *	0.042	0	0.936 *
		(0.025)		(0.001)	(0.013)
275	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
276	0.967	1 *	1	0.17*	0
		(0.234)		(0.01)	(0.028)
277	1	0.205 *	0.042	0	0.936 *
		(0.08)		(0.001)	(0.013)
278	1	0.051	0.042	0	0.936 *
		(0.136)		(0.001)	(0.013)
279	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
280	0.094	0	0	1*	0.901 *
		(0.004)		(0.001)	(0.012)
281	1	0.205 *	0.042	0	0.936 *
		(0.08)		(0.001)	(0.013)
282	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
283	1	0.016	0.469	0	0.684 *
		(0.037)		(0.001)	(0.013)
284	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)

FIGURE C.17: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.
(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
285	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
286	1	1 *	0.954	0	0.126
		(0.037)		(0.001)	(3.496)
287	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
288	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
289	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
290	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
291	1	1 *	0.954	0	0.126
		(0.037)		(0.001)	(3.496)
292	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
293	0.295	0	0.022	1*	0.166 *
		(0.01)		(0.001)	(0.039)
294	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
295	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
296	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
297	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
298	1	0.114	0.187	0	0.587 *
		(0.084)		(0.001)	(0.013)
299	0.009	0.549 *	0.736	0	0.07 *
		(0.008)		(0.001)	(0.029)
300	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
301	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
302	1	1 *	0.954	0	0.126
		(0.037)		(0.01)	(3.496)
303	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
304	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
305	0.793	1 *	1	1*	0
		(0.04)		(0.015)	(0.062)
306	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)

FIGURE C.18: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.

(Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

	Nyarko & Schotter Model		Wishful Thinking Model		
Participant	Gamma	Beta	Gamma	Beta	1- Theta
307	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
308	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
309	0.869	0	0	1*	0.034 *
		(0.005)		(0.001)	(0.013)
310	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
311	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
312	1	0.277 *	0.042	0	0.936 *
		(0.065)		(0.001)	(0.013)
313	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
314	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
315	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
316	0.293	0	0.235	1*	0.332 *
		(0.01)		(0.001)	(0.013)
317	0	1 *	1	1*	0
		(0.02)		(0.01)	(0.025)
318	0.77	0	0.954	1*	0.271 *
		(0.01)		(0.001)	(0.013)
319	0.009	0	1	0.24*	0
		(0.02)		(0.002)	(0.043)
320	1	1 *	0.837	0	0.067
		(0.075)		(0.02)	(0.076)

FIGURE C.19: Beta, Gamma and 1-Theta estimates for Attanasi (2018) Data.
Rounded to 2 Significant Figures. * indicates significance at the 5 percent level, and SEs are indicated in brackets in the below the parameter estimates.)

Bibliography

Agranov, Marina et al. (2014). “An experimental study of the impact of competition for other people’s money: the portfolio manager market”. In: *Experimental Economics* 17, pp. 564–585.

Andersson, Ola et al. (2016). “Deciding for others reduces loss aversion”. In: *Management Science* 62.1, pp. 29–36.

— (2020). “Risking other people’s money: Experimental evidence on the role of incentives and personality traits”. In: *The Scandinavian Journal of Economics* 122.2, pp. 648–674.

Attanasi, Giuseppe et al. (Dec. 2018). “Belief-dependent preferences and reputation: Experimental analysis of a repeated trust game”. In: *Journal of Economic Behavior Organization* 167. DOI: 10.1016/j.jebo.2018.12.002.

Babad (1987). “Wishful Thinking and objectivity among sports fans”. In: *Social Behaviour* 2 (4).

Baillon, A. et al. (2020). “Searching for the Reference Point”. In: *Management Science* 66.1, pp. 93–112.

Balcombe, K. et al. (2015). “Parametric Preference Functionals under Risk in the Gain Domain: A Bayesian Analysis”. In: *Journal of Risk and Uncertainty* 50.2, pp. 161–187. ISSN: 1573-0476.

