Essays in Applied and Experimental Behavioural Science

by

Jorge Lajos Kossuth Cabrejos

Thesis

Submitted to the University of Warwick

For the Degree of

Doctor of Philosophy

Warwick Business School

April 2022

Contents

List of Tables

List of Figures

Acknowledgements

I am beyond grateful to my supervisors for their guidance throughout this process. Nick P., you gave me my first experience with cool, thorough, and relevant experimental academic research, which also ended up being my first academic publication (I think we did a great job!). You were also always there to give pertinent and kind advice, especially when I needed it the most. Nick C., you have been an inspiring role model for what I aspire to be as a senior academic: humble, insightful, and empathic. And even though we didn't interact as much these four years, you were always available to have a chat and listen to my ideas and even my woes, and I thank you very much for that.

I am also grateful for having met amazing academics, professionals, and human beings throughout my journey. In particular, I would like to thank Roberto Asmat, Alessandro Castagnetti, Luciana Gaspar, Carlos Cardona, Junyang Guo, Merle van den Akker, Lara Vomfell, Ali E. Ahmed, Ahmed Maged Nofal, Elettra Latini, Silvia Granato, Erika Herrera, and Julia Wyjadlowska.

I am also thankful to the University of Warwick for having financed three and a half years of my doctoral studies via the Chancellor's International Scholarship. Without that financial help, I wouldn't have been able to move to the UK to realize one of my dreams.

Finally, I want to express my highest gratitude to Irene and Jorge, Sylvia, Stefano, and Cristina. Mom and dad, your guidance and emotional support were fundamental these four years, and I thank you for all you gave me, all your sacrifices, and for always prioritizing our education. Vovô Sylvia, although you're no longer with us, your kindness, unconditional love, and relentless strength will forever be inscribed in my heart. Stefano, brother, we have been companions since forever, and I have always admired your intelligence and academic achievements; I hope I can make you proud with this thesis. Cristina, you are and always will be the main reason behind my success. You are the smartest woman I have ever met. And the kindest. And the warmest. And the prettiest. One minute by your side is enough to forgo all my pessimism (that I stubbornly call realism). Eres mi sol. Thank you for everything.

Declaration

The work comprising this thesis is the result of my own research conducted under the supervision of Professor Nattavudh Powthdavee and Professor Nick Chater, at Warwick Business School, between 2017 and 2021. I declare it has not been submitted for an award or degree at any other university. Chapter 1 is co-authored with Roberto Asmat (PhD candidate at the University of Warwick's Department of Economics). Chapter 3 is co-authored with Nattavudh Powthdavee (Warwick Business School), Nick Chater (Warwick Business School), and Donna Harris (University of Oxford). I am the sole author of Chapter 2.

Chapter 1 is currently under review at the American Economic Journal: Applied Economics. Roberto Asmat contributed to the research design (50/100), data building (50/100), estimation analysis (100/100), conceptual framework (100/100), manuscript writing and revisions (50/100). Lajos Kossuth contributed to the research design (50/100), data building (50/100), randomization analysis (100/100), interpretation of results (100/100), manuscript writing and revisions (50/100). Lajos Kossuth's overall contribution: 50%. A signed co-authorship statement is attached to the thesis. This chapter has been uploaded as a working paper to the Social Science Research Network (SSRN)[1] repository.

The project presented in Chapter 3 was designed, conducted, analysed, written, and submitted to journals during the preparation of this thesis. We received the IRB approval from the Humanities and Social Sciences Research Ethics Committee at the University of Warwick on 15/05/2018 (Ethics Application Reference: 109/17-18). It was accepted for publication in the Journal of Economic Behavior and Organization[2] on January 9th, 2020 (received 8th September, 2019; revised 2nd December, 2019). Lajos Kossuth and Nattavudh Powdthavee contributed to the conceptualization and writing of the original draft (methodology, formal analysis), and with review and editing. Nick Chater and Donna Harris contributed to the conceptualization of the original draft and with review and editing.

[1] http://dx.doi.org/10.2139/ssrn.3964747
[2] https://doi.org/10.1016/j.jebo.2020.01.006

Abstract

This thesis comprises three essays on Applied and Experimental Behavioural Science. Broadly, it explores topics in the behavioural sciences related to gender differences in decision-making under incomplete information, income-rank social comparisons and their relationship with income-based subjective wellbeing, and *emotional hedging* – betting against the occurrence of desired outcomes. Causal analysis is achieved by using state-of-the-art experimental and quasi-experimental methodologies.

In Chapter 1, my co-author and I compare decisions by female and male judges in child support trials with complete and incomplete information where a randomly assigned judge decides on the father's income allocation to children. We contribute to the literature on gender differences in judicial decision-making, but also to a novel literature that attempts not only to detect outcome disparities, but also to study the reasoning behind these differences in decision-making processes. We find that female judges set lower awards in both cases in comparison to male judges. However, the gender gap under incomplete information is around two thirds the size of that under complete information. In exploring mechanisms underlying this difference, we propose a simple framework to elicit judges' beliefs about the unknown income in such cases by using their judicial behaviour in cases where income is known. We find that female judges estimate that the unknown income is higher. Thus, gender differences in estimated beliefs act as a countervailing force and explain the attenuation of judges' gender differences in decisions under incomplete information.

In Chapter 2, I estimate the direct and indirect effects - through changes in absolute income and income rank - of the Peruvian "JUNTOS" Programme on Income Satisfaction. Causal identification comes from a difference-in-differences analysis that exploits the variation in the rollout of the programme in different districts at different times, together with its eligibility criterion: a poverty score calculated ad hoc for JUNTOS. The degree to which JUNTOS is perceived to be satisfactory is important for two reasons. First, while a cash transfer will certainly increase income levels, it might not be enough to improve the provision of a minimum desired amount of goods and services. Second, measures of income satisfaction might provide better accounts of welfare inequality for policymakers to follow, since they are closer to people's intuitive feelings about inequality. I find that JUNTOS only affects Income Satisfaction indirectly through changes in absolute income and income rank, results that do not support a strong income rank hypothesis (absolute income is a stronger determinant of income satisfaction than income rank). They also suggest neighbours are the most important reference group and show that social comparisons (based on income rank) only appear for those above the poverty line.

Finally, in Chapter 3, my co-authors and I study whether people gain significant emotional benefits from not engaging in *emotional hedging* – deciding to bet against the occurrence of a desired outcome. While there is plenty of evidence of people being strongly averse to *emotional hedging* – which suggests a possible violation of the standard utility theories, since they would be avoiding minimizing gains and losses -, surprisingly little is known about the success of either strategy – hedging versus not hedging – at maximizing experienced utility during as well as following the realization of the outcome. After six rounds of a lab-in-the-field experiment during the 2018 FIFA World Cup, we find, first, that only a minority of subjects choose to bet against the success of the England football team in the tournament. We also find that betting for England to win produces a sharp and significant decline in happiness when England loses their matches. Conversely, post-match happiness was more stable for those who hedged regardless of the expected match outcome. These results suggested that, despite hedging being a more efficient strategy in monetary and affective terms, people did not engage in it.

1 Chapter 1: Gender Differences in Judicial Decisions under Incomplete Information: Evidence from Child Support Cases[3]

Job Market Paper

With Roberto Asmat[4]

We compare decisions by female and male judges in child support trials where a judge decides on the father's income allocation to children. We investigate two types of cases: 1) when fathers have a formal job, their income is known to judges, and awards are set as a fraction of it, and 2) when fathers work in the informal sector, their income is unknown to judges, and awards are given as a fixed amount of money. By exploiting random assignment of cases to judges, we find that female judges set lower awards in both cases in comparison to male judges. However, the gender gap under incomplete information is around two thirds the size of that under complete information. In exploring mechanisms underlying this difference, we propose a simple framework to elicit judges' beliefs about the unknown income in such cases by using their judicial behaviour in cases where income is known. We find that female judges – relative to male judges – estimate that the unknown income is higher. Thus, gender differences in estimated beliefs act as a countervailing force and explain the attenuation of judges' gender differences in decisions under incomplete information.

[3] We are grateful to Roland Rathelot, Ben Lockwood and Manuel Bagues for their supervision and guidance. We also thank James Fenske and Dan Bernhardt for their helpful suggestions. This project also benefited from feedback received from Greta Morando, Alessandro Castagnetti and Rodolfo Belleza, and participants at the European Society for Population Economics (ESPE) 2021 conference, Applied Microeconomics to Peru (AMP) 2021 seminar, XXVII Meeting on Public Economics at Universitat de Barcelona in 2020, Royal Holloway's 2019 PhD Workshop on Labour and Family Economics, Latin American Association in Law and Economics (ALACDE) 2017 conferences, and PhD Forums at Warwick University during 2017-2020. Authors thank Vera Troeger and the Centre for Competitive Advantage in the Global Economy (CAGE) for financial support to build the dataset. Diego Niño and Lidia San Miguel provided excellent research assistance.

[4] Department of Economics, Warwick University (email: r.asmat-belleza@warwick.ac.uk)

1.1 Introduction

The economics literature presents evidence on the factors affecting child support allocation when it is decided by a bargaining process between separated parents (Chiappori & Weiss, 2007). It is surprisingly less informative on how transfers from the non-custodial parent to the child are determined when there is disagreement, and they are adjudicated by a court. In this setting, the judiciary follows set legal guidelines for child support allocation with some room for discretion, leaving room for additional factors to influence the outcomes. In this chapter, we explore one of these factors: the gender of judges. In particular, would male and female judges make different decisions regarding a `fair' allocation of child support?

Gender differences in decision-making have been thoroughly studied in the literature. For instance, there is ample evidence on how gender stereotypes shape beliefs about ability in oneself and others in different categories of knowledge. They contribute to gender gaps in self-confidence, assessments of others, and behavior in cooperative games. Evidence shows both men and women tend to underestimate the ability of women relative to men in male-typed domains (Bordalo et al., 2019). These issues could play a vital role in trials in which gender stereotypes are more prevalent, such as child support trials.

There is also a large body of evidence suggesting than women and men behave differently when making decisions under uncertainty, i.e. when the probability of different events is not known (Croson and Gneezy, 2009; Byrne and Worthy, 2016). For instance, men tend to perform better than women in both the uncertainty and risk phases of the Iowa Gambling Task (Singh et al., 2020; van den Bos et al., 2013). Further, it has also been shown that uncertainty induces a significant increase in the performance of men relative to that of women in settings where subjects compete against each other (Balafoutas & Sutter, 2019), adding to the vast evidence on gender differences in competitiveness (see Niederle (2016) for a primer).

Specifically, studies on sex-related differences in judicial decision-making characterize men as objective, neutral and equidistant, while women are thought of as caring, vocational and more prone to like being involved with persons (Jeandidier et al., 2016). Naturally, men and women behave differently in social contexts, have different aspirations, backgrounds, and societal expectations, and all this ought to play a role in how they perceive the world and what they consider to be fair.

Boyd (2016), for instance, anticipates that female judges will tend to favour their own gender because of their historic marginalization in positions of power and their different perspectives in life. What she calls ‘diversity theories’ argue that a female judge would bring a ‘female perspective’ due to all the unique knowledge regarding specific cases of discrimination that male judges simply do not experience in real life (Boyd, 2016:2). Also, female judges might favour their own gender because of a sense of representation, since they might be seen as high-profile representatives of their gender, or class, to champion these groups’ wellbeing.

Further, if the social norms within the judicial decision-making context have been made historically by males, the higher the number of women that ‘join the organization’, the more acceptable it will be for them to deviate from the male behaviour that sets the social norms within said organization (Scheurer, 2014). This would mean that the more female judges there are, the more likely it could be that stereotypes regarding mothers as care-takers and fathers as providers are subverted.

Any attempt to map judges' preferences by gender from observed judicial decisions faces another important empirical challenge, however: the judiciary is often not able to observe all relevant information to make a decision. Hence, these decisions may reflect both judges' preferences about the fair amount of child support to allocate but also their beliefs about the missing information. In particular, the vital piece of information in child support cases is the income of respondents, since it is used as a reference by judges to reach a verdict. When the respondent works in the informal economy, having complete knowledge about his income might be difficult, so judges must form beliefs about the unknown income before deciding how much child support to allocate. This issue is more salient in developing countries due to the large informal labour markets, which affects judicial decision-making (Sadka et al., 2020).

In this chapter we study judicial decisions in randomly assigned child support cases with complete and incomplete information about the income of respondents. In the former, since the respondent works in the formal economy (from now on, formal cases), the judge observes his income, and decides on a child support amount that is expressed as a percentage of the respondent's income. In the latter, however, the respondent works in the informal economy (from now on, informal cases) and judges first need to form beliefs about his unknown income before deciding a child support amount, which is ultimately being awarded in absolute terms. Our sample therefore comprises judges that make decisions in both contexts, giving us a unique opportunity to inspect the degree to which incomplete information might play a role in shaping gender differences in the allocation of child support, but also in the elicitation of beliefs about the income of respondents as a potential mechanism.

We use the Peruvian judicial institutional setting for the following reasons. First, the features of the justice system allow us to identify a causal effect of gender differences in judicial decisions, since child support cases are randomly assigned to female and male judges. Second, due to the highly informal nature of the labour market, the distribution of cases across the formal and informal economies allows us to conduct the analysis described in the previous paragraph. Third, the legal objective in child support cases is narrow and measurable: the award of a reasonable amount of money to meet the needs of the children. There are no other motives such as punishment or deterrence typically found in criminal cases. Finally, child support cases are heavily gender coded. In Peru, women are generally expected to take care of children and men to provide income and physical protection. Indeed, more than 98% of the cases we analysed involve a mother suing a father.

Our database includes over 3,000 published child support cases in Lima, Peru, where the petitioner is always the mother and custodial parent, and the respondent is always the father. We extract information from two stages of a typical child support case. In the first stage, the settlement hearing, the parties are encouraged to negotiate and agree on a child support amount in order to avoid the expenses of proceeding to litigation. If the parties fail to settle, the case then proceeds to litigation. Here, judges have to decide a child support amount based on the evidence presented by the parties. Depending on the job type of the respondent, the ruling is a percentage of his income (formal job) or a fixed amount of money (informal job).

Our central finding is that, relative to their male counterparts, female judges allocate smaller amounts of child support per child. Moreover, we also find that these gender differences depend on whether the income of the father is observable. The gender-based gap is -6.8% in formal cases (0.25 standard deviations) and -5.9% in informal ones (0.16 standard deviations). These results are striking for two reasons. First, female judges seem to be more harsh towards the female petitioners, although this finding is in line with some evidence found in decisions about employment (M. F. Bagues & Esteve-Volart, 2010) and academic evaluations (M. Bagues et al., 2017). Second, results show that the gender-based gap found in formal cases - when there is complete information about the income of respondents - is 57.3% higher than that in informal ones.

To understand how the lack of information about the income of respondents attenuates the gender-based gap, we develop a simple framework of judicial decision-making in which incomplete information operates through beliefs. We interpret the verdict reached by a judge in a formal case as their revealed child support allocation preference. We then assume the same judge should exhibit the same preference in an informal case, the difference being the degree to which they would also need to form a belief about the income of the father. To illustrate, if the award given by a judge in a formal case is 30% of the respondent's income and the same judge awards S/. 300 in an informal case with observably similar respondents, we then infer that the judge must believe the respondent earns S/. 1,000. So, we use the estimated preference from the formal case, together with the child support amount allocated by that same judge in the informal case, to infer how they would form a belief about the respondent's income in that same informal case.

By restricting the sample to judges who face formal and informal cases and applying the framework described, we find a gender-based gap of around -4.7% in cases where income is unknown. This gender gap is composed of a gender gap in preferences for child support allocation of -12% and a gender gap in estimates of fathers' income of 7.33%. This means female judges - relative to their male counterparts - infer higher levels of income when information about it is incomplete or non-existent. In other words, even though female and male judges are exposed to similar sets of incomplete information (due to random allocation of cases to judges), female judges exhibit a higher reference point (estimate of the income of respondents) to decide on a 'fair' allocation of child support. Thus, gender differences in estimates about the father's income act as a countervailing force and explain the attenuation of gender differences in decisions under incomplete information.

To gain a better understanding of why judges' estimates of fathers' income differ by gender, we explore the extent to which judges rely on the mother's claim (amount of money) as a signal about the unknown income of respondents. In informal cases, mothers state their claims arguing that this amount of money is a fair estimation of what the father is able to pay. It is worth noting that claims are recorded before the case is randomly assigned to a judge, and so are independent of the gender of judges. We study the relationship between beliefs and claims by using a standard Bayesian updating framework in which judges estimate the unknown income based on their priors and the signal (claim) sent by mothers. We find that female judges rely less on the claim when estimating the unknown income. This might suggest that one underlying explanation for the lack of homophily found in the reduced-form estimates is that female judges rely less on claims made by female petitioners than male judges.

Finally, we also consider how other characteristics of judges and of the children involve in the trial influence decision-making. Regarding judges' characteristics, we collected data on age, work experience and self-reported wealth of judges from official sources. First, we check whether judges significantly differ across those characteristics and find that male and female judges are similar. Second, we include those characteristics as controls in our baseline regressions. We find that only for informal cases, judge's age and status have a significant negative and positive effect on the award, respectively. However, the gender gap remains statistically significant and we find no evidence of a change in the magnitude of the gap. Regarding children's characteristics, we collected data on whether they experienced any serious health issue, and average age and gender composition of the total number of children involved in the trial. Similar as the previous case, we find that the gender gap remains statistically significant. However, we also find that the gap differential between formal and informal cases is lower: 30.2% compared to the 57.3% found in the reduced form estimations.

This chapter contributes to the large literature on the role of gender in judicial decision-making. Most of this research has focused on criminal cases (e.g. Gruhl, Spohn and Welch (1981); Coontz (2000); Collins and Moyer (2008)), discrimination cases (e.g. Farhang and Wawro (2004)), and a range of civil rights issues such as immigration appeals (see Gill, Kagan and Marou (2015)) and issues affecting women[5] (e.g. Martin (1989); Peresie (2005); Boyd, Epstein and Martin (2010); Boyd (2016)), all in the context of Common Law. To the best of our knowledge, this study contributes to the unexplored branch of family law and child support cases, which has several advantages for the study of judges' gender differences in judicial decisions.

First, in many settings, judicial decisions are made by a panel of two or three judges (see Peresie (2005); Gill, Kagan and Marouf (2015)) and the interaction between them poses the additional problem of how to disentangle the views of female and male judges. However, in our setting, a single judge decides child support cases, so we can attribute differences to the gender of judges by exploiting random assignment. Second, the objective in child support cases is simple and narrow, as opposed to the typical issues studied in the literature (see Arnold, Dobbie and Yang (2018)). By isolating one legal objective, we can pin down preferences of judges on the single issue of income support for children. Third, as Kiser, Asher and McShane (2008) show, in most judicial instances, only 5% of cases reach litigation, so the interest should be in the settlement stage instead of the litigation stage. In child support cases in Lima, Peru, we find that 70% of cases that started the process reach litigation. Finally, in many instances there is female under-representation in courts, which poses an empirical threat when it comes to comparing decisions of male and female judges (Knepper, 2018). In stark contrast with other settings, in our setting, 60% of judges are female.

[5] The issues included in these studies are abortion, affirmative action, sex discrimination in employment and sexual harassment.

This chapter also contributes to a novel literature in attempting not only to detect outcome disparities, but also to study the reasoning and to learn about the decision-making process behind these results. For instance, Arnold, Dobbie and Yang (2018) test for racial bias in bail decisions by comparing misconduct rates of respondents for whom perceived benefits and costs of being released were equal for judges. In another recent study, Ash, Chen and Ornaghi (2021) look at gender attitudes to explain voting behaviour in gender-related cases in U.S. Circuit Courts. In the same spirit, this chapter contributes to understanding gender differences in judicial decision-making by inspecting the role of incomplete information in shaping potential differences.

We structure this chapter as follows. In Section 1.2, we describe the child support system in Peru. In Section 1.3, we review the data and construction of variables. In Section 1.4, we explain the random assignment of cases to judges, provide evidence in this regard, and then report the judge gender-based gaps. In Section 1.5, we show robustness checks and address sample selection concerns in our setting. In Section 1.6, we present a framework of judicial decision-making and discuss gender differences in estimates about the income of respondents as the main mechanism explaining results. Finally, the conclusion is in Section 1.7.