Bernstein, Daniel et al. (Feb. 2011). “Hindsight Bias From 3 to 95 Years of Age”. In: *Journal of experimental psychology. Learning, memory, and cognition* 37, pp. 378–91. DOI: 10.1037/a0021971.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag New York.

Booij, A. et al. (2009). “A Parametric Analysis of Prospect Theory’s Functionals for the General Population”. In: *Theory and Decision* 68, pp. 115–148.

Bouchouicha, R. et al. (2019). “Gender effects for loss aversion: Yes, no, maybe?” In: *Journal of Risk and Uncertainty* 59, pp. 171–184.

Brooks, P. et al. (2005). “Loss Averse Behavior”. In: *Journal of Risk and Uncertainty* 31, pp. 301–325.

Bruner, Jerome S et al. (1947). “Value and need as organizing factors in perception.” In: *The journal of abnormal and social psychology* 42.1, p. 33.

Camerer, C. (1995). “Individual Decision Making”. In: *The Handbook of Experimental Economics*. Ed. by J. Kagel et al. Princeton University Press.

Camerer, Colin et al. (Mar. 1997). “Experience-Weighted Attraction Learning in Games: A Unifying Approach”. In: 1003. URL: https://ideas.repec.org/p/clt/sswopa/1003.html.

Chakravarty, Sujoy et al. (2011). “Are you risk averse over other people’s money?” In: *Southern Economic Journal* 77.4, pp. 901–913.

Charness, G. et al. (2010). “Portfolio Choice and Risk Attitudes: an Experiment”. In: *Economic Inquiry* 48.1, pp. 133–146.

Charness, G et al. (2012). “Strong Evidence for Gender Difference in Risk Taking”. In: *Journal of Economic Behavior & Organization* 83.1, pp. 50–58.

Charness, Gary et al. (2009a). “The Role of Responsibility in Strategic Risk-Taking”. In: *Journal of Economic Behavior & Organization* 69.3, pp. 241–247.

Charness, Gary et al. (2009b). “The role of responsibility in strategic risk-taking”. In: *Journal of Economic Behavior & Organization* 69.3, pp. 241–247.

Chen, Daniel L. et al. (2016). “oTreeâ€”An open-source platform for laboratory, online, and field experiments”. In: *Journal of Behavioral and Experimental Finance* 9.C, pp. 88–97. DOI: 10.1016/j.jbef.2015.12.001. URL: https://ideas.repec.org/a/eee/beexfi/v9y2016icp88-97.html.

Cheung, Yin-Wong et al. (1997). “Individual learning in normal form games: Some laboratory results”. In: *Games and economic behavior* 19.1, pp. 46–76.

Conte, Anna et al. (2011). “Mixture models of choice under risk”. In: *Journal of Econometrics* 162.1, pp. 79 –88. ISSN: 0304-4076.

Croson, R. et al. (2009). “Gender Differences in Preferences”. In: *Journal of Economic Literature* 47.2, pp. 448–474.

Eckel, Catherine C et al. (2002). "Sex differences and statistical stereotyping in attitudes toward financial risk". In: *Evolution and human behavior* 23.4, pp. 281–295.

Eckel, Catherine C. et al. (2008). "Men, Women and Risk Aversion: Experimental Evidence". In: ed. by Charles R. Plott et al. Vol. 1. Handbook of Experimental Economics Results. Elsevier, pp. 1061 –1073. DOI: https://doi.org/10.1016/S1574-0722(07)00113-8. URL: http://www.sciencedirect.com/science/article/pii/S1574072207001138.

Engelmann, Jan B et al. (2024). "Anticipatory anxiety and wishful thinking". In: *American Economic Review* 114.4, pp. 926–960.

Erev, Ido et al. (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria". In: *The American Economic Review* 88.4, pp. 848–881. ISSN: 00028282. URL: http://www.jstor.org/stable/117009 (visited on 09/29/2022).

Eriksen, Kristoffer W et al. (2010). "Myopic investment management". In: *Review of Finance* 14.3, pp. 521–542.