1.2 Overview of the Peruvian child support system

The goal in child support cases is for the judge to set an award (the specific monetary amount for child support the respondent will have to pay) based on the claims of both parties and all the available information. The criteria for setting an award are provided by the Peruvian Civil Procedure Law: a) the needs of the children who are trial matter; b) the respondent's income; and c) additional children the respondent must support. However, the way to balance these variables and how to determine the award is not stated explicitly and is at the discretion of each judge. The only firm rule is that the amount, in total, should not represent more than 60% of the respondent's income.

A crucial aspect of the legislation is that it is not necessary for a judge to know the respondent's income to set an award. About 70% of the workers in Peru work in the informal sector, and often a child support case involves adjudicating two opaque income streams. Therefore, judges' awards depend on the respondent's job status. First, in cases where the respondent has a formal job and his income is known to the judges, they set an award as a percentage of income. For the sake of simplicity, we define these cases as 'formal'. Second, in cases where the respondent works in the informal sector, his income is unknown to the judge, who sets an award as a fixed sum. We define such cases as ‘informal’. In these types of cases, a reference of the respondent's income is supposed to be made by judges, although they do not often report it. In this regard, the legislation states that the reference cannot be less than the legal minimum wage.

The process that a case goes as follows. Once a child support case is filed, it will first be revised by a court and subsequently admitted for trial. If admitted, the respondent is immediately notified and has up to five days to respond to the claim. If the respondent does not respond, he is declared a ‘rebel’ and loses his right to present evidence that supports his position. After this step, a date is set for the first stage of the process under study: the settlement hearing. The judge’s goal in this stage is to get both petitioner and respondent to settle, so they can avoid incurring in the monetary and time-consuming costs of litigating. If the petitioner and the respondent fail to settle on a specific amount for child support, the case proceeds to litigation where the judge will decide the award. Finally, if any or both parties disagree with the award, they could appeal the decision and proceed to a final stage.

1.3 Data

This chapter uses data from two Peruvian administrative sources: virtual archives of judicial records (“Consulta de Expedientes Judiciales" [CEJ]) and the national transparency agency (“La Contraloría").

1.3.1 Judicial Records

We use publicly available documents from the CEJ website relating to child support cases filed in the capital city of Lima during 2017 and 2018. CEJ provides all records of each action taken by the parties and the judge assigned in each case. For a given case, we collected documents corresponding to the two stages described in section 1.2: the settlement hearing (first stage) and the litigation (second stage). The final data set was built based on 3,015 child support cases from the website. It was collected and built in a one-year period. It required querying documents (out of a broader sample of cases), identifying cases of interest (verdicts of child support cases), and reading/extraction of variables from each document to build the data set. In the following subsections, we describe how we extracted variables from both documents.

Settlement hearing: This document is signed by the judge assigned to the case and records the attendance of parties and their attorneys (names and IDs), and characteristics of children who are involved in the trial (full names and age). As the main goal is to promote agreement between the parties, the following steps of the settlement process are conducted by the judge: i) the petitioner asks for an amount for child support, ii) the respondent either accepts or offers a different amount, iii) the judge suggests an amount, and iv) parties accept or reject the proposition. Unfortunately, most cases do not record this process and only show a no settlement/settlement result. We infer the judges' gender from their full names shown in the digital signature.

Litigation: This document is also signed by the judge assigned to the case[6] and contains the verdict and the judge's arguments. We extracted data from the 'case analysis' section[7] that contains the judge's analysis on the three criteria established by Peruvian law to determine an award. For the first criterion, the needs of the children who are involved in the trial, judges typically state their age, gender, how much the mothers spend on them, and if they have any special needs such as health conditions. For the second criterion, the economic capacity of the respondent, the judge examines all his income sources, such as salary, businesses, and properties, should there be information about them. As explained in section 1.2, according to the law it is not necessary to thoroughly investigate the income of the respondent to provide an award, and this is especially relevant for cases where the respondent works in the informal sector. Finally, for the third criterion, the judge investigates whether the respondent has other dependent children to support[8].

Table 1.1 shows a summary of the data sets used in this study. Since only districts that have more than one court apply randomisation of cases to courts, we discarded cases from districts with only one court. At the settlement hearing stage, there are 2,371 cases in total. These cases were assigned among 149 judges, 59% of whom were women. 27.4% of the cases reached an agreement and did not proceed to the ligation stage. At the litigation stage, there are 1,736 cases in total. These cases were assigned among 153 cases, 61% of whom were women. In 22.7% of cases, the respondent had a formal job.

---

[6] We only found a judge in the litigation stage different than the one observed in the settlement hearing in a few cases. This is mostly explained by an abnormal delay between these two stages such that the judge leading the court changed.

[7] The complete analysis consists of three parts: legal framework, case analysis, and verdict. In all cases, the legal framework section contains the same information about judicial principles followed and, therefore, there is no variation in this regard across cases.

[8] Descendants who are 18 years old or older, and other relatives such as parents or siblings are not considered the responsibility of the defendant, although respondents often claim to have such responsibilities.

1.3.2 Transparency Agency

We supplement the data of judicial cases by further characteristics of judges from two publicly available sources. Given their important public responsibilities, judges are closely supervised by the Peruvian Transparency Agency to detect irregularities related to corruption. From their CVs, we collected judges' ages, job position (principal judge or alternate judge) and years of experience as principal at the time of the study. The position of a judge (principal/alternate) depends on professional achievements and experience as a judge. Moreover, principal judges earn higher salaries and are held in higher esteem than alternate judges. From the second source (financial situation), we collected judges' wealth. It worth mentioning that this is self-report information, which should include savings at the financial system and valuation of their assets (mostly real estate).

1.4 Empirical Analysis

1.4.1 Identification condition

Identification is achieved by cases being randomly assigned to judges within a judicial district. According to the Peruvian Civil Law, child support cases are randomly assigned to courts within a judicial district[9]. Most judicial districts contain more than one court, and each court is led by only one judge. Thus, districts with more than one court will follow a random algorithm for assigning cases to courts. This ensures a fair distribution of caseload across courts and also prevents petitioners from targeting their cases to more favourable judges. As courts are led by one judge, court randomisation means that cases are randomly assigned to judges.

The randomisation process is conducted as follows. First, to sue for child support, the petitioner must attend a Peace Court in the judicial district corresponding to the geographical district where she lives. The forms to file a lawsuit are designed to be simple and accessible even to those who do not have the means to pay for legal services. The form is entered into the IT system in an office called “Mesa de Partes”. The case is randomly assigned to one court office out of several within the judicial district. Finally, the court office receives the lawsuit file and the process described in 1.2 starts.

[9] Due to different population sizes of geographical districts, the judicial system sets ‘judicial districts’ to aggregate small population-sized districts. For instance, two small population-sized geographical districts ‘X’ and ‘Y’ can be merged into the judicial district called ‘X-Y’.

We provide a screenshot of the randomisation step in the system as is written in its user guide[10]. This shows how the person in charge must register the case into the system (see the Appendix). We highlight the fact that it is impossible for the officer to manipulate the assignment of the case to a court. This randomisation pipeline has been confirmed by two separate sources in interviews conducted in Peru in 2019[11].

To corroborate that cases are indeed randomly assigned to judges in our data set, we conduct balancing tests. Table 1.2 presents the balance check of cases characteristics observed at the settlement hearings. For each group of judges, it presents the means and standard errors of the variables used for the analysis. The balance check is determined by the p-values of the differences-in-means two-tailed t-tests shown in the last column. It is important to note that these calculations only contemplate judicial offices in which there is at least one male and one female judge, as randomisation of cases to judges could be conducted. A statistically significant p-value suggests that there is enough statistical evidence to reject the null hypothesis of balance. As can be observed in Table 1.2, there is no evidence of imbalance in any of the observed variables.

As our main analysis is based on cases that reached the litigation stage, we also check whether cases' characteristics are balanced across female and male judges at the litigation stage. Table 1.3 and Table 1.4 present balance checks for explanatory variables in the litigation stage for formal and informal cases, respectively. Analogously to Table 1.2 they present the means and standard errors of the relevant variables used for the analysis for male judges and female judges. The last column shows the p-values of the differences-in-means two-tailed t-tests. Again, these calculations only contemplate judicial offices in which there is at least one male and one female judge. In formal cases, Table 1.3 shows no imbalance except for one variable (with 90% of confidence): the number of children outside of trial that the respondent has to support. For informal cases, Table 1.4 shows no imbalance in all cases' characteristics.

All these pieces taken together constitute robust evidence that cases were randomised and the registration of cases followed the user guide described above.

[10] The user guide is available here: http://www.csjjunin.gob.pe/archivos/modulos/pagina_web/servicios/Estadistica/Documento-10Apr2019-091826.pdf
[11] We interviewed the assistant of a judge who works in one of the courts in our study and a lawyer who had served as an attorney in child support cases.

1.4.2 Econometric specifications and results

Before we present the estimation of gender differences in decision-making, we provide a graphical analysis of decisions by gender of judges. Figures 1.1 and 1.2 show kernel distributions of child support allocations given by male and female judges in formal and informal cases. In both cases it can be clearly seen that the mean of the distribution for female judges is lower than that for male judges, indicating that female judges award lower amounts of child support on average. Moreover, the differences between female and male judges seem to be larger in formal cases than in informal cases. We estimate the judge gender gap in formal and informal cases by exploiting random assignment of cases to judges as follows:

$$(1)\quad \log\left(\frac{\alpha_{ij}}{N_i^T}\right) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \varepsilon_{ij}, \quad i \in F$$

$$(2)\quad \log\left(\frac{A_{ij}}{N_i^T}\right) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \varepsilon_{ij}, \quad i \in I$$

In equation (1), $\boldsymbol{\alpha_{ij}}$ is the award (as percentage of the respondent's income) in formal case $\boldsymbol{i} \in \boldsymbol{F}$ assigned to judge *j*. $\boldsymbol{Female_{j(i)}}$ is an indicator variable for whether formal case $\boldsymbol{i} \in \boldsymbol{F}$ was assigned to a female judge *j*. $\boldsymbol{N_i^T}$ and $\boldsymbol{N_i^{-T}}$ denote the number of children involved in formal case $\boldsymbol{i} \in \boldsymbol{F}$ and the additional number of children the respondent in formal case $\boldsymbol{i} \in \boldsymbol{F}$ needs to support, respectively. In equation (2), A_{ij} is the award (as a fixed amount of money) in informal case $\boldsymbol{i} \in \boldsymbol{I}$ assigned to judge *j*. $\boldsymbol{Female_{j(i)}}$, $\boldsymbol{N_i^T}$ and $\boldsymbol{N_i^{-T}}$are analogous variables for informal case $\boldsymbol{i} \in \boldsymbol{I}$. Finally, both equations include district $\boldsymbol{\gamma_d}$ and year $\boldsymbol{\gamma_t}$ fixed effects. All standard errors are clustered at the judge level. The main coefficient of interest is $\boldsymbol{\beta_1}$ which estimates a semi-elasticity: the percentage change in the award when the case is assigned to a female judge relative to a male judge.

Table 1.5 contains the results of the pooled OLS regressions for the formal and informal cases. Columns (i) and (ii) show that the gender-based gap is -6.8% in the formal cases, and -5.9% in the informal cases. For the formal cases, this means that female judges set an award per child that is on average 6.8% lower than that of the male judges. Looking at the other explanatory variables provides additional depth in understanding the judicial decision-making process in child support cases. Both the number of children included in the trial and the number of additional children the respondent has to support are negatively associated with the award in both formal and informal cases, since the maximum award a judge can give is 60% of the respondent's income. The coefficient associated with the number of children in trial is bigger when compared to the additional children the respondent needs to support in both formal and informal cases.

While the estimates of the gender gaps are easily interpreted (semi-elasticity), the limitation is that they cannot be comparable between formal and informal cases. Indeed, while the judge gender differences in formal cases correspond to differences in ratios (% of income awarded), in informal cases they correspond to differences in levels (amounts of money). To make them comparable we re-estimate equations (1) and (2) by standardising the awards given in formal and informal cases. Table 1.5 shows these results in columns (iii) and (iv). When a formal case is assigned to a female judge, the allocated child support amount per child is around -0.247 standard deviations relative to when it is assigned to a male judge. The analogous figure for informal cases is around -0.157 standard deviations. This means that the female judge effect is 57.3% stronger in formal cases than in informal cases. This striking result raises the question of how incomplete information has this attenuation effect in judge gender differences in child support decisions. We develop a simple framework to address the role of incomplete information in decision-making in Section 1.6.

1.5 Robustness Analysis

1.5.1 Going beyond gender

We have provided evidence of a gender-based gap when judges make child support rulings. ‘Gender’ might not be the only story behind these results, however. For instance, a female judge who is 60 years old at the time of trial might have views about what she considers a ‘fair’ allocation of child support, not because of her being female, but because she belongs to a generation with distinct social norms.

Thus, we check whether disparities are attributable to judges' characteristics other than their gender. As described in Section 1.3.2, we collected data on their ages, job status (principal or alternate), years of experience as principal and self-reported wealth. First, we inspect whether male and female judges differ along these characteristics. Second, we include all judges' characteristics in our baseline regressions.

Table 1.6 presents a balance analysis for characteristics of judges by their gender. This test shows that female and male judges do not differ in other characteristics beyond gender except for their job status: 49.2% of male judges work as principal judges while for female judges it is 35.2%.

Table 1.7 shows the estimates of equations (1) and (2) with all judges' characteristics as control variables in addition to the variables used in Table 1.5. The gender effect prevails after including these controls. For formal cases, none of the other judge's characteristics are statistically associated with awards given by judges. For informal cases, judge's age, job status, and wealth have a significant effect on awards. However, there is no evidence that the magnitude of the gender gap changes after including them.

We also check for any heterogeneity in judges' decisions by adding as covariates three important characteristics of the children who are involved in the trial: i) whether they have a serious health issue; ii) the gender composition of the total number of children involved in the trial; and iii) their average age. Table 1.8 shows the estimates of equations (1) and (2) with the addition of these characteristics as covariates. Our main result still holds: female judges appear to be harsher towards female petitioners (although we lose some statistical significance in the formal cases). The main difference, however, is that the gap in informal cases is considerably larger (-0.182 standard deviations compared to the -0.157 found in Table 1.5) and the one in formal cases is smaller (-0.237 standard deviations compared to the -0.247 found in Table 1.5). Adding children's characteristics as covariates reduces the difference in gaps between formal and informal cases: the formal gap is 30.2% bigger than the informal one in these estimations (compared to 57.3% in the standard model).

A couple of additional results are worth mentioning. First, neither the age of the children involved in the trial nor them having a serious health issue affects the amount per child given by judges as child support. However, the gender composition of the total number children involved in trial, when this number is large, does seem to matter. Compared to the case where every child is male, in cases where there is 1 of 4 or 3 of 4 female children, the average child support amount awarded is larger (except in the informal non-standardized estimations). These results are not found in cases where either half or the total amount of children involved in the trial are female.

Finally, another plausible story is the degree to which the father having had experience in previous child support cases can modify the judicial decision in the current one (via his behaviour in court, his choice of lawyer, or his strategy in selecting which information to reveal, for example). We model judicial decision making as in equations (1) and (2) but restrict the sample to those cases in which there are no additional (external) children involved in the trial. Our based assumption is that fathers with no additional children to support are experiencing a child support case for the first time. Table 1.9 shows these estimates. There is still a gender gap to be observed, albeit larger than when considering the whole sample. The gap is -0.267 standard deviations in formal cases (compared to the -0.247 found in Table 1.5) and -0.172 standard deviations in informal cases (compared to the -0.157 found in Table 1.5). However, the ratio between formal and informal gaps is slightly smaller in the restricted sample estimation: while the gap in formal cases is 55.2% bigger than the one in informal cases for these specifications, it was 57.3% when considering the full sample. These results would suggest that the magnitude of the gender gap suffers no significant alterations, but that fathers having experience in these types of trials attenuates the gender gap slightly.

1.5.2 Sample selection bias

In this section we investigate whether the litigation stage in child support cases is suitable for detecting gender-based differences in judicial decisions. As Knepper (2018) pointed out, to detect judge gender-based differences in decisions, one must also inspect judge gender–based settlement rates. Indeed, the prior literature has focused on the trial stage and has ignored the fact that, optimistically, cases reach trial 20% of the time (Kiser et al., 2008). This could make the trial stage unsuitable for detecting gender-based differences if the judge’s gender influences the likelihood of settlement and creates a sample selection bias in the trial stage, as Knepper (2018) finds in workplace sex discrimination cases.

Thus, we start with a hypothesis that the judge's gender has an effect on the probability of settlement. Exploiting random assignment of cases, we test this hypothesis as follows:

(3) $$\Pr\left(Settle_{ij} = 1\right) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 Formal_i + \gamma_d + \gamma_t + \varepsilon_{ij}$$

Where $\boldsymbol{Settle_{ij}}$ is an indicator variable for whether case $\boldsymbol{i}$ assigned to judge $\boldsymbol{j}$ settles or avoids litigation. $\boldsymbol{Female_{j(i)}}$ is an indicator variable for whether case $\boldsymbol{i}$ was assigned to a female judge $\boldsymbol{j}$. We control for $\boldsymbol{N_i^T}$, the number of children involved in case $\boldsymbol{i}$ and $\boldsymbol{Formal_i}$, an indicator variable for whether the respondent has a formal job. Regarding the latter, for cases that do not settle, the type of respondent's job is captured in litigation. While for cases that settle, we infer the type of job by the format that the agreed amount is expressed in[12]. Finally, $\boldsymbol{\gamma_d}$ and $\boldsymbol{\gamma_t}$ are district and year fixed effects, respectively.

[12] If the agreement is in percentage of income, we assume the respondent has a formal job, whereas when the agreement is a fixed amount of money, we assume he has an informal job.

Table 1.10 shows the marginal effects of regression (3). We find no gender effect on the likelihood of settlement, regardless of whether one or both parties attended. Moreover, in stark contrast with figures from discrimination cases suggested by Knepper (2018) where only 5% of cases reach litigation, in our data-set over 70% of cases failed to settle and proceeded to litigation. Interestingly, Table 1.10 shows that when the respondent has a formal job the likelihood of settlement decreases. Although the evidence indicates that the judge's gender does not affect the likelihood of settlement (extensive margin), this raises a concern about the potential effect of the judge's gender on the level of agreement (intensive margin) for cases that settle.

In light of that result, we also test the impact of the judge's gender on the level of agreement for cases that settle as follows:

$$(4) \qquad Agreement_{ij} = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 Formal_i + \gamma_d + \gamma_t + \varepsilon_{ij}$$

Where $\boldsymbol{Agreement_{ij}}$ is the level of agreement (either in fixed amount of money or in percentage terms) for case $\boldsymbol{i}$ assigned to judge $\boldsymbol{j}$ that settles in the settlement hearing. $\boldsymbol{Female_{j(i)}}$ is an indicator variable for whether case $\boldsymbol{i}$ was assigned to a female judge $\boldsymbol{i}$, $\boldsymbol{N_i^T}$ is the number of children involved in case $\boldsymbol{i}$ and $\boldsymbol{Formal_i}$ is an indicator variable for whether the respondent has a formal job. Finally, $\boldsymbol{\gamma_d}$ and $\boldsymbol{\gamma_t}$ are district and year fixed effects, respectively.

Table 1.11 shows the estimates of equation (4). Again, we do not find a significant effect of the judge's gender on the level of agreement for cases that settle. This means that there is no evidence that cases would self-select into litigation depending on the judge's gender. Thus, we provide strong evidence that the litigation stage in child support cases is suitable for detecting gender-based differences.

1.6 Understanding the effect of incomplete information

The main result from the previous section is that the gender gap in formal cases is around 56% bigger than in informal cases. We propose a simple model of how incomplete information shrinks gender differences in child support decisions and emphasise beliefs about income as the mechanism. We elicit these beliefs based on the methodological premise that the pool of judges who are making decisions in informal cases are revealing their preferences in formal cases during the period of analysis.

1.6.1 Conceptual Framework of Judicial Decision-making

For the sake of simplicity, let us assume a situation in which one female judge and one male judge have to make their decisions in trials where there is only one child involved, and where the respondent supports no other children. A case $\boldsymbol{i}$ is randomly assigned to a judge with gender $\boldsymbol{g = m,f}$ who sets an award to be paid by the respondent. There are two types of cases, as mentioned in the previous section: the formal case $\boldsymbol{i} \in \boldsymbol{F}$ and the informal case $\boldsymbol{i} \in \boldsymbol{I}$.

In a formal case, a judge with gender $\boldsymbol{g}$ observes $\boldsymbol{y_i}$- the respondent's income in the formal case $\boldsymbol{i} \in \boldsymbol{F}$ - and awards a monthly percentage deduction from the respondent’s salary $\boldsymbol{\alpha^g} \in (\boldsymbol{0}, \boldsymbol{0.6}]$. Note that we allow $\boldsymbol{\alpha^g}$ to vary between gender $\boldsymbol{g}$ and that it is constant within gender $\boldsymbol{g}$ since it does not depend on any case-specific characteristic. Let us reiterate that, legally, the judge only needs three pieces of information to decide on the case: the number of children who are trial matter, the respondent’s income, and any additional children the respondent has to support. Given that the judge is able to observe $\boldsymbol{y_i}$ and that, by assumption, trials only involve one child and no other children to support, $\boldsymbol{\alpha^g}$ is the fraction that the judge considers fair to deduct from the respondent’s salary to support one child who is trial matter. Thus, the fraction of income per child is our outcome of interest in formal cases and what we define as the judge’s allocation preferences.

In an informal case, on the other hand, a judge with gender $\boldsymbol{g}$ does not observe $\boldsymbol{y_i}$ - the respondent's income in the informal case $\boldsymbol{i} \in \boldsymbol{I}$ - and this time sets an award $\boldsymbol{A_i^g}, \boldsymbol{i} \in \boldsymbol{I}$ to be paid monthly by the respondent. To do so, the judge forms a belief about the respondent's income $\boldsymbol{b_i^g}, \boldsymbol{i} \in \boldsymbol{I}$ and sets a percentage $\boldsymbol{\alpha_i^g}, \boldsymbol{i} \in \boldsymbol{I}$ to be deducted from it. It is worth noting that neither $\boldsymbol{b_i^g}$ nor $\boldsymbol{\alpha_i^g}$ are observed in $\boldsymbol{i} \in \boldsymbol{I}$. We only observe the given award $\boldsymbol{A_i^g}$. Thus, a judge with gender $\boldsymbol{g}$ will choose an award $\boldsymbol{A_i^g}$ in an informal case $\boldsymbol{i} \in \boldsymbol{I}$ as follows:

(5) $$A_i^g = \alpha^g b_i^g \quad \forall i \in I$$

Again, the main outcome of interest is the award per child in an informal case. This is implicit since we assume there is only one child involved in the trial and no other children to support. Thus, under this framework, a judge with gender *g* grants an award that is constituted by their allocation preferences $\boldsymbol{\alpha^g}$ and their beliefs about the respondent’s income $\boldsymbol{b_i^g}$.

This framework allows for interpreting the gender-based gaps in judicial decision-making both in formal and informal cases. In the formal cases, the gap is straightforward to calculate since the award $\boldsymbol{\alpha^g}$ only varies between gender $\boldsymbol{g}$ and does not depend on case-specific characteristics:

(6) $$GAP_{formal} = \frac{\alpha^f}{\alpha^m}\,, \qquad i \in F$$

If the gap in equation (6) were less than 1, it would mean that female judges have preferences for lower shares of income to be allocated to a child than their male counterparts in formal cases. As can be seen, this gap is entirely driven by differences in allocation preferences between genders.

In the informal cases, however, the gender-based gap would take the following expression:

(7) $$GAP_{informal} = \left(\frac{\alpha^f}{\alpha^m}\right)\left(\frac{b^{-f}}{b^{-m}}\right), \quad i \in I$$

If the gap in equation (7) were less than 1, it would mean that female judges are more lenient than their male counterparts towards the respondents in informal cases. However, the gap encompasses two different gaps: the gap in allocation preferences and the gap in beliefs about the respondents' income. We use this framework to interpret the results shown in section 1.4.2. Moreover, this framework is the starting point in understanding how judges' gender-based differences in beliefs might expand or shrink the gender-based gap in allocation preferences under some assumptions, as we explain in the next section.

1.6.2 Methodology

Intuitively, this methodology uses the revealed allocation preferences $\boldsymbol{\alpha_{ij}}$ in the formal cases $\boldsymbol{i} \in \boldsymbol{F}$ assigned to judge $\boldsymbol{j}$ to infer judge's $\boldsymbol{j}$ beliefs about the income of respondents in informal cases $\boldsymbol{i} \in \boldsymbol{I}$. To illustrate, if the award given by a judge in a formal case represents 30% of the respondent's income and awards S/. 300 in a similar but informal case, we can then infer that the judge believes the respondent earns S/. 1,000 by assuming that the judge should maintain, *ceteris paribus*, the same allocation preference (30%). Thus, the award in informal cases breaks down into allocation preferences and beliefs about the respondent's income and we inspect the role of each of these factors in determining the gender-based gap.

Let us remember the gender-based gap expression in informal cases under the simple framework we developed in Section 1.6.1:

$$GAP_{informal} = \left(\frac{\alpha^f}{\alpha^m}\right)\left(\frac{b^{-f}}{b^{-m}}\right), \quad i \in I$$

Equation (7) computes the gender-based gap in informal cases $\boldsymbol{I}$ based on the given awards $\boldsymbol{A_i^g}$ which are observable. However, the factors of that decision, $\boldsymbol{b_i^g}$ and $\boldsymbol{\alpha_i^g}$, are unobserved by the researcher. If those variables were observable, we could inspect the role of uncertainty in shaping the gap in informal cases $\boldsymbol{I}$, by taking log of equation (7).

(8) $$\log(GAP_{informal}) = \log\left(\frac{\alpha^f}{\alpha^m}\right) + \log\left(\frac{b^{-f}}{b^{-m}}\right)$$

If the gap in allocation preferences were less than 1 but the gap in beliefs more than 1, then we could conclude that incomplete information would attenuate the gap in allocation preferences. To conduct such an analysis, we propose the next methodology to calibrate $\boldsymbol{\alpha_i^g}$ in order to estimate the parameter $\boldsymbol{b_i^g}$.

We first calibrate the allocation preferences from the decisions made by judges in formal cases $\boldsymbol{i} \in \boldsymbol{F}$ as follows:

(9) $$\frac{\alpha_{ij}}{N_i^T} = \mu_j + \beta' X_{ij} + \varepsilon_{ij}, \quad i \in F$$

Where $\boldsymbol{\alpha_{ij}}$ is the award given by judge $\mathbf{j}$ in formal case $\boldsymbol{i} \in \boldsymbol{F}$; $\boldsymbol{X_{ij}}$ is a vector that contains the main criteria for award-giving in child support cases (number of children in trial and additional children in need of support by the respondent), and $\boldsymbol{\mu_j}$ is the judge fixed-effect.

Second, we infer how those same judges would have decided the award $\boldsymbol{\tilde{\alpha}_{ij}}$ (as a percentage of income to be deducted) in informal cases. By using the coefficients from regression (9), we predict the allocation preferences $\boldsymbol{\tilde{\alpha}_{ij}}$ in informal cases as follows:

(10) $$\tilde{\alpha}_{ij} = (\hat{\mu}_j + \hat{\beta} X_{ij}) N_i^T, \quad i \in I$$

Where $\widetilde{\boldsymbol{\alpha}}_{\boldsymbol{ij}}$ is the calibrated award made by judge $\boldsymbol{j}$ in informal case $\boldsymbol{i} \in \boldsymbol{I}$ as a percentage of the respondent's income; vector $\boldsymbol{X}_{\boldsymbol{ij}}$ contains the same set of variables as in equation (9) but for informal case $\boldsymbol{i} \in \boldsymbol{I}$; $\widehat{\boldsymbol{\beta}}$ is the vector of coefficients estimated in equation (9), and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{j}}$ is the estimated judge fixed effect also taken from equation (9). Note that the total estimated award is the multiplication of the calibrated award per child and the number of children involved in the informal trial. Figure 1.3 shows the kernel densities of the calibrated awards in informal cases by judge's gender. It can be seen that awards vary from 0.1 to 0.5 in general and that distributions seem to have the same variance but not the same mean: the first moment for male judges might be higher than for female judges.

The third and last step is to estimate the judge's belief about the respondent's income $\boldsymbol{b}_{\boldsymbol{i}}$ in informal case $\boldsymbol{i} \in \boldsymbol{I}$ by combining the calibrated award $\widetilde{\boldsymbol{\alpha}}_{\boldsymbol{ij}}$ (as percentage of income) and the observed award $\boldsymbol{A}_{\boldsymbol{ij}}$ (fixed amount of money) in informal case $\boldsymbol{i} \in \boldsymbol{I}$:

$$(11) \qquad \hat{b}_{ij} = \frac{A_{ij}}{\widetilde{\alpha}_{ij}}\,, \qquad i \in I$$

Figure 1.4 shows the kernel distributions of estimates by judge's gender. As opposed to Figure 1.3, the mean and the variance of distribution of estimates seem to vary by judge's gender. Note that in all these estimations we are not interested in the effect that the judge's gender has on the awards: the judge's time-invariant characteristics (such as gender) are captured by the judge fixed effect $\widehat{\boldsymbol{\mu}}_{\boldsymbol{j}}$. Instead, our focus is on modelling the award in formal cases to predict how judges would have awarded a percentage deduction in informal cases.

1.6.3 Econometric specifications and results

Given the inputs provided by the expressions (10) and (11), we decompose the awards in informal cases $\boldsymbol{A}_{\boldsymbol{ij}}$ into allocation preferences $\widetilde{\boldsymbol{\alpha}}_{\boldsymbol{ij}}$ and beliefs $\widehat{\boldsymbol{b}}_{\boldsymbol{ij}}$. Then we calculate the gender-based gap in both dimensions to measure the relative contribution of both sources of variation to the total gap in awards under incomplete information.

$$(12) \qquad \log\left(\frac{A_{ij}}{N_i^T}\right) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \varepsilon_{ij}\,, \qquad i \in I$$

$$(13) \qquad \log\left(\frac{\tilde{\alpha}_{ij}}{N_i^T}\right) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \varepsilon_{ij}\,, \qquad i \in I$$

$$(14) \qquad \log \hat{b}_{ij} = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \varepsilon_{ij}\,, \qquad i \in I$$

The main coefficient of interest $\boldsymbol{\beta_1}$ in equation (13) estimates the difference in the average allocation preference displayed by female versus male judges, while the main coefficient of interest $\boldsymbol{\beta_1}$ in equation (14) estimates the difference in the average estimate of the respondents' income of female versus male judges for observably similar respondents. The results are in Table 1.12.

As can be seen in Column 1, the observed gap suggests that female judges are 4.7% more lenient towards respondents. Interestingly, however, the gap is bigger if we only take into account allocation preferences (12%, as shown in Column 2). This suggests that incomplete information about the respondent's income attenuates the gender gap in allocation preferences. Indeed, female and male judges respond differently to incomplete information: the former (compared to the latter) estimate that the income of respondents is 7.3% higher on average.

These estimates, moreover, are not driven by differences in specific case characteristics between genders. As it was shown in the balance tests (Section 1.4.1), these are balanced between male and female judges. In particular, we look at the claim of petitioners as a possible source of variation. Figure 1.3 presents the kernel distributions of estimates by the judge's gender which show that the first and second moments do not differ by the judges' gender. In the next section, we develop a simple framework to investigate factors that could drive this result.

1.6.4 Estimates of beliefs

In this section, we propose a simple framework to estimate why male and female judges form different beliefs about the income of respondents. The starting point is based on the fact that randomisation of cases ensures that male and female judges decide on cases with the same set of (incomplete) information about cases' characteristics when respondents work in the informal sector. Therefore, it must be the case that male and female judges process the signals of the unknown income of respondents differently. In particular, we focus on the claim (the amount of money requested from the respondent by the petitioner) which is the main signal given by petitioners to judges on this regard. In fact, when setting their claims, petitioners argue that the amount asked corresponds to the level of income of respondents. Thus, we investigate to what extent male and female judges rely on this piece of information to estimate the unknown income.

This framework is based on the standard Bayesian updating model in which judges form beliefs about the respondent's unknown income ***Y*** based on a prior belief and a signal (the petitioner's claim amount ***C***). However, we introduce one important feature to this simple model: the fact that petitioners exaggerate their claims to signal that the respondent has higher disposable income in order to increase the chances of obtaining higher awards. It is important to note that this signal, by definition, cannot depend on the gender of judges since it is set by the petitioner before the randomisation of the case. Thus, to create different income predictions based on the gender of judges, we allow the prior of judges about the unknown income to depend on their gender.

1.6.4.1 The model

We assume that the judge with gender $\boldsymbol{g}$ does not know the true respondent's income $\boldsymbol{y} = \mathbf{ln}\ (\boldsymbol{Y})$ but has a prior $\boldsymbol{y_0^g}$ and a fixed variance $(\boldsymbol{\sigma_0^g})^2$ of the prior about $\boldsymbol{y}$:

(15) $$y = y_0^g + \sigma_0^g \delta \qquad \delta \sim N(0,1)$$

While the petitioner's claim amount $\boldsymbol{c} = \mathbf{ln}\ (\boldsymbol{C})$ is an upward biased signal of the unknown income in the following fashion:

(16) $$c = By + \sigma\varepsilon \qquad \varepsilon \sim N(0,1), B > 1$$

Since the randomisation of cases occurs after the respondent makes the claim, the exaggeration rate $\boldsymbol{B}$ cannot depend on the gender of judges. Also, note that, as a simplification, the exaggeration rate $\boldsymbol{B}$ is constant and must be more than 1 to reflect the fact that the signal is an inflated version of the true income. Since the judge knows the claim has an exaggeration rate $\boldsymbol{B}$, the judge cares about the deflated signal $\tilde{\boldsymbol{c}} = \frac{\boldsymbol{c}}{\boldsymbol{B}}$. Thus, a more intuitive way of writing down equation (16) is:

(17) $$\tilde{c} = y + \frac{\sigma}{B}\varepsilon \qquad \varepsilon \sim N(0,1), B > 1$$

Since $(\boldsymbol{y}, \tilde{\boldsymbol{c}})$ is distributed according to a bivariate Gaussian distribution, we pin down the judge's belief formation (posterior) based on the prior mean and the petitioner's deflated signal of the unknown income as follows:

(18) $$E(y|\tilde{c}) = E(y) + \frac{Cov(y,\tilde{c})}{Var(\tilde{c})}\big(\tilde{c} - E(\tilde{c})\big)$$

Working with equations (15) and (17) to find the elements of equation (18):

(19) $$E(y) = y_0^g, E(\tilde{c}) = y_0^g, Var(\tilde{c}) = (\sigma_0^g)^2 + \frac{\sigma^2}{B^2}, Cov(y,c) = \ (\sigma_0^g)^2$$

Plugging these results in equation (18):

(20) $$E(y|\tilde{c}) = y_0^g + \frac{(\sigma_0^g)^2}{(\sigma_0^g)^2 + \frac{\sigma^2}{B^2}}\ (\tilde{c} - y_0^g)$$

Equation (20) shows that if the deflated signal $\tilde{\boldsymbol{c}}$ exceeds the judge's prior $\boldsymbol{y_0^g}$, the judge's guess is adjusted upwards. Conversely, if the deflated claim $\tilde{\boldsymbol{c}}$ is lower than the judge's prior $\boldsymbol{y_0^g}$, the judge's guess is adjusted downwards. In other words, given the judge's prior $\boldsymbol{y_0^g}$, the updating direction depends on whether the claim $\boldsymbol{c}$ exceeds or is less than the threshold $\boldsymbol{By_0^g}$.

1.6.4.2 Estimations and results

To estimate parameters of this model, we replace $\tilde{\boldsymbol{c}}$ in terms of $\boldsymbol{c}$ in equation (20) and we obtain the following:

(21) $$E(y|c) = \theta y_0^g + \omega c$$

Where $\theta = \frac{\sigma^2}{(\sigma_0^g)^2 B^2+\sigma^2}$ and $\omega = \frac{(\sigma_0^g)^2 B}{(\sigma_0^g)^2 B^2+\sigma^2}$

In these expressions, the weights $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ on $\boldsymbol{y_0^g}$ and $\boldsymbol{c}$, respectively, can be interpreted in terms of the variance (or precision) of the judge's prior about $\boldsymbol{y}$, relative to the variance (or precision) of the petitioner's signal about $\boldsymbol{y}$. For instance, the more accurate or precise the petitioner's signal (i.e. the lower $\boldsymbol{\sigma}$ is), the greater is $\boldsymbol{\omega}$. However, it is important to note that the exaggeration rate affects these weights through $\boldsymbol{\sigma_0^g}$: *B* expands the negative effect of $\boldsymbol{\sigma_0^g}$ on $\boldsymbol{\theta}$ and shrinks the positive effect of $\boldsymbol{\sigma_0^g}$ on $\boldsymbol{\omega}$.

Since the parameter $\boldsymbol{\sigma}$ allows for some noise in the claim as a signal of income and cannot depend on the gender of judges, we set it equal to 1 as a simplification and focus on the comparison of $\boldsymbol{y_0^g}$ and $\boldsymbol{\sigma_0^g}$: across gender given a constant exaggeration rate ***B***. Thus, we estimate equation (21) separately for male and female judges as follows:

$$\ln(\hat{b}_{ij}) = \beta_1 + \beta_2 \ln(C_i) + \gamma_d + \gamma_t + \varepsilon_{ij} \tag{22}$$

Where $\boldsymbol{\hat{b}_{ij}}$ is the belief set by judge ***j*** about the income of respondent-case ***i*** estimated from equation (11), $\boldsymbol{C_i}$ is the claim amount made by the petitioner in case ***i***. In addition, both equations include district $\boldsymbol{\gamma_d}$ and year $\boldsymbol{\gamma_t}$ fixed effects. Thus, we interpret $\boldsymbol{\hat{\beta}_1}$ and $\boldsymbol{\hat{\beta}_2}$ in equation (22) as $\boldsymbol{\theta y_0^g}$ and $\boldsymbol{\omega}$ from equation (21), respectively. By assuming that $\boldsymbol{\sigma = 1}$, we recover the parameters of interest $\boldsymbol{\sigma_0^g}$ and $\boldsymbol{y_0^g}$:

$$y_0^g = \frac{\hat{\beta}_1 B}{B-\hat{\beta}_2 B^2} \tag{23}$$

$$(\sigma_0^g)^2 = \frac{\hat{\beta}_2 B}{B-\hat{\beta}_2 B^2} \tag{24}$$

Table 1.13 shows the estimation of equation (22) for each gender separately. It can be seen that the estimates differ by the gender of judges. Replacing $\boldsymbol{\hat{\beta}_1}$ and $\boldsymbol{\hat{\beta}_2}$ for each gender in the previous expressions, it can be shown that $\boldsymbol{y_0^m} > \boldsymbol{y_0^f}$ and $\boldsymbol{\sigma_0^m} > \boldsymbol{\sigma_0^f}$ for any value of $\boldsymbol{B} \in$

$(\mathbf{1}, \mathbf{3})$[13]. These results suggest that, although female judges have a lower prior mean, the associated weight on the prior mean is higher because of the higher precision (lower $\boldsymbol{\sigma_0^g}$) of the prior mean in comparison to male judges. Regarding the weight on the signal, female judges rely less on the petitioner's claim because the precision of the signal relative to the precision of prior mean, $\frac{\boldsymbol{\sigma}}{\boldsymbol{\sigma_0^g}}$, is lower for female judges (given the higher precision of the prior mean) which decreases the weight on the signal in comparison to male judges.