Eriksen, Kristoffer W et al. (2020). "Risk-taking on behalf of others". In: *Journal of Behavioral and Experimental Finance* 26, p. 100283.

Ferecatu, A. et al. (2016). "Heterogeneous Risk and Time Preferences". In: *Journal of Risk and Uncertainty* 53.1, pp. 1–28.

Filippin, A. et al. (2016). "A Reconsideration of Gender Differences in Risk Attitudes". In: *Management Science* 62.11, pp. 3138–3160.

Fornasari, Federico et al. (2020). "Interpersonal risk assessment and social preferences: An experimental study". In: *Journal of Economic Psychology* 77, p. 102183.

Fox, Craig R. et al. (2009). "Chapter 11 - Prospect Theory and the Brain". In: *Neuroeconomics*. Ed. by Paul W. Glimcher et al. London: Academic Press, pp. 145 –173. ISBN: 978-0-12-374176-9. DOI: https://doi.org/10.1016/B978-0-12-374176-9.00011-7. URL: http://www.sciencedirect.com/science/article/pii/B9780123741769000117.

Füllbrunn, Sascha C et al. (2017). "Decision making for others: The case of loss aversion". In: *Economics Letters* 161, pp. 154–156.

Füllbrunn, Sascha et al. (2020). "Current issues in decision making for others". In: *Journal of Economic Psychology* 77. Decision Making for Others, p. 102250. ISSN:

0167-4870. DOI: https://doi.org/10.1016/j.joep.2020.102250. URL: https://www.sciencedirect.com/science/article/pii/S0167487020300076.

Gao, X. et al. (2020). "Estimating Risk Preferences for Individuals: A Bayesian Approach". CEAR Working Paper 2020-15.

Gaudecker, H. von et al. (2011). "Heterogeneity in Risky Choice Behaviour in a Broad Population". In: *American Economic Review* 101.2, pp. 664–694.

Gelman, A. et al. (1992). "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4, pp. 457–472. ISSN: 08834237.

Gneezy, Uri et al. (1997). "An experiment on risk taking and evaluation periods". In: *The quarterly journal of economics* 112.2, pp. 631–645.

Gneezy, Uri et al. (2023). "Can wishful thinking explain evidence for overconfidence? An experiment on belief updating". In: *Oxford Economic Papers* 75.1, pp. 35–54.

Gurdal, Mehmet Y et al. (2013). "Why blame?" In: *Journal of Political Economy* 121.6, pp. 1205–1247.

Harrison, G. et al. (2008). "Expected Utility Theory and Prospect Theory: one Wedding and a Decent Funeral". In: *Experimental Economics* 12.2, p. 133. ISSN: 1573-6938.

Heller, Yuval et al. (June 2020). "Biased-Belief Equilibrium". In: 2006.15306. URL: https://ideas.repec.org/p/arx/papers/2006.15306.html.

Hey, J. (2014). "Choice Under Uncertainty: Empirical Methods and Experimental Results". In: *Hanbook of the Economics of Risk and Uncertainty*. Ed. by M. Machina et al. Eslevier B.V., pp. 809–850.

Hey, J. et al. (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data". In: *Econometrica* 62.6, pp. 1291–1326.

Hoelzl, Erik et al. (2005). "Overconfident: Do You Put Your Money on It?" In: *The Economic Journal* 115.503, pp. 305–318. ISSN: 00130133, 14680297. URL: http://www.jstor.org/stable/3590396 (visited on 10/31/2022).

Holt, Charles A et al. (2002a). "Risk Aversion and Incentive Effects". eng. In: *The American economic review* 92.5, pp. 1644–1655. ISSN: 0002-8282.

— (2002b). "Risk aversion and incentive effects". In: *American economic review* 92.5, pp. 1644–1655.

Holt, Charles A. et al. (2014). "Chapter 4 - Assessment and Estimation of Risk Preferences". In: *Handbook of the Economics of Risk and Uncertainty*. Ed. by Mark Machina et al. Vol. 1. Handbook of the Economics of Risk and Uncertainty. North-Holland, pp. 135 –201. DOI: https://doi.org/10.1016/B978-0-444-53685-3.00004-0. URL: http://www.sciencedirect.com/science/article/pii/B9780444536853000040.