1.7 Discussion

The literature revised in section 1.1 on gender differences in decision making would predict clear differences also in the judicial context and, concretely, evidence of homophily when there is enough discretion: female judges tending to benefit female plaintiffs/defendants. This is the reason why the results found in this study, that female judges seem to exhibit less leniency towards females than their male counterparts, might seem surprising at first. Even though there is not enough data to test for additional hypotheses, we believe a potential explanation lies at the intra-gender interaction between the judge and the female petitioner. For women, becoming a judge means having to overcome a system that has excluded them historically. In the US, for example, only after 140 years after the establishment of the federal court was a woman appointed to the federal bench. The same can easily be said for judicial systems around the world. Thus, women in the profession might have to dissociate themselves from their gender to thrive in male-dominated contexts, become Queen Bees (Derks et al., 2011) and overcompensate their 'masculinity' by being less lenient towards their own gender.

This decoupling from expected gender norms might also make intra-gender interactions more difficult, insofar as the traditional kind of femininity is interpreted as the 'wrong kind' of femininity (S. Mavin & Grandy, 2012). Female judges, both as being influenced by a mainly male-oriented context (S. A. Mavin et al., 2014), and also being forced to 'masculinise' to be successful in the profession, could perceive a more traditional woman (stay-at-home mother, who does not work and takes care of the children, and is asking in court for the father to be the provider) as the 'wrong kind', and thus 'punish' her for this.

On the same page, it is also possible that female petitioners are simply experiencing what is known as 'female misogyny'. This concept is thought of as the act of women subjugating,

[13] An exaggeration rate *B* larger than 3 would generate negative values of mean and variance priors of judges.

repressing, undermining, excluding or stigmatising other women (S. Mavin, 2008), and arises -theoretically - when a female subverts the gendered order. For female judges, this order could be interpreted as the 'masculinised' profession. The 'typical' female would then be subverting it and thus becomes discriminated against based on her non-conformity.

Unfortunately, due to the nature of our dataset, testing hypotheses based on this theoretical framework has not been possible. Further, even if our data allowed it, we believe experimental evidence (in the lab or in the field) is better suited to address these conjectures. Further research is necessary to understand the nuances related to the mechanisms behind the gender differences in judicial decision-making found in this study.

1.8 Conclusions

In this article, we inspect whether there are gender differences when child support is decided by a court rather than by negotiation between the parents. By exploiting random assignment of cases to judges, we find that female judges decide on a lower allocation of child support than male judges do in formal and informal cases. Moreover, we find that the effect of assigning a female judge to a formal case is 57.3% stronger than for an informal case.

To understand this result we use a simple model of incomplete information. We assume that a judge has the same preferences for child support allocation regardless of whether a case is formal or informal. By estimating this component when the judge makes decisions in formal cases, we are able to calibrate their preferences in informal cases and infer the degree to which the amount of child support they allocate in the latter is influenced by their beliefs about the income of respondents. We find that, relative to male judges, female judges infer that the respondent has higher levels of income when they cannot observe it during trial, explaining why the gender gap in informal cases is smaller than in formal cases. By using a simple Bayesian updating framework in which judges form beliefs about the unknown income based on their priors and the signal (claim) sent by the petitioner, we provide a possible explanation for this fact: data shows that female judges rely less on the signal sent by the petitioner and put more weight on their priors. These findings highlight the fact that information asymmetries might play a role in influencing the outcomes of different types of judicial settings. For instance, could the lack of information explain racial disparities in the outcomes of criminal cases?

Finally, the evidence found in this chapter has vital policy implications. There is evidence that parents transport less economic resources after parental separation (Björklund &

Sundström, 2006). For example, since the father has reduced access to the child, he has less incentives to provide resources. Further, if the mother remarries, the father has fewer incentives to support his child because part of the transfer spills over to the new husband (Chiappori & Weiss, 2007). Hence, child support allocation is not a trivial matter, so a discussion about the predictability of the judicial system in these types of cases is necessary, given that verdicts depend so much on variables such as the gender of the judge or the lack of information during trial. A potential solution to reduce discretion in child support cases could be for judges to rely on benchmarks based on, for instance, the type of respondent's job or the cost of living of the district where the child resides. This is an important issue for further research.

Table 1.1: Sample characteristics

	All cases (i)	Male judges (ii)	Female judges (iii)
Panel A. Hearings			
Settlement (%)	27.4	24.5	29.7
Respondent is formal (%)	20.2	20.8	19.6
Number of judges	149	61	88
Observations	2371	1061	1310
Panel B. Litigations			
Respondent is formal (%)	22.7	22.2	21.3
Number of judges	153	59	94
Observations	1736	856	880

Note*:* This table describes samples corresponding to the two stages (hearing and litigation) of child support cases. Samples contain cases filed in districts with at least one court led by a male judge and one court led by a female judge. *Settlement* indicates the proportion of hearings that reach settlement. *Respondent is formal* indicates the proportion of cases in which the respondent has a formal job.

Table 1.2: Balance Table, Case Characteristics by Judges' Gender (Hearing)

	Male Judge		Female Judge		Difference	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**p**
Number of children (in trial)	1.33	0.024	1.32	0.019	0.007	0.816
Petitioner attends (%)	97.8	0.5	98.2	0.4	-0.4	0.591
Petitioner's attorney attends (%)	67.7	2	64	1.7	3.7	0.168
Respondent attends (%)	64.6	1.7	65.2	1.6	-0.6	0.785
Respondent's attorney attends (%)	35.3	1.7	35.9	1.7	-0.6	0.823
Respondent is rebel (%)	49.2	2.6	49.8	2.8	-0.6	0.881
Respondent is formal	20.8	1.6	19.6	1.4	1.2	0.582
Observations	1061		1310		2371	
Number of judges	61		88		149	

<u>Note</u>: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents a balance table of cases' characteristics in the hearing stage. The value for t-tests are p-values of the difference across groups. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Petitioner attends* indicates the proportion of hearings in which the petitioner is present. *Petitioner's attorney attends* indicates the proportion of hearings in which the petitioner's attorney is present. *Respondent attends* indicates the proportion of hearings in which the respondent is present. *Respondent's attorney attends* indicates the proportion of hearings in which the respondent's attorney is present. *Respondent is rebel* indicates the proportion of hearings in which the respondent is declared a 'rebel'. *Respondent is formal* indicates the proportion of hearings in which the respondent has a formal job.

Table 1.3: Balance Table, Case Characteristics by Judges' Gender (Litigation - Formal)

	Male Judge		Female Judge		Difference	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**p**
Number of children (in trial)	1.38	0.048	1.35	0.019	0.029	0.644
Number of children (off trial)	0.32	0.041	0.489	0.4	-0.171	0.079*
Claim (%)	55.5	0.6	56	1.7	-0.5	0.543
Petitioner reports respondent's income (%)	48	5.3	51.1	1.6	-3.1	0.658
Respondent reports his income (%)	43.9	4.3	46.6	1.7	-2.7	0.658
Petitioner has assets (%)	0	0	1.1	2.8	-1.1	0.318
Respondent has assets (%)	4	1.6	5.7	1.4	-1.7	0.517
Observations	173		176		349	
Number of judges	43		57		100	

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents a balance table of cases' characteristics for cases where the respondent has a formal job. Sample restricted to cases held at districts where there is at least one female and one male judge and that reached the litigation stage. The value for t-tests are p-values of the difference across groups. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Number of children (off trial)* is a variable that contains the additional number of children the respondent must take care of. *Claim* is the amount of child support claimed by the petitioner at the beginning of the process (as a fraction of the respondent's income). *Petitioner reports respondent's income* indicates the proportion of petitioners in the sample who give information about the income of respondents. *Respondent reports his income* indicates the proportion of respondents who give information about their income. *Petitioner has assets* indicates the proportion of cases for which there is information of whether the petitioner has any assets. *Respondent has assets* indicates the proportion of cases for which there is information of whether the respondent has any assets.

Table 1.4: Balance Table, Case Characteristics by Judges' Gender (Litigation - Informal)

	Male Judge		Female Judge		Difference	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**p**
Number of children (in trial)	1.38	0.033	1.33	0.025	0.009	0.810
Number of children (off trial)	0.36	0.045	0.34	0.029	0.021	0.702
Claim (PEN)	1425.5	97.6	1340.7	57.4	84.8	0.454
Petitioner reports respondent's income (%)	53.8	3.2	52.8	3.9	1	0.836
Respondent reports his income (%)	44.3	3.1	45	3.1	-0.7	0.866
Petitioner has assets (%)	1	0.4	1.2	0.5	-0.2	0.722
Respondent has assets (%)	6.8	1.3	7.8	0.9	-1	0.536
Observations	585		642		1227	
Number of judges	50		74		124	

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents a balance table of cases' characteristics for cases where the respondent has an informal job. Sample restricted to cases held at districts where there is at least one female and one male judge and that reached the litigation stage. The value for t-tests are p-values of the difference across groups. *Number of children (in trial*) is a variable that contains the number of children involved in the trial. *Number of children (off trial)* is a variable that contains the additional number of children the respondent must take care of. *Claim* is the amount of child support claimed by the petitioner at the beginning of the process (in PEN). *Petitioner reports respondent's income* indicates the proportion of petitioners in the sample who give information about the income of respondents. *Respondent reports his income* indicates the proportion of respondents who give information about their income. *Petitioner has assets* indicates the proportion of cases for which there is information of whether the petitioner has any assets. *Respondent has assets* indicates the proportion of cases for which there is information of whether the respondent has any assets.

Table 1.5: Judges' Gender Effects on Child Support Decisions – pooled OLS estimates

	Log (award per child)		Z-score (award per child)	
	Formal (i)	**Informal (ii)**	**Formal (iii)**	**Informal (iv)**
Female judge	-0.068**	-0.059**	-0.247**	-0.157***
	(0.033)	(0.027)	(0.117)	(0.060)
Number of children (in trial)	-0.343***	-0.291***	-0.997***	-0.357***
	(0.019)	(0.019)	(0.058)	(0.037)
Number of children (off trial)	-0.169***	-0.098***	-0.490***	-0.146***
	(0.011)	(0.016)	(0.043)	(0.028)
Observations	349	1227	349	1227
Number of judges	100	124	100	124
R^2	0.616	0.383	0.54	0.239

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents the pooled OLS estimates of the effects of gender on child support decisions. Columns (i) and (ii) use the log of the award per child as a dependent variable. Columns (iii) and (iv) use the standardised award per child as a dependent variable. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Number of children (off trial)* is a variable that contains the additional number of children the respondent must take care of. Each regression includes district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Table 1.6: Balance Table, Judges' Characteristics by Gender

	Male Judge		Female Judge		Difference	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**p**
Age (years)	43.571	1.105	42.64	0.857	0.931	0.503
Judge as principal (%)	49.2	0.066	35.2	0.051	14	0.093*
Experience as principal	6.379	1.026	5.871	0.913	0.508	0.712
Wealth (standardised)	0.061	0.118	-0.04	0.114	0.101	0.55
Observations	59		88		146	

<u>Note</u>: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table shows the balance test for all characteristics of judges available. The value for t-tests are p-values of the difference across groups. *Age* is a variable that describes how old (in years) the judge is at the time of the trial. *Judge is principal* indicates the proportion of male and female judges that are a principal judge at the time of the trial. *Experience as principal* is a variable that indicates for how long (in years) the judge has been a principal judge at the time of the trial. *Wealth* is a standardised measure of the wealth amounts (in local currency) reported by the judges.

Table 1.7: Judges' Gender Effects on Child Support Decisions with all Judges' Characteristics - pooled OLS Estimates

	Z-score (award per child)			
	Formal		Informal	
	(i)	(ii)	(iii)	(iv)
Female judge	-0.247***	-0.261**	-0.157***	-0.161**
	(0.117)	(0.124)	(0.059)	(0.069)
Age (years)		-0.003		-0.013**
		(0.008)		(0.007))
Judge is principal		0.110		-0.287***
		(0.114)		(0.106)
Experience as principal (years)		-0.005		0.004
		(0.017)		(0.013)
Wealth (standardised)		0.035		-0.114**
		(0.039)		(0.054)
Observations	349	334	1227	1172
Number of judges	100	98	124	119
R^2	0.540	0.554	0.239	0.256

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents pooled OLS estimates of the standardised gender gap in child support decisions when including all judges' characteristics as controls. Columns (i) and (ii) use the standardised award per child as a dependent variable in formal cases without and with additional judges' characteristics as covariates, respectively. Columns (iii) and (iv) use the standardised award per child as a dependent variable in informal cases without and with additional judges' characteristics as covariates, respectively. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Age* is a variable that describes how old (in years) the judge is at the time of the trial. *Judge is principal* is an indicator variable for whether the judge is a principal judge or not. *Experience as principal* is a variable that indicates for how long (in years) the judge has been a principal judge at the time of the trial. *Wealth* is a standardised measure of the wealth amounts (in local currency) reported by the judges. District and year fixed effects are also included. Standard errors in parentheses are clustered at the judge level.

Table 1.8: Judges' Gender Effects on Child Support Decisions with Children's Characteristics – pooled OLS Estimates

	Log (award per child)		Z-score (award per child)	
	Formal (i)	Informal (ii)	Formal (iii)	Informal (iv)
Female judge	-0.064*	-0.066**	-0.237*	-0.182***
	(0.037)	(0.031)	(0.131)	(0.068)
Number of children (in trial)	-0.364***	-0.307***	-1.042***	-0.365***
	(0.025)	(0.022)	(0.073)	(0.055)
Number of children (off trial)	-0.169***	-0.104***	-0.492***	-0.157***
	(0.012)	(0.017)	(0.047)	(0.028)
Health issues	0.024	-0.031	0.131	0.015
	(0.052)	(0.031)	(0.198)	(0.069)
<u>Sex of children in trial</u>				
1 in 4 children is female	0.336***	0.231	0.988***	0.444**
	(0.105)	(0.222)	(0.257)	(0.193)
Half of children is female	-0.043	0.003	-0.174	-0.033
	(0.038)	(0.039)	(0.114)	(0.093)
3 in 4 children is female	0.288***	0.084	0.840***	0.400**
	(0.087)	(0.118)	(0.280)	(0.169)
Every child is female	-0.046	-0.005	-0.148	0.027
	(0.028)	(0.027)	(0.101)	(0.078)
Other combinations	0.097*	0.031	0.236	-0.046
	(0.058)	(0.078)	(0.153)	(0.197)
Average age of child	-0.001	0.004**	-0.008	0.007
	(0.003)	(0.002)	(0.009)	(0.006)
Observations	328	1151	328	1151
Number of judges	99	121	99	121
R^2	0.623	0.379	0.552	0.23

<u>Note</u>: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents the pooled OLS estimates of the effects of gender on child support decisions adding children's characteristics as covariates. Columns (i) and (ii) use the log of the award per child as a dependent variable. Columns (iii) and (iv) use the standardised award per child as a dependent variable. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Number of children (off trial)* is a variable that contains the additional number of children the respondent must take care of. *Child has health issues* is an indicator variable for whether any of the children involved in the trial exhibits a serious health issue. *Sex of children in trial* is a categorical variable that measures the proportion of female children out of the total of children in trial. The base category is *Every child is male*. *Average age of child* Each regression includes district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Table 1.9: Judges' Gender Effects on Child Support Decisions – pooled OLS Estimates no additional children to support

	Log (award per child)		Z-score (award per child)	
	Formal (i)	**Informal (ii)**	**Formal (iii)**	**Informal (iv)**
Female judge	-0.072*	-0.055*	-0.267*	-0.172**
	(0.041)	(0.031)	(0.149)	(0.074)
Number of children (in trial)	-0.327***	-0.282***	-0.981***	-0.354***
	(0.021)	(0.019)	(0.065)	(0.043)
Observations	252	953	252	953
Number of judges	92	116	92	116
R^2	0.589	0.402	0.506	0.245

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table presents the pooled OLS estimates of the effects of gender on child support decisions when there are no additional children to support. Columns (i) and (ii) use the log of the award per child as a dependent variable. Columns (iii) and (iv) use the standardised award per child as a dependent variable. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. Each regression includes district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Table 1.10: Marginal effects for settlement – probit estimates

	Likelihood of settlement in hearings					
	All cases			Both parties attended		
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Female judge	0.074	0.078	0.102	0.076	0.080	0.100
	(0.083)	(0.082)	(0.086)	(0.116)	(0.116)	(0.120)
Number of children (in trial)		-0.062	-0.053		-0.052	-0.036
		(0.041)	(0.043)		(0.049)	(0.051)
Respondent is formal			-0.110*			-0.241***
			(0.064)			(0.076)
Observations	2404	2391	2253	1520	1510	1419
Number of judges	0.023	0.024	0.027	0.045	0.045	0.052
R^2	147	147	142	134	134	129

<u>Note</u>: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table shows the marginal effects of case characteristics on the likelihood of settlement. Columns (i) to (iii) use the whole set of cases. Columns (iv) to (vi) use the restricted set of cases where both parties attended the hearings. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Respondent is formal* is an indicator variable for whether the respondent has a formal job. Each regression controls for district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Table 1.11: Judges' gender effects on agreed amount of child support – pooled OLS estimates

	Agreed amount	
	(i)	**(ii)**
Female judge	-24.95	-24.04
	(27.32)	(28.19)
Number of children (in trial)		115.9***
		(19.79)
Respondent is formal	-514	-526.9***
	(17.90)	(19.30)
Observations	653	651
Number of judges	104	104
R^2	0.343	0.383

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table shows pooled OLS estimates of the effects of the gender gap and other case characteristics on agreed amounts of child support. The dependent variable is measured as the logarithm of the agreed amount of child support per child. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Respondent is formal* is an indicator variable for whether the respondent has a formal job. Each regression controls for district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Table 1.12: Gender gap decomposition - pooled OLS estimates

	$\log(A)$	$\log(\alpha/n^t)$	$\log(b)$
	(i)	(ii)	(iii)
Female judge	-0.047^{***}	-0.121^{***}	0.073^{**}
	(0.023)	(0.036)	(0.036)
Number of children (in trial)	-0.295^{***}	-0.352^{***}	0.057^{***}
	(0.013)	(0.008)	(0.014)
Number of children (off trial)	-0.102^{***}	-0.183^{***}	0.080^{***}
	(0.014)	(0.009)	(0.011)
Observations	1382	1382	1382
Number of judges	107	107	107
R^2	0.444	0.82	0.195

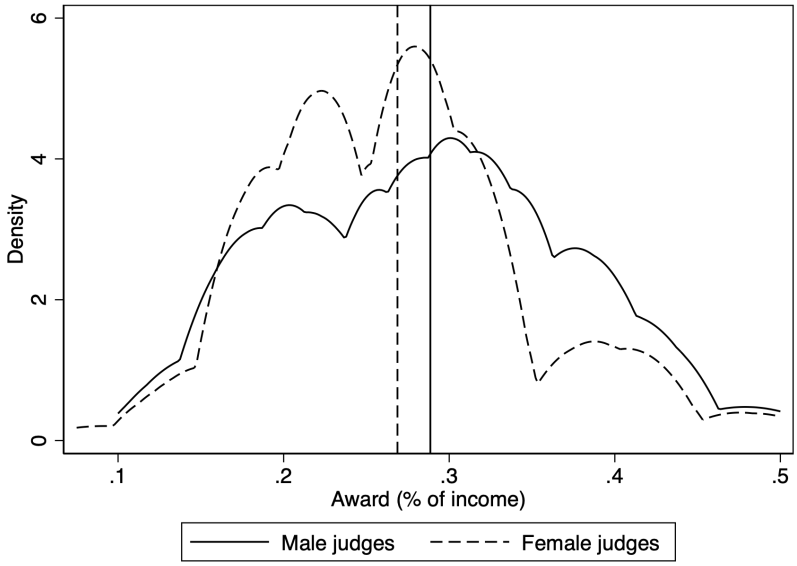
Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table shows pooled OLS estimates of the gender gap decomposition in informal cases. Column (i) uses the log of the award per child as dependent variable. Column (ii) uses the calibrated percentage of the respondent's income awarded per child as dependent variable. Column (iii) uses the log of the belief about the respondent's income as dependent variable. *Female judge* is an indicator variable for whether the case was assigned to a female judge. *Number of children (in trial)* is a variable that contains the number of children involved in the trial. *Number of children (off trial)* is a variable that contains the additional number of children the respondent must take care of. Each regression controls for district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Table 1.13: Belief formation - pooled OLS estimates

	Log (belief)	
	Male judge	**Female judge**
	(i)	**(ii)**
Log of Claim	0.370***	0.253***
	(0.044)	(0.034)
Constant	4.692***	5.385***
	(0.292)	(0.218)
Observations	728	641
Number of judges	50	59
R^2	0.435	0.318

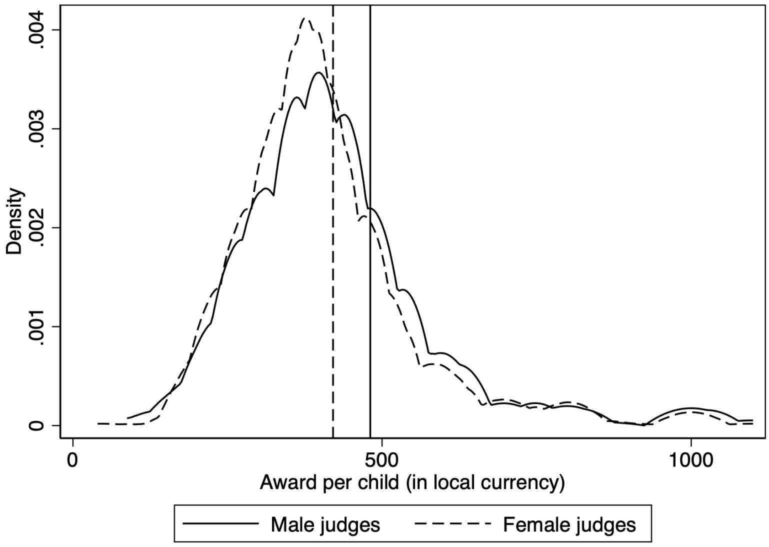
Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. This table shows pooled OLS estimates of the belief formation framework. In columns (i) and (ii), the dependent variable is the logarithm of the estimated belief. *Log of Claim* is the logarithm of the claim presented by the petitioner. Each regression controls for district and year fixed effects. Standard errors in parentheses are clustered at the judge level.