Huber, J. et al. (2001). "On the Similarity if Classical and Bayesian Estimates of Individual Mean Partworths". In: *Marketing Letters* 12.3, pp. 259–269.

Humphrey, Steven J et al. (2011). *The social costs of responsibility*. Tech. rep. CeDEx discussion paper series.

Ioannou, Christos A. et al. (2014). "A generalized approach to belief learning in repeated games". In: *Games and Economic Behavior* 87, pp. 178–203. ISSN: 0899-8256. DOI: https://doi.org/10.1016/j.geb.2014.05.007. URL: https://www.sciencedirect.com/science/article/pii/S0899825614000888.

Jacquement, N. et al. (2018). *Experimental Economics: Methods and Applications*. Cambridge University Press.

König-Kersting, Christian et al. (2016). "Ambiguity attitudes in decisions for others". In: *Economics Letters* 146, pp. 126–129.

Kvaløy, Ola et al. (2014). "Playing the trust game with other people's money". In: *Experimental Economics* 17, pp. 615–630.

Luzuriaga, Miguel et al. (2017). "Taking Risk with Other People's Money: Does Information about the Others Matter?" In: *Review of Behavioral Economics* 4.2, pp. 107–133.

Mayraz, Guy (2011). "Wishful thinking". In: *Available at SSRN 1955644*.

Moffatt, P. (2016). *Experimetrics*. Macmillan Palgrave.

Montinari, Natalia et al. (2013). *Social preferences under risk: the role of social distance*. Tech. rep. Jena Economic Research Papers.

Murphy, R. et al. (2018). "Hierarchical Maximum Likelihood Parameter Estimation for Cumulative Prospect Theory: Improving the Reliability of Individual Risk Parameter Estimates". In: *Management Science* 64, pp. 308–326.

Nilsson, H. et al. (2011). "Hierarchical Bayesian Parameter Estimation for Cumulative Prospect Theory". In: *Journal of Mathematical Psychology* 55, pp. 84–93.

Nyarko, Yaw et al. (July 2002). "An Experimental Study of Belief Learning Using Elicited Beliefs". In: *Econometrica* 70. DOI: 10.1111/1468-0262.00316.

Oechssler, Jörg et al. (2009). "Cognitive abilities and behavioral biases". In: *Journal of Economic Behavior Organization* 72.1, pp. 147–152. ISSN: 0167-2681. DOI: https://doi.org/10.1016/j.jebo.2009.04.018. URL: https://www.sciencedirect.com/science/article/pii/S0167268109001103.

Offerman, Theo et al. (Oct. 2002). "Imitation and Belief Learning in an Oligopoly Experiment". In: *The Review of Economic Studies* 69.4, pp. 973–997. ISSN: 0034-6527. DOI: 10.1111/1467-937X.00233. eprint: https://academic.oup.com/restud/article-pdf/69/4/973/4431109/69-4-973.pdf. URL: https://doi.org/10.1111/1467-937X.00233.

Paetzel, Fabian et al. (2018). "Cognitive ability and in-group bias: An experimental study". In: *Journal of Public Economics* 167, pp. 280–292. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2018.04.006. URL: https://www.sciencedirect.com/science/article/pii/S0047272718300677.

Pahlke, Julius et al. (2012). "Risk-taking for others under accountability". In: *Economics Letters* 114.1, pp. 102–105.

Plummer, M. (2017). *JAGS Version 4.3.0 User Manual*. Tech. rep.

Pollmann, Monique MH et al. (2014). "Risk taking by agents: The role of ex-ante and ex-post accountability". In: *Economics Letters* 123.3, pp. 387–390.

Polman, Evan (2012). "Self–other decision making and loss aversion". In: *Organizational Behavior and Human Decision Processes* 119.2, pp. 141–150.