Figure 1.1: Kernel distributions of awards by judges' gender in formal cases



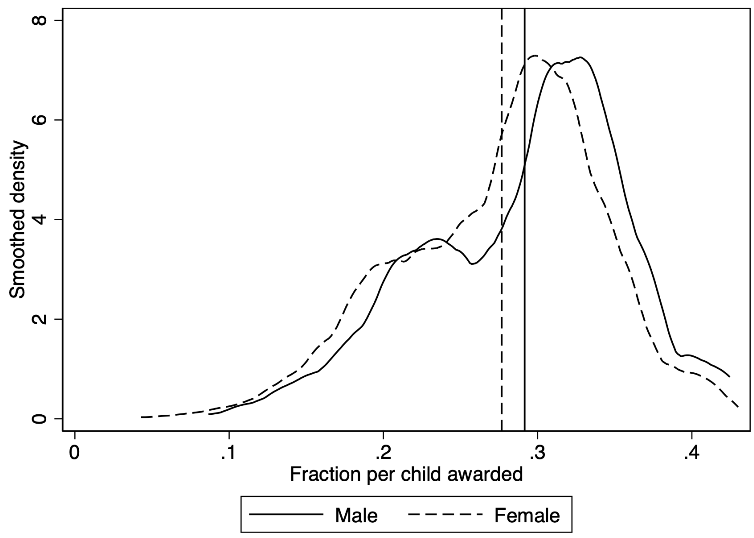
Note: Lines show the kernel densities of child support (as a percentage of the respondent's income) awarded by male and female judges in formal cases.

Figure 1.2: Kernel distributions of awards by judges' gender in informal cases



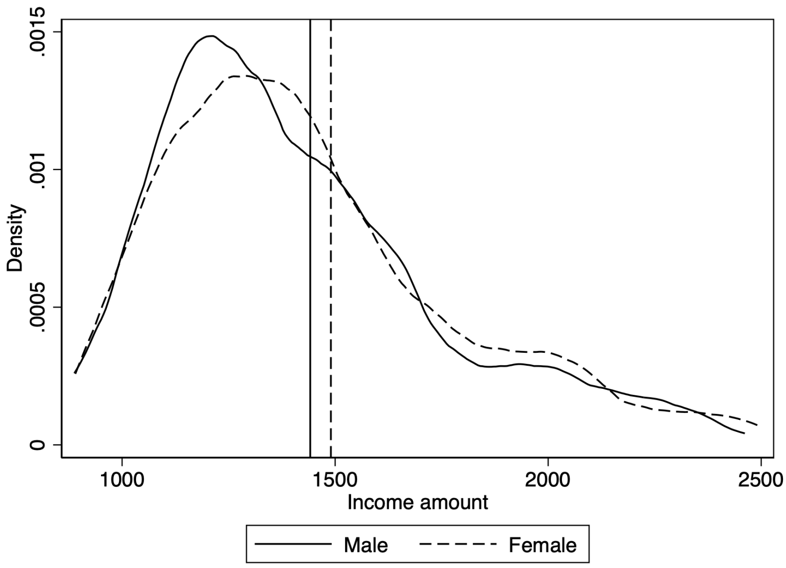
Note: Lines show the kernel densities of child support (as fixed amounts of money to be transferred by the respondent to the petitioner) awarded by male and female judges in informal cases.

Figure 1.3: Kernel distribution of calibrated awards in informal cases



Note: Lines show the kernel densities of calibrated child support amounts (as a percentage of the respondent's income) awarded by male and female judges in informal cases. A calibrated award is the hypothetical allocation set by judges when the respondent's income is not observable (informal job) based on the judges' revealed allocation preferences when the respondent's income is observable (formal job). It captures weights assigned by judges to the respondent's number of children in and out of trial, and the judges' fixed-effect extracted from formal cases.

Figure 1.4: Kernel distribution of estimated judges' beliefs about the respondent's income



Note: Lines show the kernel densities of the estimated judges' beliefs about the respondent's unknown income (in local currency) for male and female judges in informal cases. The estimated belief is the ratio of the award (given in absolute terms) divided by the calibrated award (allocation preference – share of the respondent's income to be allocated) when the respondent's income is not observable (informal job).

Figure 1.5: Kernel distribution of petitioners' claims

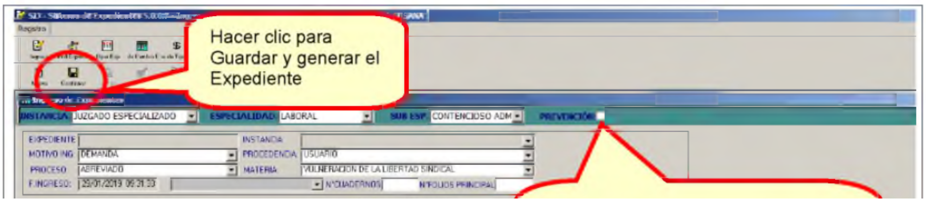
Note: Lines show the kernel densities of the petitioners' claims in absolute terms (in local currency) presented to male and female judges in informal cases. The claim of petitioners is the amount of money they say the father should transfer to cover the needs of their children.

Appendix

Figure 1.6: Randomisation of cases in the Judicial System (before)

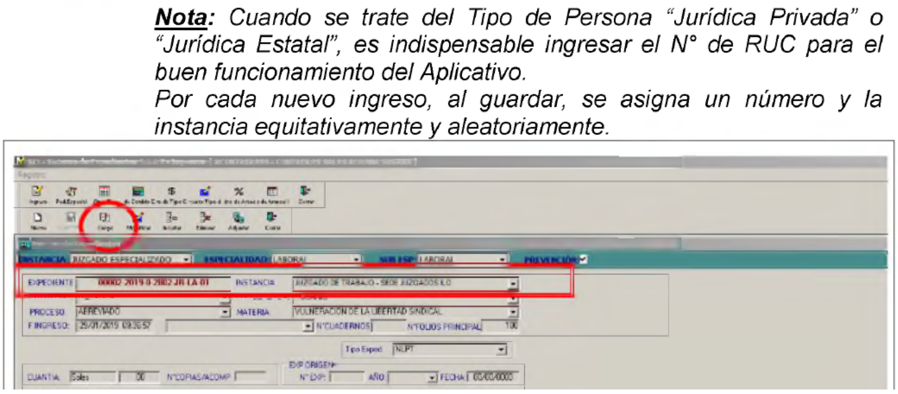
6.3.2. Registro de Información

Para continuar, seleccionar el Órgano Jurisdiccional, la Especialidad y Subespecialidad; luego indicar el Motivo de Ingreso, la Procedencia, el Proceso, la Materia, número de cuadernos, folios del Principal, el Monto de la cuantía si la hubiera, Número de copias, Número, año y fecha del expediente de origen de provenir de 1ra instancia; ingresar la Sumilla y los Tipos de Partes; finalmente, para guardar hacer clic en la opción <Confirmar> de la Barra de Herramientas.



Note: This image shows the step before the system randomly assigns a case to a court within a district as shown in the user guide. As can be seen, the user, when registering the case into the system, cannot choose the court nor the ID of the case. Short translation: "To continue, select categories and characteristics of the case, then click on the save button".

Figure 1.7: Randomisation of cases in the Judicial System (after)



Note: This image shows how the randomisation of cases works as shown in the user guide. As can be seen, once the user saved a case, the system randomly assigns the case to a court and assigns an ID number (increasing order). Short translation: "For each new case registered, when saved, a number is given, and a court is randomly assigned".

2 Chapter 2: The direct and indirect effects of the Peruvian Conditional Cash Transfers Programme “JUNTOS” on Income Satisfaction[14]

This chapter evaluates the direct and indirect effects - through changes in absolute income and income rank with respect to diverse reference groups - of the Peruvian Conditional Cash Transfers Programme “JUNTOS” on income satisfaction. Identification is achieved through the implementation of a quasi-experimental difference-in-differences design that exploits the staggered rollout of JUNTOS in districts across Peru since 2005 together with its eligibility criterion, a poverty score calculated with the Peruvian National Households Survey (ENAHO). I find no direct effects of JUNTOS on income satisfaction. However, JUNTOS generates a statistically significant increase in household income and in the ranked position of the household within the different reference groups, which contributes to the finding of indirect effects of JUNTOS on income satisfaction through these changes. Finally, the results do not support a strong income rank hypothesis (absolute income is a stronger determinant of income satisfaction than income rank); they suggest neighbours are a more important reference group than age group, educational level, or employment status; and they show that social comparisons (based on income rank) only appear in the non-poor household sub-sample.

[14] I am grateful to Gabriela Smarrelli, Roberto Asmat, Carlos Cardona, Sebastian Calvo, Víctor Saldarriaga, Liyang Sun and Clément de Chaisemartin for their invaluable help throughout the development of this project - and the changes it went through - since its inception back in 2018.

2.1 Introduction

Since their inception in Latin America over 20 years ago, Conditional Cash Transfers Programmes (CCT) have proliferated all across the developing world -and some developed countries- with two main objectives: i) reducing current poverty levels, and ii) breaking the inter-generational transmission of poverty (Ibarrarán et al., 2017). Nowadays, there is substantive evidence of their success at attaining those objectives, but also of their impact on more diverse socio-economic indicators. Studies have found mixed impacts on future employment and earnings for their beneficiaries (Millán et al., 2019), no impact on depressive symptoms (Zimmerman et al., 2021), positive short-term effects on general mental health and subjective wellbeing (Haushofer et al., 2021; McGuire et al., 2020), positive significant effects on stunting reduction and human capital accumulation (Cahyadi et al., 2020), and large increases in dietary diversity (Hidrobo et al., 2014), to name a few.

Surprisingly less focus has been given in the literature to the effect of CCTs on household income satisfaction. Theoretically, this indicator measures the perceived discrepancy between earned and desired income: an individual ought to be more satisfied the narrower this gap is (Miething, 2013). At the same time, desired income expectations are set by multiple possible reference standards relative to past income trajectories, the earnings of others in a reference group, or one's own assessment of earning capabilities, for example (Crawford Solberg et al., 2002). For programmes designed to alleviate poverty, the degree to which the transfer amount is perceived to be satisfactory ought to be important enough for policy makers to consider. First, while a cash transfer will certainly increase income levels, it might not be enough to improve the provision of a minimum desired amount of goods and services. Second, measures of income satisfaction might provide better accounts of welfare inequality for policymakers to follow, since they are closer to people's intuitive feelings about inequality (A. Ferrer-i-Carbonell, 2003).

The effect of a CCT on income satisfaction could be direct or indirect. Since 'treatment' is given as a transfer of cash to a beneficiary household subject to some conditionalities, the transfer amount itself might be the one affecting income satisfaction. However, there might also be an effect of the CCT on income satisfaction through how it affects other measures of household income. Concretely, the transfer should increase the beneficiary households' absolute income levels, but it should also change - *ceteris paribus* - their ranked standing relative to those non-eligible households in a reference group. These changes might then be what ends up affecting income satisfaction, regardless of the direct effect of the absolute transfer amount.

So far, the literature is not settled with respect to how changes in different types of income measures affect subjective satisfaction with living standards. Regarding the relationship between absolute income and life satisfaction[15], for instance, studies have found that higher levels of income are associated with better life evaluations and less money with more emotional pain (Diener & Biswas-Diener, 2002; Frank, 2006). It has also been found that this association seems to be stronger in poorer nations, or when it means avoiding poverty regardless of the development status of the country (Diener & Biswas-Diener, 2002). In general, cross-sectional analysis shows that wealthier countries have higher levels of subjective wellbeing and that the wealthiest within societies are more satisfied with their lives (Clark, 2017). Time-series analysis, however, shows a less clear picture. The well-known 'Easterlin Paradox', for instance, suggests that while income and happiness are positively and strongly correlated within a country in one specific period, over time this correlation loses strength (Diener et al., 1993; R. A. Easterlin, 1974, 2001). This finding has led to the suggestion that perhaps it is not absolute levels of income that are the drivers of happiness, but relative concerns based on social comparisons (Clark et al., 2008).[16] However, most of the evidence has historically come from samples that have surpassed 'post-materialistic concerns' (Kuegler, 2009).

Studies incorporating relative concerns based on social comparisons have typically added a 'relative income' variable measured as the average or median income of a reference group in the equations. The evidence by and large shows a negative correlation between the average income of the reference group and subjective wellbeing: increases of income in the reference group are associated with drops in subjective wellbeing for the individual (see Blanchflower and Oswald (2004), using state income as reference; Luttmer (2005b), using neighbours as reference; Graham and Felton (2006), using country and city average income as reference; Kingdon and Knight (2007a), using local and distant others as references; and Dolan *et al.* (2021) for a succinct primer on the literature).

[15] Theoretically, life satisfaction is comprised of people's satisfaction with different domains of their lives (R. Easterlin & Sawangfa, 2007; Praag & Ferrer-i-Carbonell, 2005), satisfaction with income being one of them (Crawford Solberg et al., 2002). While some scholars argue that this association is more complicated than it seems (Rojas, 2006), most of the literature is centered around the study of the relationship between life satisfaction in general (or general subjective wellbeing indicators) and different income measures.

[16] Alternative explanations suggest the paradox might arise because of the asymmetric experience of positive and negative economic growth (de Neve et al., 2018), and that it could be dependent on the measure of subjective well-being used (D. Kahneman & Deaton, 2010).

More recently, however, the literature on the relationship between relative income concerns and life satisfaction has posited that a more theoretically sound measure of relative income is income rank. This notion argues that people do not compare themselves with a reference income level, but that they gain utility by occupying a higher position within a relevant income distribution (Brown et al., 2008; Clark et al., 2009, 2017; Clark & Oswald, 1996; Smith et al., 1989; Stutzer, 2004; Walasek & Brown, 2015). Concretely, this hypothesis posits that status comparisons are based on the relative ordered position of the individual within a rank comprising of members of a reference group.

Theory from different disciplines helps to support this hypothesis. In psychology (and psychophysics in particular), studies have found that individuals tend to judge specific characteristics of different items based on its ranked position along a particular dimension (weight, height or pitch, for example) (Parducci, 1995; Stewart et al., 2006). Furthermore, in biology, on one hand, there is evidence that rank-based status comparisons have evolutionary underpinnings (Samuelson, 2004; Zizzo, 2002): as Powdthavee (2009) suggests, it is likely that in sexual selection processes females would tend to choose mates based on their position in specific hierarchies, such as access to resources for future offspring. On the other hand, correlations have been found between occupying lower status positions in social ranks and higher levels of stress (Sapolsky, 2004).

This relatively new body of literature presents an important empirical challenge. Some studies argue for a ‘strong’ income (or social) rank hypothesis, where comparisons completely negate the effect of absolute income measures on life satisfaction (Boyce et al., 2010; Brown et al., 2017). Others, such as Macchia, Plagnol and Powdthavee (2019), find that income rank does not necessarily negate the association of absolute income with life satisfaction, but that it shows a stronger correlation with subjective wellbeing. Finally, a third strand of studies finds no significant differences between absolute income and income rank, and their association with life satisfaction (FitzRoy & Nolan, 2021).

A plausible explanation for these mixed results might be the selection of reference groups. In simple terms, reference groups are comparison points used by individuals to gauge their relative standing in some domain (Michalos, 1985; Pettigrew, 1967). Therefore, the reference group with which each income rank is constructed might be fundamental for understanding the strength of its association with subjective wellbeing. Indeed, it seems colleagues are the most frequently-cited reference group, that comparisons with friends are associated with low subjective wellbeing, and that individuals tend to choose social groups with whom they interact the most as references (Clark & Senik, 2010).

This chapter aims to understand the direct and indirect effects - through changes in absolute income and income rank with respect to different reference groups - of the Peruvian CCT called "JUNTOS" on income satisfaction. Aided by a quasi-experimental difference-in-differences design, I estimate the direct and indirect effects of JUNTOS on the income satisfaction of a sample of Peruvian households as follows: identification is achieved by exploiting the variation in the rollout of the programme in different districts at different times together with its eligibility criterion, a poverty score constructed ad hoc for JUNTOS with data from the Peruvian National Households Survey (ENAHO). The objective: to compare eligible and non-eligible households before and after JUNTOS arrived in each district. The sample comprises a total of 90,660 households in 1,246 districts across a period of 8 years, between 2004 and 2011[17].

Absolute income is measured as the logarithm of the rounded total household income, which includes both monetary and in-kind income sources. The Income Rank indicators are measured simply as the relative ordered position of the household within the income rank of all the households that comprise the reference group, which are: a) immediate neighbours; b) age group within the district; c) educational attainment within the district; and d) employment status within the district. This distinction is based on the theoretical literature on reference groups, which taxonomizes them into exogenous and endogenous. The first of the categories refers to those reference groups not necessarily actively chosen by the individual (neighbours, parents, siblings) (Ada Ferrer-i-Carbonell, 2005; Kingdon & Knight, 2007b; Ravallion & Lokshin, 2002). Endogenous reference groups, on the other hand, are thought to be actively chosen by the individual for comparison (Diener & Fujita, 1997; Falk & Knell, 2004). The assumption in this study is that reference group a) is exogenously determined, while groups b), c) and d) can be actively chosen by the individual and can thus be considered endogenous.

Any effect discovered should be interpreted as an intent-to-treat effect, since JUNTOS has structural problems of filtration and under-coverage. For instance, according to the most recent institutional report on the programme, JUNTOS has filtration rates (households that should not be receiving the programme, but do) of around 4.1% (Valenzuela, 2015). The first set of results for the difference-in-differences regressions shows no direct effects of JUNTOS on income satisfaction, regardless of econometric specification. This suggests that the transfer amount by itself is not enough to significantly improve perceived living standards.

[17] While JUNTOS started in 2005 and is ongoing to this date, starting in 2012 the algorithm for calculating the poverty score changed drastically and it is no longer calculated based on the ENAHO, but on census data that is not public.