Polman, Evan et al. (2020). "Decision making for others involving risk: A review and meta-analysis". In: *Journal of Economic Psychology* 77. Decision Making for Others, p. 102184. ISSN: 0167-4870. DOI: https://doi.org/10.1016/j.joep.2019.06.007. URL: https://www.sciencedirect.com/science/article/pii/S016748701930306X.

Qui, J. et al. (2011). "Understanding the two Components of Risk Attitudes: an Experimental Analysis". In: *Management Science* 57, pp. 193–199.

Quiggin, J. (1982). "A Theory of Anticipated Utility". In: *Journal of Economic Behavior and Organization* 3.4, pp. 323–343.

Reynolds, Douglas B et al. (2009). “Risky shift versus cautious shift: determining differences in risk taking between private and public management decision-making”. In: *Journal of business & economics research (JBER)* 7.1.

Rouder, J. et al. (2005). “An Introduction to Bayesian Hierarchical Models with an Application in the Theory of Signal Detection”. In: *Psychonomic Bulletin & Review* 55, pp. 84–93.

Schmidt, U. et al. (2002). “An Experimental Test of Loss Aversion”. In: *Journal of Risk and Uncertainty* 25.3, pp. 233–249.

Schmidt, U et al. (2005). “What is Loss Aversion”. In: *Journal of Risk &Uncertainty* 30.2, pp. 157–167.

Stahl, D. (2014). “Heterogeneity of Ambiguity Preferences”. In: *The Review of Economics and Statistics* 96.5, pp. 609–617.

Starmer, C. (2000). “Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk”. In: *Journal of Economic Literature* 38.2, pp. 332–382. DOI: 10.1257/jel.38.2.332.

Stott, H. (2006). “Cumulative Prospect Theory’s Functional Menagerie”. English. In: *Journal of Risk and Uncertainty* 32.2, pp. 101–130. ISSN: 0895-5646. DOI: 10.1007/s11166-006-8289-6.

Toubia, O. et al. (2013). “Dynamic Experiments for Estimating Preferences: an Adaptive Method of Eliciting Time and Risk Parameters”. In: *Management Science* 59, pp. 613–640.

Train, K. (2009). *Discrete Choice Methods with Simulation*. 2nd. Cambridge University Press.

Tversky, A. et al. (1992). “ Advances in Prospect Theory: Cumulative Representation of Uncertainty”. In: *Journal of Risk and Uncertainty* 5.4, pp. 297–323.

Vieider, Ferdinand M. et al. (2015). “Common Componetns of Risk and Uncertainty Attitudes Across Contexts and Domains: Evidence from 30 Countries”. In: *Journal of the European Economic Association* 13.3, pp. 421–452. DOI: https://doi.org/10.1111/jeea.12102. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jeea.12102. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jeea.12102.

Vieider, Ferdinand M et al. (2016). “Risk taking for oneself and others: A structural model approach”. In: *Economic Inquiry* 54.2, pp. 879–894.

Wakker, P. (2010). *Prospect Theory*. Cambridge University Press.

Weinstein, Neil D. (1980). “Unrealistic optimism about future life events”. In: *Journal of Personality and Social Psychology* 39, pp. 806–820.

Wilcox, N. (2011). “Stochastically more Risk Averse: A Contextual Theory of Stochastic Discrete Choice under Risk”. In: *Journal of Econometrics* 162.1, pp. 89 –104.

Xie, Zhixin et al. (2017). “Investigating Gender Differences under Time Pressure in Financial Risk Taking”. In: *Frontiers in Behavioural Neuroscience* 11.246.

Yildiz, Muhamet (Jan. 2007). “Wishful Thinking in Strategic Environments”. In: *The Review of Economic Studies* 74.1, pp. 319–344. ISSN: 0034-6527. DOI: 10.1111/j.1467-937X.2007.00423.x. eprint: https://academic.oup.com/restud/article-pdf/74/1/319/18347267/74-1-319.pdf. URL: https://doi.org/10.1111/j.1467-937X.2007.00423.x.