To test for indirect effects of JUNTOS on income satisfaction, I first test whether the cash transfer of JUNTOS significantly increased absolute income and the household position within the income rank of all the different reference groups, otherwise it would be perfectly clear that JUNTOS does not improve income satisfaction neither directly nor indirectly. For all econometric specifications, I find positive effects of JUNTOS on absolute income and every measure of income rank, except for when the reference group is the age group within the district (for which I could not discard pre-intervention parallel trends different than zero). In the preferred specification (adding a full set of controls), receiving JUNTOS improves income levels by around 13.15% on average. Also, households that receive JUNTOS exhibit upward rank mobility, regardless of reference group, of about 5.38 positions (in a 0-100 ranking). All difference-in-differences specifications are robust to pre-intervention parallel trends and spill-over effects of the programme on the non-eligible.

The second set of results shows positive and statistically significant indirect effects of JUNTOS on income satisfaction through changes in absolute income and income rank. These results also provide evidence disproving a strong income rank hypothesis, since not only do the coefficients associated with absolute income not disappear, but these are, on average across specifications, over 7 times stronger than the coefficient associated with the different income rank measures. Finally, the strength of the indirect effect of JUNTOS on life satisfaction through income rank seems to be determined by how close geographically the reference group is (which suggests reference group exogeneity), and not by some characteristic individuals within the household might choose to compare with, such as age group, education level, or employment status (endogenous reference groups).

Interestingly, the Peruvian setting allows for testing one more hypothesis related to the relationship between absolute income, relative income concerns, and income satisfaction: the degree to which social comparisons arise in those households at the bottom of the income distribution. Theoretically, two strands of the literature would suggest the poorest in society would only pay attention to absolute income concerns. First, income-deprived households in developing countries rely much more on kinship relationships and social cohesion with their neighbours and immediate family as a means of financial insurance and risk-sharing (Akay & Martinsson, 2011).

Second, parallel to many theories in developmental psychology, Maslow's hierarchy of needs (Maslow, 1954) would predict it is more likely that richer populations have overcome some basic deficiencies and/or some pressing needs; thus it is also more likely they would be able to care about more ‘superfluous’ needs, such as prestige or status. People in the lower levels of the income hierarchy, on the other hand, are likely to exhibit more ‘absolute’ needs. For example, an optimal calorie/nutrition intake is clearly determined by absolute facts of human biology, rather than social comparison. The same can be said for adequate shelter, access to health services, warmth, or basic comfort, among other basic needs. However, evidence of Maslow’s framework is mixed for developing countries. In Malawi, for instance, Ravallion and Lokshin (2010) find that only the comparatively well-off exhibit concerns for social comparisons. On the other hand, studies in Bangladesh (Asadullah & Chaudhury, 2012) and Peru (Guillen-Royo, 2011) find that the poor also lose subjective wellbeing when engaging in social comparisons with different reference groups.

To test this hypothesis, I re-run the regressions testing for the indirect effects of JUNTOS on income satisfaction through changes in absolute income and income rank in sub-samples of poor and non-poor households, as determined by the ENAHO. I find that social comparisons only appear in the non-poor subsample: for poor households, there is no indirect effect of JUNTOS on income satisfaction through income rank, only through changes in absolute income.

This study contributes, more generally, to the vast literature on CCT impact evaluation. Specifically, it provides original evidence on their potential direct and indirect effects on an understudied (but important) indicator: income satisfaction. It also contributes with causal evidence to the growing literature on the relationship between subjective wellbeing, absolute income, and income rank, in developing countries (where evidence is scarce, since studies have mostly been focused on western rich countries), with varied reference groups for rank construction, and all across the income distribution.

This chapter is structured as follows. Section 2.2 provides a brief overview of the nature and history of the JUNTOS programme in Peru. In Section 2.3 I describes the data sources and the construction of the main variables of interest for the study. Section 2.4 details the empirical strategy, and the econometric specifications and results for each of the research questions in this chapter. In section C, I analyse the results of the two main research questions of this study. Finally, in Section 2.5 I provide some conclusions and the implications of the findings.

2.2 What is JUNTOS?

The Conditional Cash Transfers Programme JUNTOS was first introduced in 2005, in the district of Chuschi, Ayacucho, Peru, with two objectives: to reduce current poverty levels (through the cash transfer) and to minimize the inter-generational transmission of poverty (through its conditionalities aimed at improving human capital). The rollout of the programme was first focused on the poorest districts in Peru (110 in 2005) and, year by year, it has been expanding to include, as of 2016, around 1300 districts (Saldarriaga & Diaz, 2021).

Between 2005 and 2009, beneficiary households received a lump-sum payment of 100 PEN (around US$ 26 at the current exchange rate), which changed to a bi-monthly payment of 200 PEN starting in 2010. To be eligible for JUNTOS, households need to comply with the following requirements. First, they need to be in districts targeted by JUNTOS. This targeting system was used to select districts based on i) exposure to terrorism provoked by the Shining Path, ii) poverty levels (measured as a percentage of the population with unsatisfied basic needs), iii) poverty gaps, iv) malnutrition in children, and v) presence of extreme poverty (Perova & Vakis, 2011). Second, due to programme's objectives, households need to comprise at least one child below the age of 15 (between 2005-2010) or 19 (since 2011), or a pregnant woman. Third, potentially eligible households need to score above a certain poverty score's cut-off point. Between 2005-2011, this poverty score was calculated ad hoc for JUNTOS using the Peruvian National Household Survey (ENAHO); starting in 2012, all social programmes in Peru share the same focalization index.

Finally, payments are made to the household (through the mother or pregnant woman) subject to the following conditionalities. For children under 6 years old, pregnant women and mothers who are breastfeeding: the conditionalities involve pre-natal and post-natal check-ups for mothers and pregnant women, and development and growth controls (CRED) for children. For children over 6 years old, they need to show they have attended at least 85% of the school year. It is worth mentioning that neither eligibility nor conditionalities focus on income satisfaction.

2.3 Data

I construct the sample using repeated cross-sections of districts from the Peruvian National Households Survey (ENAHO) over the period 2004-2011, merged with JUNTOS administrative data for the same period. ENAHO is representative at the national level and is publicly available. JUNTOS administrative data provides district paydays, which I use to assess when JUNTOS reached each district.

The variables of interest are taken from the ENAHO and constructed as follows. The main outcome variable of interest, *income satisfaction*, is measured as a four-point ordered categorical variable that asks the household head ***"Do you deem to live Very Well / Well / Badly / Very Badly with your household income?"***. *Log of Income* is calculated as the logarithm of the rounded total household income, which includes both monetary and in-kind income sources. The *Income Rank* variables are measured simply as the relative ordered position of the household within the income rank of all the households that comprise the reference group. It has been normalized so that its values range from 0 to 100 for all neighbourhoods and thus make them comparable (Boyce et al., 2010; Powdthavee, 2009). The closer a household is to 100, the higher ranked it is.

a) $Rank_{ijt} = \frac{i-1}{n-1}$

The reference groups used for the purposes of this study are immediate neighbours, different age groups within the district, education level within the district and employment status also within the district. Table 2.1 in the Appendix summarizes their mean estimates by JUNTOS eligibility.

Finally, the covariates used as controls for the empirical analysis are the number of household members, and the sex, age, employment status, education level and relationship status of the respondent. I clean for outliers by dropping the top 1% of the distribution in income, spending, age, and number of household members. Again, summary measures of these variables are found in Table 2.1. All these variables are taken from the ENAHO.

2.4 Empirical strategy

2.4.1 Identification condition

The main objective of this study is to estimate the direct and indirect effects (through changes in absolute income and income rank) of JUNTOS on income satisfaction. Since the selection of beneficiaries for JUNTOS is non-random, I exploit the variation in the rollout of the programme in different districts at different times, together with its eligibility criterion, to perform a quasi-experimental difference-in-differences analysis, similarly to Ritter Burga (2014) or Saldarriaga and Diaz (2021). The idea is to compare eligible and non-eligible households before and after JUNTOS arrived in their districts.

Treatment status is given by a household both being exposed to JUNTOS and being eligible for the programme simultaneously. A household starts being exposed to JUNTOS since the first time the programme arrives to the district, information which is collected from JUNTOS administrative data. Also, I determine programme eligibility through the replication of the algorithm used by the programme for defining poverty. Households above the cut-off are deemed eligible and those below it, ineligible. Appendix A describes the specifics for the calculation of this algorithm, which was used by JUNTOS between 2005 and 2011, and which used the ENAHO as the sole source of information. Starting in 2012, all social programmes in Peru changed their eligibility methodology to a unique Household Focalization Index (IFH), which is calculated with census data, which is the main reason that I could not continue the empirical strategy after 2011.

2.4.2 Econometric specifications and results – direct effects

The direct effect of JUNTOS on income satisfaction is modelled by the following preferred specification:

(1) $$IS_{ijt} = \beta_0 + \beta_1 Eligible_{ijt} + \beta_2 Exposed_{ijt} + \beta_3 (Eligible_{ijt} * Exposed_{ijt}) + \beta_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$$

where $\boldsymbol{IS_{ijt}}$ is *income satisfaction* of individual $\boldsymbol{i}$ for the entire household in district $\boldsymbol{j}$ and surveyed at *t*. $\boldsymbol{Eligible_{ijt}}$ and $\boldsymbol{Exposed_{ijt}}$ are dummies indicating whether a household is eligible for JUNTOS or exposed to the programme, respectively. $\boldsymbol{X_{ijt}}$ is the set of covariates acting as controls. $\boldsymbol{I_j}$ and $\boldsymbol{I_y}$ are district and year fixed effects, respectively, and $\boldsymbol{\varepsilon_{ijt}}$ is the error term. I add district and year fixed effects to control for time-invariant factors specific to each district, and time-varying factors common to all of them, respectively.

The coefficient of interest is $\boldsymbol{\beta_3}$, which should be interpreted as the intent-to-treat (ITT) effect of JUNTOS on income satisfaction. It measures the pre-post-treatment change in income satisfaction for eligible households relative to those non-eligible ones, regardless of noncompliance, filtration, withdrawal, or deviation from treatment. The ITT interpretation comes from the fact that JUNTOS has structural problems of filtration and under-coverage. The most recent study suggests filtration rates (households that should not be receiving the programme, but do) of 4.1%, and under-coverage rates (households that should be receiving the programme, but don't) of 79.3% (Valenzuela, 2015).

For $\boldsymbol{\beta_3}$ to be unbiased, two assumptions need to hold: parallel trends in outcomes before exposure, and no spill-over effects of JUNTOS on the non-eligible. However, a quick look at Table 2.2 provides enough statistical evidence of no significant direct associations between JUNTOS and income satisfaction, regardless of specification ((i) does not include controls; (ii) does include controls). This directly implies that the mere reception of the cash transfer does not improve income satisfaction.

Regression (1) in Table 2.2 shows regression results not including income variables. Regressions (2) - (5) show results including the log of absolute income and income rank (with different reference groups) as covariates. I find strong positive statistical associations between the measure of absolute income (log of total income) and income satisfaction for every regression from (2) to (5). Holding the JUNTOS transfer constant, absolute income is always a positive determinant of income satisfaction, regardless of specification. On average, every 10% increase in absolute income is associated with an improvement in income satisfaction of 0.012 in a four-point scale. Income rank, on the other hand, is only consistently significant and shows a positive association with income satisfaction in the regressions that control for the number of household members, and the sex, age, employment status, education level and relationship status of the respondent. However, the magnitude of the coefficient is too small to make meaningful interpretations.

2.4.3 Econometric specifications and results – indirect effects

The lack of direct effects of JUNTOS on income satisfaction might stem from the fact that the transfer amount is not big enough to be perceived as significant by households. Conceptually, however, this cash transfer should be increasing both the household's absolute income levels and the household's relative position within a reference group's income rank, compared to those households that did not receive the transfer, *ceteris paribus*. If JUNTOS generates statistically meaningful increases on these income measures, it would then be plausible to hypothesize that JUNTOS might have an indirect effect on income satisfaction through changes in absolute income and income rank.

To first test the effect of JUNTOS on absolute income and income rank, I modelled the following preferred econometric specifications:

(2) $Y_{ijt} = \beta_0 + \beta_1 Eligible_{ijt} + \beta_2 Exposed_{ijt} + \beta_3 (Eligible_{ijt} * Exposed_{ijt}) + \beta_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

(3) $RankY_{rijt} = \beta_0 + \beta_1 Eligible_{ijt} + \beta_2 Exposed_{ijt} + \beta_3(Eligible_{ijt} * Exposed_{ijt}) + \beta_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

where $\boldsymbol{Y_{ijt}}$ is the Log of Total Household Income for household ***i*** in district *j* and surveyed at ***t***; $\boldsymbol{RankY_{rijt}}$ is Total Income Rank for reference group ***r*** (the district, age group within the district, educational level within the district, and employment status within the district) and household ***i*** in district ***j*** and surveyed at ***t***. $\boldsymbol{Eligible_{ijt}}$ and $\boldsymbol{Exposed_{ijt}}$ are dummies indicating whether a household is eligible for JUNTOS or exposed to the programme, respectively. $\boldsymbol{X_{ijt}}$ is the set of covariates acting as controls. $\boldsymbol{I_j}$ and $\boldsymbol{I_y}$ are district and year fixed effects, respectively, and $\boldsymbol{\varepsilon_{ijt}}$ is the error term. I add district and year fixed effects to control for time-invariant factors specific to each district, and time-varying factors common to all of them, respectively.

Similarly to 2.4.2, the coefficient of interest is $\boldsymbol{\beta_3}$, which should be interpreted as the intent-to-treat effect of JUNTOS on absolute income and income rank. It measures the intertemporal change of income and income rank for those households eligible for JUNTOS relative to their non-eligible counterparts. A positive and significant coefficient would suggest a causal effect of JUNTOS on income and income rank, only if two assumptions hold: parallel trends in outcomes before exposure, and no spill-over effects of JUNTOS on the non-eligible.

2.4.3.1 Parametric event studies

Regarding the first set of robustness checks, Figures 2.1 – 2.5 show event study estimates for the effect of JUNTOS on the log of Total Income and Income Rank (with the district, district and age group, district and educational level, and district and employment status as reference groups). These estimates are given by the following regressions:

(4) $Y_{ijt} = \delta_1 Eligible_{ijt} + \sum_{t \neq t_0} \delta_2 Exposed_{ijt} + \sum_{t \neq t_0} \delta_3(Eligible_{ijt} * Exposed_{ijt}) + \delta_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

(5) $RankY_{rijt} = \delta_1 Eligible_{ijt} + \sum_{t \neq t_0} \delta_2 Exposed_{ijt} + \sum_{t \neq t_0} \delta_3(Eligible_{ijt} * Exposed_{ijt}) + \delta_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

where ***t*** denotes the number of years since the advent of JUNTOS in each district, and $\boldsymbol{t_0}$ indexes the year prior to arrival as a reference point. For parallel trends in outcomes before exposure to hold, we should expect to find no joint effects different than zero in $\boldsymbol{\delta_3}$ for $\boldsymbol{t < t_0}$.

This would mean there were no differences in absolute income and income rank between eligible and non-eligible households before JUNTOS arrived at each district. Again, I add district and year fixed effects to control for time-invariant factors specific to each district, and time-varying factors common to all of them, respectively.

Figures 2.1 – 2.5 show pre-intervention parallel trends for the effect of JUNTOS on the following income measures: log of total income (F-statistic of joint significance 0.67; p-value 0.6456), income rank with the district as a reference (F-statistic of joint significance 1.18; p-value 0.3163), income rank with educational level within the district as a reference (F-statistic of joint significance 0.58; p-value: 0.7183), and income rank with employment status within the district as a reference (F-statistic of joint significance 1.00; p-value 0.4149). Only for the effect of JUNTOS on rank of income with the district and age group as a reference are pre-intervention parallel trends not to be found (F-statistic of joint significance is 2.33; p<0.05).

On the other hand, all event studies show post-intervention trends F-statistics of joint significance different than zero, suggesting the programme did have an effect. The post-intervention trends for the effect of JUNTOS on log of income show an F-statistic of joint significance of 8.43 (p<0.01); on rank of income with the district as a reference, 3.51 (p<0.01); on rank of income with age group within the district as a reference, 2.71 (p<0.05); on rank of income with educational level within the district as a reference, 3.52 (p<0.01); and on rank of income with employment status within the district as a reference, 3.59 (p<0.01).

2.4.3.2 Spill-over effects of exposure to JUNTOS on the non-eligible

Spill-over effects of JUNTOS on the non-eligible, for the purposes of this study, might arise due to problems with filtration. As explained in 2.4.2, the most recent institutional overview of JUNTOS calculates a filtration rate of 4.1% (Valenzuela, 2015). This suggests that around 4.1% of recipient households should not be eligible for the programme due to its poverty score being below the threshold. A formal test of spill-over effects is thus needed to guarantee unbiased causal estimations.

I test spill-over effects of JUNTOS on the sub-sample of non-eligible households as follows:

(6) $Y_{ijt} = \beta_0 + \beta_1 Exposed_{ijt} + \beta_2 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

(7) $RankY_{rijt} = \beta_0 + \beta_1 Exposed_{ijt} + \beta_2 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

where econometric specifications are the same as 2.4.3, the only difference being the sample on which the regressions are run: non-eligible households. For spill-overs not to be present, the coefficient of interest, $\boldsymbol{\beta_1}$, should not be statistically significant. This would indicate that there is no effect of JUNTOS on the different income measures for the sub sample of non-eligible, or, in other words, that the rate of filtration of the programme is negligible. Table 2.3 presents the results of the spill-over regressions. There is no evidence of spill-over effects of JUNTOS on any of the income measures, regardless of specification (controls and no controls).

2.4.3.3 Results

Having taken care of the assumptions necessary for an unbiased estimation of the effect of JUNTOS on absolute income and the different measures of income rank, Table 2.4 presents the results. Regressions (1) through (5) estimate the effect of JUNTOS on log of total income, income rank with the district as a reference, income rank with age group within the district as a reference, income rank with educational level within the district as a reference, and income rank with employment status within the district as a reference, respectively.

All regressions show a positive and statistically significant ($p<0.01$) coefficient of interest, suggesting that JUNTOS exerts an effect on all the different income measures (except for income rank with the district and age group as a reference, since we could not confirm pre-intervention parallel trends in outcomes). In the preferred specification (adding a full set of controls), receiving JUNTOS improves income levels by around 13.15% on average. Also, households that receive JUNTOS exhibit upward rank mobility, regardless of reference group, of about 5.38 positions (in a 0-100 ranking).

2.4.3.4 Indirect effects

Since JUNTOS has been shown to affect all the different income measures, but not income satisfaction directly, this sub-section aims at testing whether there is an indirect effect of JUNTOS on income satisfaction through changes in absolute income and income rank. To this avail, first I model the following set of equations:

(8) $IS_{ijt} = \theta_0 + \theta_1 Eligible_{ijt} + \theta_2 Exposed_{ijt} + \theta_3(Eligible_{ijt} * Exposed_{ijt}) + \theta_4 Y_{ijt} + \theta_5 RankY_{ijt} + \theta_6 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

(9) $Y_{ijt} = \varphi_0 + \varphi_1 Eligible_{ijt} + \varphi_2 Exposed_{ijt} + \varphi_3 (Eligible_{ijt} * Exposed_{ijt}) + \varphi_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

(10) $RankY_{rijt} = \gamma_0 + \gamma_1 Eligible_{ijt} + \gamma_2 Exposed_{ijt} + \gamma_3 (Eligible_{ijt} * Exposed_{ijt}) + \gamma_4 X_{ijt} + I_j + I_y + \varepsilon_{ijt}$

The indirect effect of JUNTOS on $\boldsymbol{IS_{ijt}}$ through changes in $\boldsymbol{Y_{ijt}}$ and $\boldsymbol{RankY_{ijt}}$ is then given by $\boldsymbol{\theta_4}$ **x** $\boldsymbol{\varphi_3}$ and $\boldsymbol{\theta_5}$ **x** $\boldsymbol{\gamma_3}$, respectively. To make estimations comparable, all variables of interest $\boldsymbol{IS_{ijt}}$, $\boldsymbol{Y_{ijt}}$, and $\boldsymbol{RankY_{rijt}}$, are standardized with mean 0 and standard deviation of 1. It is also worth noting again that $\boldsymbol{RankY_{ijt}}$ is measured considering different reference points (the district, district and age group, district and education level, and district and employment status). All regressions include a set of controls and district and year fixed effects. The standard errors for all estimated effects are bootstrapped (200 replications), as suggested by Hayes (2009).

Table 2.5 presents these estimations. Specification (1) estimates indirect effects of JUNTOS on income satisfaction through changes in the log of total income and income rank with the district as the reference group. Specification (2) estimates indirect effects of JUNTOS on income satisfaction through changes in the log of total income and income rank with age group within the district as the reference group. Specification (3) estimates indirect effects of JUNTOS on income satisfaction through changes in the log of total income and income rank with educational level within the district as the reference group. Finally, specification (4) estimates indirect effects of JUNTOS on income satisfaction through changes in the log of total income and income rank with employment status within the district as the reference group.

Both the coefficients associated to the measures of absolute income and income rank are strongly and positively significant (p <0.01) across all specifications. Receiving JUNTOS improves income satisfaction through changes in absolute income an income rank. Contrary to the strong income rank hypothesis, the indirect effects of JUNTOS on income satisfaction through changes in income rank do not make the indirect effects through changes in absolute income disappear. In fact, in every specification, the indirect effects on income satisfaction through absolute income are, on average across specifications, over 7 times as strong.

Additionally, specification (1) is where the indirect effect of income rank is the strongest (0.0051 standard deviations) and that of absolute income the weakest (0.0263 standard deviations). In general, there seems to be a trade-off between the strength (or weakness) of the indirect effect of one explanatory variable versus the other, since the same pattern is observed in every specification. Finally, the strength of the indirect effect of income rank seems to be determined by the geographical proximity of the reference group (neighbours) - which suggests exogenous reference groups are more determinant - and not some qualitative property of social comparisons (age group, education level, or employment status).

2.4.3.5 Indirect effects – poor and non-poor sub-sample

To explore the degree to which social comparisons arise in those households at the bottom of the income distribution, or whether these exist at all, I model the econometric specifications of 2.4.3.4 in sub-samples of poor and non-poor households, as determined by the ENAHO. Table 2.6 presents these results. Regressions (1) – (4) are run on the poor sub-sample; and regressions (5) – (8) on the non-poor sub-sample. I find no indirect effects of JUNTOS on income satisfaction through income rank, regardless of reference group, in the poor sub-sample. Relative concerns only appear, with a positive and significant sign, in the regressions on the non-poor sub-sample. Absolute income, on the other hand, is always significant regardless of sub-sample.

2.5 Limitations

The standard difference-in-differences (DID) estimator compares outcome evolution between two periods and between two groups: a treatment group that switches from untreated to treated, and a control group that is untreated at both dates (de Chaisemartin & d'Haultfoeuille, 2021). It is unbiased if parallel trends hold; that is, if in absence of treatment, the outcome evolution for both the treated and untreated groups would have been the same. However, recent research has shown that when treatment is staggered at different points in time - what in the literature is known as a Two-way Fixed Effects estimation (TWFE) - an additional assumption needs to hold: homogenous treatment effects between all groups and over time. This is often implausible (Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2020).

The problem seems to originate in the composition of the estimator. Goodman-Bacon (2021), for example, has shown that the TWFE estimator, $\boldsymbol{\widehat{\beta}_{fe}}$, is the weighted average of two types of DIDs: i) a DID comparing switchers (untreated to treated) to the never-treated; and ii) a DID comparing switchers to the always-treated. This second group of DIDs - which are known as the 'forbidden comparisons' - would be the ones responsible for altering the weights (even contributing to negative weights) and thus biasing the TWFE estimator.

The empirical strategy used in this chapter is essentially a TWFE estimation of the direct and indirect effects of JUNTOS on income satisfaction. A proper estimation of the treatment effects is thus necessary for the unbiased causal interpretation of results. The new literature on TWFE presents some solutions for this problem. For instance, de Chaisemartin and D'Haultfœuille (2020) construct a new estimator altogether, which is significantly different from the linear regression estimator, and which estimates the treatment effect on groups that switch to treatment only during the period when they switch. Wooldridge (2021), on the other hand, performs a simple extension of the Mundlak (1978) device reproducing TWFE estimates by adding both unit-specific time series averages and period-specific cross-sectional averages in pooled OLS regressions.

A better fit for the nature of the pooled cross-section of this study, however, is the estimator devised by Sun and Abraham (2021). Their alternative method estimates the share of each cohort as weights to deal with dynamic heterogeneous treatment effects, which they find to be more interpretable than the weights underlying the standard TWFE estimation. To provide a robustness check of the DID estimations in 2.4.3 (the causal effect of JUNTOS on absolute income and income rank), I re-estimate regressions (1) and (2) interacting $\boldsymbol{Eligible_{ijt} * Exposed_{ijt}}$ with a cohort indicator, so as to avoid using later eligible households as controls. I then calculate cohort shares. Finally, I construct the Sun and Abraham (2021) estimator as the weighted average of the new regression coefficients.[18]

Table 2.7 presents the results of these estimations. In general, the effects of JUNTOS on absolute income and income rank (regardless of reference group) seem to be smaller but statistically significant and in the same direction as the original estimates. For example, in Table 2.4 I show that receiving JUNTOS improves absolute income levels in 12.63% when including covariates. On the other hand, the Sun and Abraham (2021) estimator suggests JUNTOS improves absolute income levels in 12.41%. This difference seems to originate from the treatment effects associated with the 2005 cohort (those eligible households that started being exposed to JUNTOS in 2005). The rest of the cohorts exhibit relatively homogeneous treatment effects.

[18] This method of constructing the Sun and Abraham (2021) estimator in Stata was discussed thoroughly with Liyang Sun, co-author of the paper, in a fluid e-mail exchange in July 2021. This is an example of how novel the recent literature is.

While this robustness check would suggest that the dynamic treatment effects I found are relatively homogeneous, the literature on the TWFE estimator is changing rapidly and discussions are happening almost in real-time, so I would not discard other kinds of problems arising in the (not so distant) future.

2.6 Conclusion

This chapter has evaluated the direct and indirect effects - through changes in absolute income and income rank with respect to different reference groups - of the Peruvian CCT JUNTOS on income satisfaction. Results show no direct effects of JUNTOS on income satisfaction. However, they reveal statistically significant increases in absolute income and in the ranked position of the household within exogenous and endogenous reference groups, which contributes to a second set of findings of positive and significant indirect effects of JUNTOS on income satisfaction through these changes. As detailed in the introduction, the degree to which the transfer amount provided by the CCT is perceived to be satisfactory by the eligible households ought to be important enough for policy makers to consider. While a cash transfer will certainly increase income levels, it might not be enough to satisfy a household's desired amount that could at least cover basic needs.

This chapter has also shown that income-rank comparisons only appear among those above the poverty line, which goes in line with the predictions implied by Maslow's hierarchy of needs (Maslow, 1954), but it still stands that absolute income is always important as a life satisfaction determinant. In fact, I do not find support of neither for a strong nor a weak income rank hypothesis: the effect of JUNTOS on income satisfaction through changes in absolute income is over seven times stronger than through changes in income rank. It is only after households pass a certain income threshold that social comparisons arise. This finding gives strength to the relevance of rising absolute income levels for the worse-off in society, both under the objective and subjective frameworks of wellbeing, as a valuable objective of social policy.

Finally, this study shows that the strength of social comparisons depends on the nature of the reference group. The indirect effect of JUNTOS on income satisfaction through changes in income rank is stronger when the reference group is the district in general (neighbours), when compared to age group, educational level, and employment status. This would suggest that, contrary to what recent evidence has found (Clark & Senik, 2010), geographical proximity (and thus relative group exogeneity) seems to be more relevant to determine the importance of reference groups than some other endogenous comparison point (like friends or colleagues). Granted, however, the rest of reference groups in this study (age group, educational level, and employment status), can only be inferred as endogenous, since their calculation does not correspond to direct questions on reference groups in the ENAHO.

Limitations arise due to the recent econometric developments around difference-in-differences estimations with staggered timing in treatment. It is possible that the causal estimates found in this study are biased because of this, although the robustness checks performed with the Sun and Abraham (2020) estimations suggest otherwise. Further research might focus on the policy implications on the nature of relative concerns and their relationship with subjective wellbeing and on the relevance of different reference groups in determining this association.

Table 2.1: Summary statistics of main variables

	Non-Eligibles	Eligibles
Variable	Mean/SE	Mean/SE
Income Satisfaction	2.624	2.524
(1=Very bad - 2 Bad - 3 Good - 4 Very Good)	[0.006]	[0.009]
Total Household Income (PEN)	22324.398	8306.067
	[321.329]	[96.248]
Rank Total Income (0-100)	56.112	46.438
(Community as reference group)	[0.232]	[0.450]
Rank Total Income (0-100)	55.527	47.907
(Community and education as reference group)	[0.203]	[0.381]
Rank Total Income (0-100)	55.554	44.932
(Community and age group as reference group)	[0.194]	[0.425]
Rank Total Income (0-100)	56.030	46.546
(Community and employment status as reference group)	[0.232]	[0.451]
Age (Years)	42.772	42.619
	[0.143]	[0.170]
Employment status	0.238	0.102
(0 Employed - 1 Unemployed)	[0.006]	[0.004]
Education level - Primary or Secondary	0.727	0.831
	[0.007]	[0.006]
Education level - Tertiary	0.243	0.012
	[0.007]	[0.001]
Number of Household Members	4.743	5.350
	[0.018]	[0.024]
Sex	0.517	0.445
(0 Male - 1 Female)	[0.004]	[0.005]
Relationship Status	0.206	0.313
(0 Single 1 With Partner)	[0.005]	[0.009]

Table 2.2: Direct effects of JUNTOS on income satisfaction

	(1)		(2)		(3)		(4)		(5)	
	Income Satisfaction		Income Satisfaction		Income Satisfaction		Income Satisfaction		Income Satisfaction	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls
Exposed * Eligible to JUNTOS	0.003	0.004	-0.012	-0.013	-0.012	-0.013	-0.011	-0.013	0.003	-0.013
	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)
Eligible to JUNTOS	-0.106***	-0.090***	-0.040***	-0.014	-0.041***	-0.014	-0.038***	-0.014	-0.040***	-0.014
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Exposed to JUNTOS	-0.036	-0.034	-0.025	-0.024	-0.024	-0.024	-0.025	-0.024	-0.025	-0.024
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Log of Income			0.113***	0.126***	0.098***	0.128***	0.134***	0.131***	0.113***	0.131***
			(0.008)	(0.008)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Rank of Income (District as reference)			0.001	0.001**						
			(0.001)	(0.001)						
Rank of Income (District and Age Group as reference)					0.001***	0.001***				
					(0.001)	(0.001)				
Rank of Income (District and Education Level as reference)							-0.001***	0.001***		
							(0.001)	(0.001)		
Rank of Income (District and Employment Status as reference)									0.001	0.001**
									(0.001)	(0.001)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660
Number of clusters	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246
R^2	0.077	0.087	0.082	0.109	0.082	0.109	0.082	0.109	0.082	0.109

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Each column shows the coefficient of interest for the OLS linear regressions identifying the direct intent-to-treat effect of JUNTOS on income satisfaction. The dependent variable in (1), (2), (3), (4) and (5) measures satisfaction with income in a four-point ordered categorical scale. In (1), regressions do not include neither absolute nor relative income variables. In (2), regressions include the log of total income and income rank with the district as reference group as covariates. In (3), regressions include the log of total income and income rank with the district and age group as reference group as covariates. In (4), regressions include the log of total income and income rank with the district and educational level as reference group as covariates. In (5), regressions include the log of total income and income rank with the district and employment status as reference group as covariates. (i) is the basic specification; (ii) adds individual and household level controls. Errors are clustered at the district level. Regressions estimated with the *reghdfe* command in Stata. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Table 2.3: Spill-over effects of JUNTOS on absolute income and income rank

	(1)		(2)		(3)		(4)		(5)	
	Log of Income		**Rank of income - Community as reference**		**Rank of income - Community and age group as reference**		**Rank of income - Community and education as reference**		**Rank of income - Community and employment as reference**	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls
Exposed to JUNTOS	-0.033 (0.038)	-0.049 (0.035)	0.281 (1.249)	-0.383 (1.110)	0.590 (1.344)	0.197 (1.231)	0.661 (1.363)	-0.153 (1.265)	0.383 (1.311)	-0.217 (1.197)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	66,020	66,020	66,020	66,020	66,020	66,020	66,020	66,020	66,020	66,020
Number of clusters	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099
R^2	0.230	0.386	0.032	0.224	0.025	0.162	0.025	0.169	0.031	0.210

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Each column shows the coefficient of interest for the spill-over OLS linear regressions of JUNTOS on the different income measures. The dependent variable in (1) measures the natural logarithm of the household's total income. In (2), the dependent variable denotes the relative ranked position of a household within an income rank taking the district as the reference point. In (3), (4) and (5), the dependent variables are constructed analogously as in (2), only with changes in reference group: in (3) the reference group is the district and the age group; in (4), district and educational levels; and in (5), district and employment status. (i) is the basic specification; (ii) adds individual and household level controls. All regressions use the sample of non-eligible for JUNTOS. Errors are clustered at the district level. Regressions estimated with the *reghdfe* command in Stata. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Table 2.4: Effect of JUNTOS on absolute income and income rank

	(1)		(2)		(3)		(4)		(5)	
	Log of Income		**Rank of income - Community as reference**		**Rank of income - Community and age group as reference**		**Rank of income - Community and education as reference**		**Rank of income - Community and employment as reference**	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls
Exposed * Eligible to JUNTOS	0.128***	0.119***	5.590***	5.208***	5.539***	5.055***	5.601***	5.692***	5.690***	5.256***
	(0.031)	(0.026)	(1.175)	(0.994)	(1.223)	(1.084)	(1.011)	(0.973)	(1.194)	(1.015)
Eligible to JUNTOS	-0.566***	-0.533***	-20.549***	-19.342***	-20.331***	-19.334***	-16.656***	-19.272***	-20.385***	-19.083***
	(0.016)	(0.013)	(0.583)	(0.487)	(0.623)	(0.534)	(0.535)	(0.524)	(0.593)	(0.496)
Exposed to JUNTOS	-0.077**	-0.069**	-3.027***	-2.625***	-3.475***	-2.824***	-3.029***	-2.821***	-3.241***	-2.805***
	(0.036)	(0.030)	(0.985)	(0.867)	(1.036)	(0.944)	(0.931)	(0.842)	(0.994)	(0.885)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660
Number of clusters	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246
R^2	0.423	0.534	0.061	0.232	0.044	0.161	0.038	0.172	0.056	0.215

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Each column shows the coefficient of interest for the OLS linear regressions identifying the intent-to-treat effect of JUNTOS on the outcome variables of interest. The dependent variable in (1) measures the natural logarithm of the household's total income. In (2), the dependent variable denotes the relative ranked position of a household within an income rank taking the district as the reference point. In (3), (4) and (5), the dependent variables are constructed analogously as in (2), only with changes in reference group: in (3) the reference group is the district and the age group; in (4), district and education levels; and in (5), district and employment status. (i) is the basic specification; (ii) adds individual and household level controls. Errors are clustered at the district level. Regressions estimated with the *reghdfe* command in Stata. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Table 2.5: Indirect Effects of JUNTOS on income satisfaction through changes in absolute income and income rank

	(1)	(2)	(3)	(4)
Log Income	0.0263***	0.0275***	0.0282***	0.0281***
	(0.0039)	(0.0037)	(0.0038)	(00038)
Total Income Rank - Community as Reference	0.0051***			
	(0.0018)			
Total Income Rank - Community and Age Group as Reference		0.0038***		
		(0.0102)		
Total Income Rank - Community and Education Level as Reference			0.0033***	
			(0.0013)	
Total Income Rank - Community and Employment Status as Reference				0.0033***
				(0.0012)
Total Indirect Effects	0.0314***	0.0313***	0.0316***	0.0314***
	(0.0040)	(0.0041)	(0.0040)	(0.0040)

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Table shows estimates of indirect effects of JUNTOS on income satisfaction through changes in absolute income and income rank. Variables of interest are standardized with mean 0 and standard deviation 1. Estimations correspond to the basic specifications with individual and household level controls. Errors are bootstrapped with 200 replications. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Table 2.6: Indirect Effects of JUNTOS on income satisfaction through changes in absolute income and income rank – Poor and Non-Poor Subsample

	Poor				Non-Poor			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Log Income	0.0122***	0.0121***	0.0120***	0.0125***	0.0220***	0.0245***	0.0261***	0.0255***
	(0.0032)	(0.0029)	(0.0029)	(0.0030)	(0.0054)	(0.0057)	(0.0061)	(0.0060)
Total Income Rank - Community as Reference	0.0008				0.0094**			
	(0.0011)				(0.0038)			
Total Income Rank - Community and Age Group as Reference		0.0011				0.0051**		
		(0.0008)				(0.0022)		
Total Income Rank - Community and Education Level as Reference			0.0012				0.0041*	
			(0.0009)				(0.0022)	
Total Income Rank - Community and Employment Status as Reference				0.0004				0.0052*
				(0.0006)				(0.0026)
Total Indirect Effects	0.0130***	0.0133***	0.0132***	0.0130***	0.0314***	0.0297***	0.0302***	0.0308***
	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0071)	(0.0070)	(0.0070)	(0.0071)

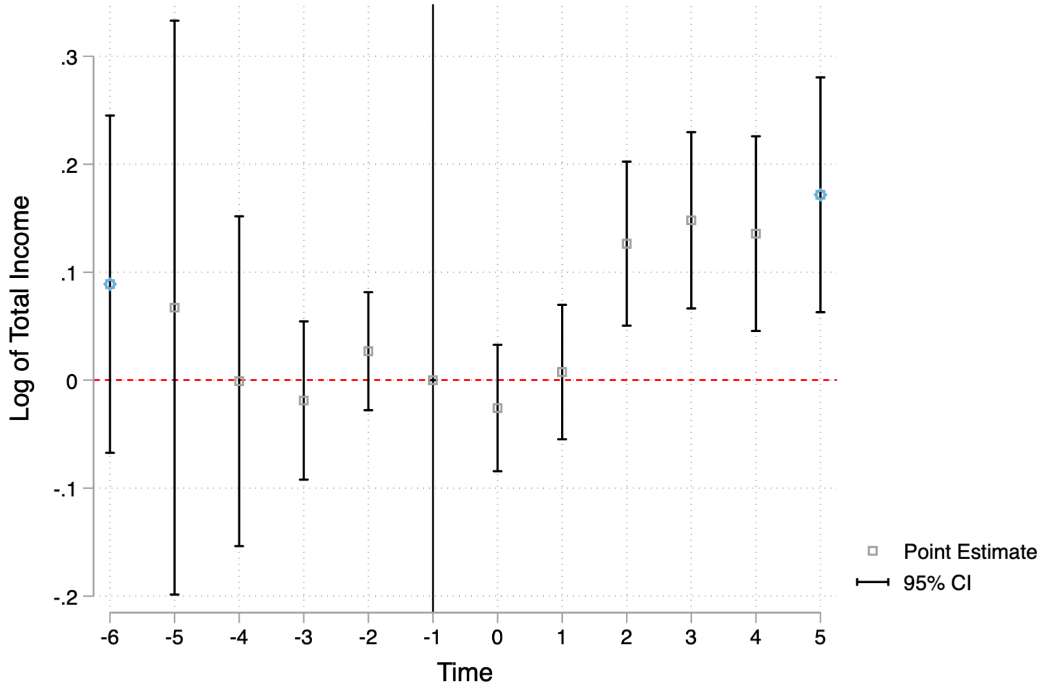
Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Table shows estimates of indirect effects of JUNTOS on income satisfaction through changes in absolute income and income rank. Variables of interest are standardized with mean 0 and standard deviation 1. Estimations correspond to the basic specifications with individual and household level controls. Errors are bootstrapped with 200 replications. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Table 2.7: Effect of JUNTOS on outcome variables – Ad-hoc calculation of IW estimator (Sun and Abraham, 2020)

	(1)		(2)		(3)		(4)		(5)	
	Log of Income		**Rank of income - Community as reference**		**Rank of income - Community and age group as reference**		**Rank of income - Community and education as reference**		**Rank of income - Community and employment as reference**	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls	Basic Specification	With Controls
Exposed * Eligible to JUNTOS	0.127***	0.117***	5.600***	5.200***	5.544***	5.043***	5.617***	5.701***	5.694**	5.242**
	(0.030)	(0.025)	(1.171)	(0.989)	(1.218)	(1.081)	(1.010)	(0.970)	(1.191)	(1.010)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660	90,660
Number of clusters	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246	1,246
R^2	0.423	0.534	0.061	0.232	0.044	0.161	0.038	0.172	0.056	0.215

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Each column shows the coefficient of interest for the linear regressions identifying the intent-to-treat effect of JUNTOS on the outcome variables of interest. The dependent variable in (1) measures the natural logarithm of the household's total income. In (2), the dependent variable denotes the relative ranked position of a household within an income rank taking the community as the reference point. In (3), (4) and (5), the dependent variables are constructed analogously as in (2), only with changes in reference group: in (3) the reference group is the community and the age group; in (4), community and education levels; and in (5), community and employment status. (i) is the basic TWFE specification; (ii) adds individual and household level controls. Errors are clustered at the district level. Regressions estimated with the *reghdfe* command in Stata. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

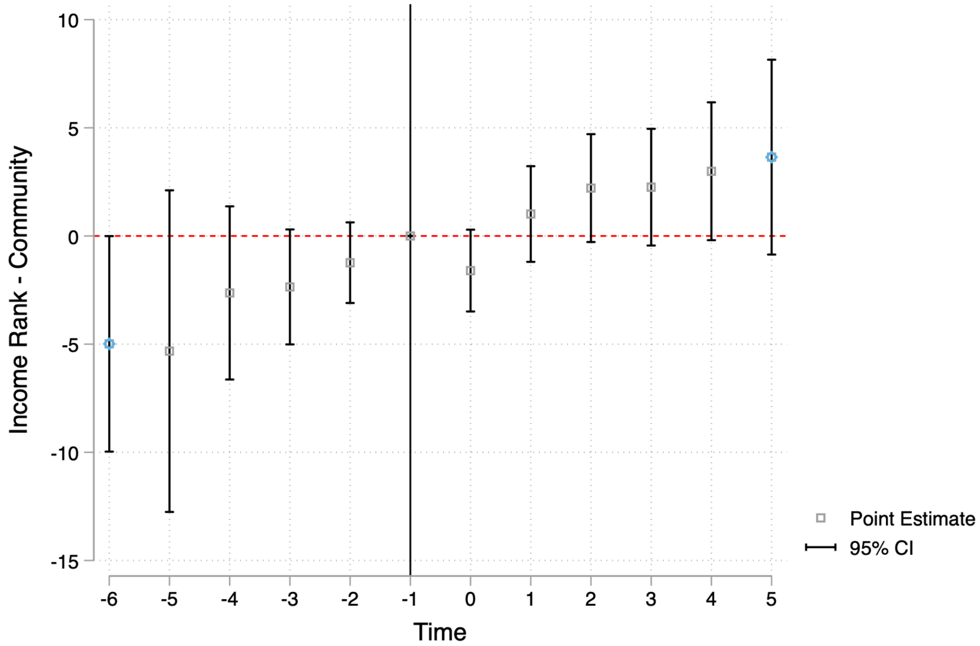
Figure 2.1: Event Study Estimates of the Effect of JUNTOS on Log of Total Income



F-statistic of pre-treatment trends = 0.67 ; p-value=0.6456

Note: Event Study estimates of the effect of JUNTOS on the Log of Total Household Income along with 95% confidence intervals and standard errors clustered at the district level. Graph includes individual and household controls. Crucial for the identification of parallel trends before treatment is the impossibility to reject that the coefficients of pre-intervention effects are equal to 0. The F-statistic for a joint significance test of pre-intervention trends is reported below the graph and suggests parallel trends. Conversely, for the post-treatment period, the F-statistic of joint significance is 8.43 (p-value<0.01), suggesting joint effects different than 0. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

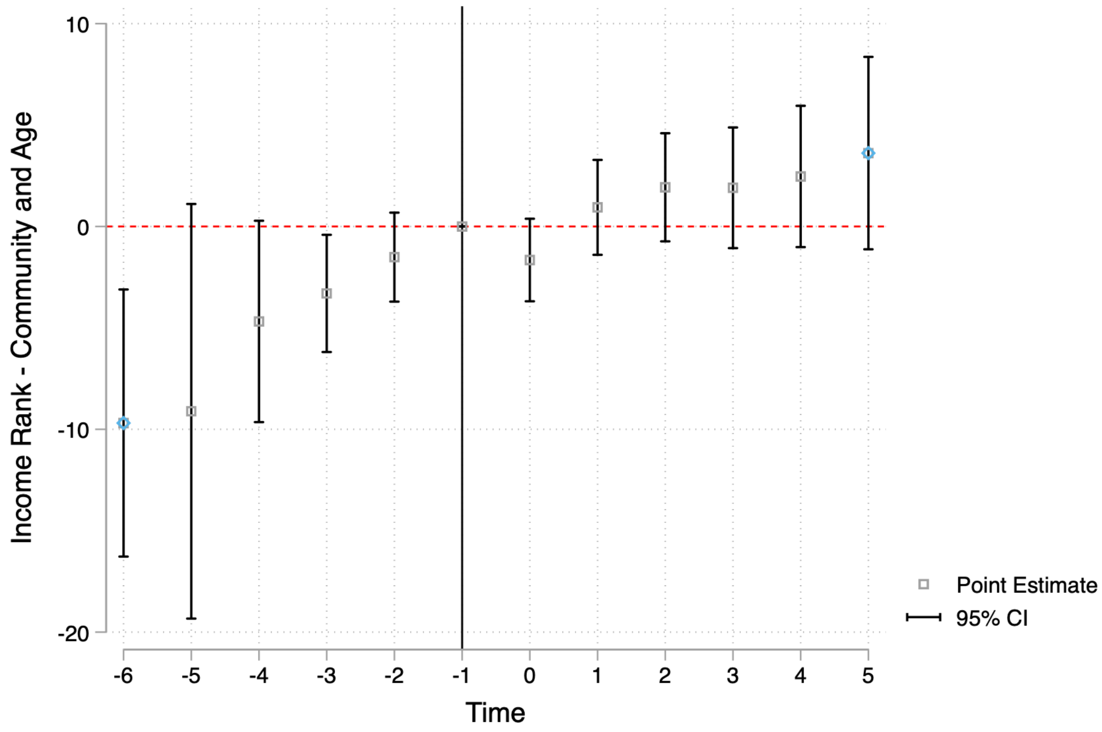
Figure 2.2: Event Study Estimates of the Effect of JUNTOS on Total Income Rank (Community as Reference)



F-statistic of pre-treatment trends = 1.18 ; p-value=0.3163

Note: Event Study estimates of the effect of JUNTOS on Income Rank with the Community as a Reference along with 95% confidence intervals and standard errors clustered at the district level. Graph includes individual and household controls. Crucial for the identification of parallel trends before treatment is the impossibility to reject that the coefficients of pre-intervention effects are equal to 0. The F-statistic for a joint significance test of pre-intervention trends is reported below the graph and suggests parallel trends. Conversely, for the post-treatment period, the F-statistic of joint significance is 3.51 (p-value<0.01), suggesting joint effects different than 0. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

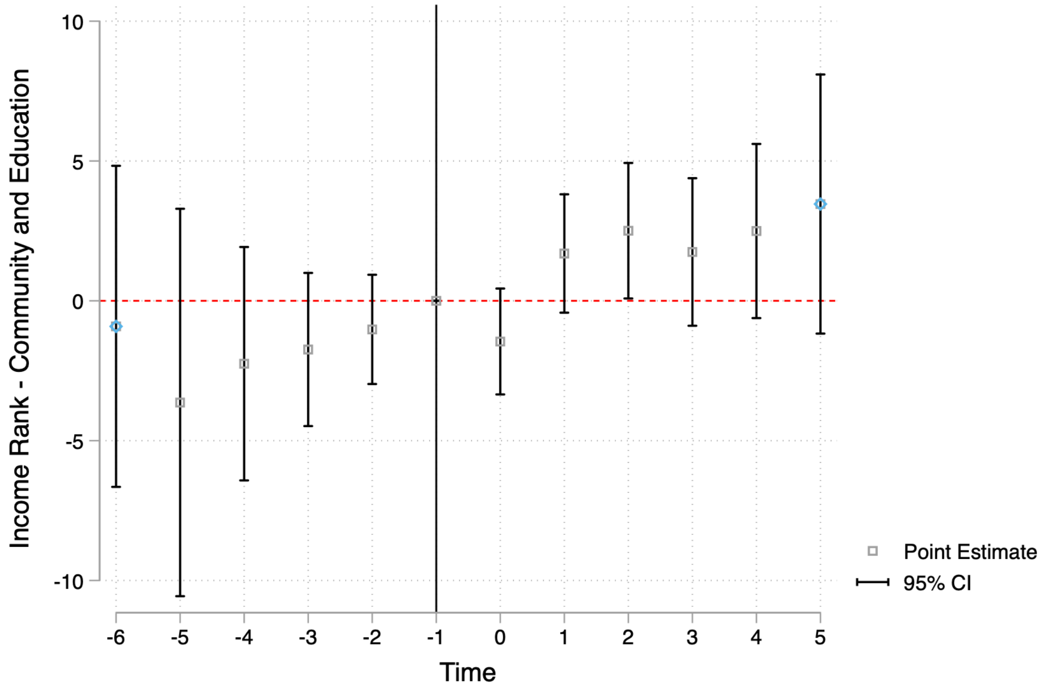
Figure 2.3: Event Study Estimates of the Effect of JUNTOS on Total Income Rank (Community and Age Group as Reference)



F-statistic of pre-treatment trends = 2.33 ; p-value=0.0407

Note: Event Study estimates of the effect of JUNTOS on Income Rank with the Community and Age Group as a Reference along with 95% confidence intervals and standard errors clustered at the district level. Graph includes individual and household controls. Crucial for the identification of parallel trends before treatment is the impossibility to reject that the coefficients of pre-intervention effects are equal to 0. The F-statistic for a joint significance test of pre-intervention trends is reported below the graph and does not suggest parallel trends. For the post-treatment period, the F-statistic of joint significance is 2.71 (p-value<0.05), suggesting joint effects different than 0. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

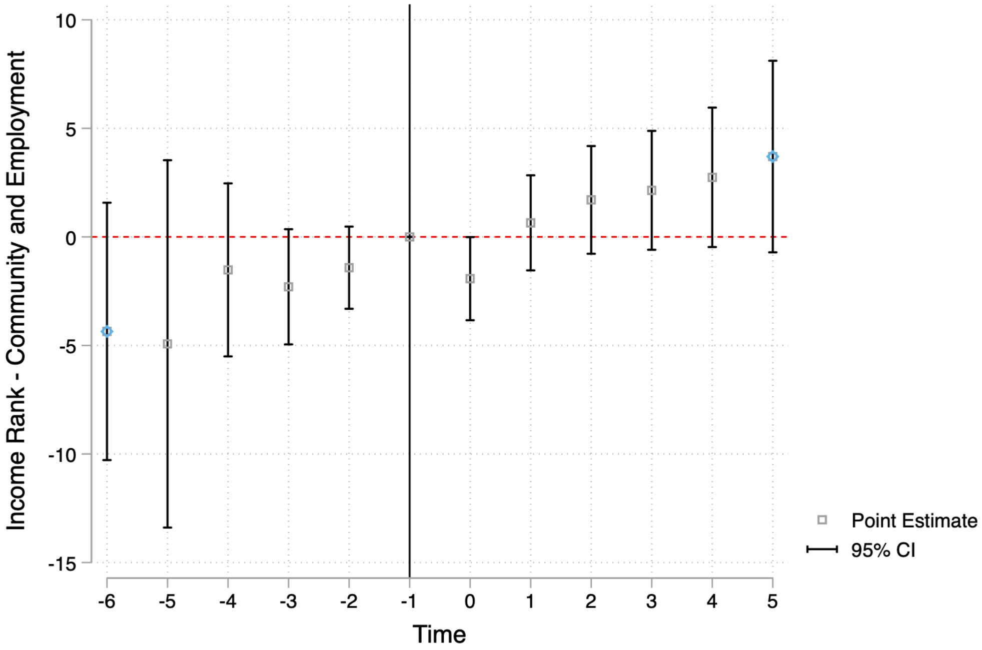
Figure 2.4: Event Study Estimates of the Effect of JUNTOS on Total Income Rank (Community and Education Level as Reference)



F-statistic of pre-treatment trends = 0.58 ; p-value=0.7183

Note: Event Study estimates of the effect of JUNTOS on Income Rank with the Community and Education Level as a Reference along with 95% confidence intervals and standard errors clustered at the district level. Graph includes individual and household controls. Crucial for the identification of parallel trends before treatment is the impossibility to reject that the coefficients of pre-intervention effects are equal to 0. The F-statistic for a joint significance test of pre-intervention trends is reported below the graph and suggests parallel trends. Conversely, for the post-treatment period, the F-statistic of joint significance is 3.52 (p-value<0.01), suggesting joint effects different than 0. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Figure 2.5: Event Study Estimates of the Effect of JUNTOS on Total Income Rank (Community and Employment Status as Reference)



F-statistic of pre-treatment trends = 1.00 ; p-value=0.4149

Note: Event Study estimates of the effect of JUNTOS on Income Rank with the Community and Employment Status as a Reference along with 95% confidence intervals and standard errors clustered at the district level. Graph includes individual and household controls. Crucial for the identification of parallel trends before treatment is the impossibility to reject that the coefficients of pre-intervention effects are equal to 0. The F-statistic for a joint significance test of pre-intervention trends is reported below the graph and suggests parallel trends. Conversely, for the post-treatment period, the F-statistic of joint significance is 3.59 (p-value<0.01), suggesting joint effects different than 0. Data comes from the Peruvian National Household Survey (ENAHO) and from JUNTOS administrative data.

Appendix

Algorithm for the calculation of the likelihood of being poor

Step 1

We run the following logit regression on a pool of households taken from the Peruvian National Households Survey (ENAHO) between 2001 and 2004:

$$\text{Logit: } Y=\alpha + \beta X + \mu$$

Where the dependent variable, *Y*, is constructed as follows:

Y=1, if the household is considered poor
Y=0, if the household is not poor

On the other hand, α is a constant, μ represents the error term and *X* consists on the following exogenous variables:

1. analf_m: percentage of illiterate adult women within the household
2. edu_men: percentage of underage members within the household that attend any type of schooling
3. combust0: access to industrial sources of cooking fuel (gas, kerosene, electricity)
4. no_equip: number of absent household appliances
5. serv3: possession of public lightning service, water supply and toilet services
6. tipom2: dwelling type 2
7. tipom3: dwelling type 3
8. tipom4: dwelling type 4

The regression results are:

Variable	Coefficient
analf_m	1.1832
	[12.66]***
edu_men	0.2276
	[5.13]***
combust0	-0.7624
	[12.84]***
no_equip	0.4446
	[27.40]***
serv3	-0.3769
	[3.23]***
tipom2	-0.2593
	[5.55]***
tipom3	-0.8584
	[14.86]***
tipom4	-1.3172
	[17.53]***
Constant	-1.3461
	[12.48]***
Observations	17980
Pseudo R2	0.182

Absolute value of Z-statistic in brackets
* statistical significance at 10%; ** statistical

significance at 5%; *** statistical significance at 1%

Step 2

We create the following variables from Questionnaire 100, Population Characteristics:

Objective

Takes value of 1 if person is “under 14 years of age” or “pregnant women between 12 and 49 years old”, 0 otherwise, starting in question 115.

Adult women illiteracy

Takes a value of 1 for a woman, over 18 years of age, who doesn’t know to write or read, according to question 108. Takes a value of 0 otherwise.

Underage household members attending school

Takes a value of 1 if there is an underage household member who is currently attending any type of school, according to question 110, 0 otherwise.

Step 3

We aggregate the household database with the sum of adult household members, underage members, illiterate adult women and underage household members attending school.

Afterwards, we filter the database by those households complying with the **Objective** of having at least one household member under 14 years of age or any pregnant woman.

Step 4

We construct ratios for:

- Adult illiterate women as a percentage of total adult household members, and
- Underage household members attending school as a percentage of the total of underage household members.

Step 5

We obtain the following variables from Questionnaire 200, Household and Dwelling Characteristics:

No equipment

Indicates the total amount of equipment not possessed by a household. Takes values from 1 to 7. Equipment is defined as: black and white TV, color TV, refrigerator, electric iron, gas kitchen, motorized vehicle, pedal vehicle.

Services

Takes values from 1 to 3, depending on whether the households has public lightning service, public water service, and toilet.

Fuel

Takes a value of 1 if the fuel most used by the household to cook has an industrial origin, 0 otherwise.

Step 6

We aggregate this database with the previous one. Then we filter those households that comply with the **Objective** and that show “survey complete”.

Posteriorly, if there were missing values for **No equipment, Services or Fuel,** we compute the district average so as not to lose information.

Step 7

We aggregate variables comprising dwelling types obtained by the predominant materials for walls, ceilings, and floors (questions 203, 204 and 205, respectively.

After generating 294 possible combinations of these materials, we chose 22 comprising 91.1% of all the existing dwellings. These were grouped in the following variables:

Variable	Type	Wall	Ceiling	Floor
Dwelling type 1	102	Adobe	Tile	Dirt
	126	Adobe	Straw	Dirt
	294	Mat	Straw	Dirt
	210	Stone w/mud	Straw	Dirt
	114	Adobe	Cane	Dirt
	168	Quincha	Straw	Dirt
Dwelling type 2	108	Adobe	Calamine	Dirt
	150	Quincha	Calamine	Dirt
	252	Wood	Straw	Dirt
	276	Mat	Calamine	Dirt
	113	Adobe	Cane	Cement
	101	Adobe	Tile	Cement
	192	Stone w/mud	Calamine	Dirt
Dwelling type 3	234	Wood	Calamine	Dirt
	107	Adobe	Calamine	Cement
	250	Wood	Straw	Wood
	106	Adobe	Calamine	Wood
	24	Brick	Calamine	Dirt
Dwelling type 4	232	Wood	Calamine	Wood
	23	Brick	Calamine	Cement
	5	Brick	Concrete	Cement
	233	Wood	Calamine	Cement

If there were missing values for **Dwelling Type** we compute the district average so as not to lose information.

Step 8

We multiply each generated variable by the coefficient found in the logit regression in **Step 1** and we find the logistic distribution expressed as *Y*, which gives the likelihood of a household being poor.

Step 9

We predict the likelihood of each household being poor and then we order them by that likelihood.

Knowing that rural poverty is 65.9% between 2001-2004, the threshold associated with that percentage is 0.7567447.

That threshold represents the cut above which households are poor, and under which they are not poor.

3 Chapter 3: Does it Pay to Bet on your Favourite to Win? Evidence on Experienced Utility from the 2018 FIFA World Cup experiment[19]

Published as Kossuth, Lajos; Nattavudh Powdthavee; Donna Harris; and Nick Chater (2020) **'Does it pay to bet on your favourite to win? Evidence on experienced utility from the 2018 FIFA World Cup experiment'**, *Journal of Economic Behavior & Organization*, 171, pp. 35-58, DOI: https://doi.org/10.1016/j.jebo.2020.01.006 [20]

With Nattavudh Powdthavee[21], Donna Harris[22], and Nick Chater[23]

This study examined whether people gained significant emotional benefits from not engaging in emotional hedging –betting against the occurrence of desired outcomes. Using the 2018 FIFA World Cup as the setting for our exploratory study, we found substantial reluctance among England supporters to bet against the success of the England football team in the tournament. This decision not to offset a potential loss through hedging did not pay off in people's happiness following an England win. However, it was associated with a sharp decrease in people's happiness following an England loss, which was a similar experience among subjects who were randomly assigned to bet for an England win. Post-match happiness was relatively more stable among those who chose to hedge or were randomly allocated to hedge. We conclude that people do not hedge enough partly because they tend to overestimate the expected diagnostic cost of betting against their social identity, while underestimate the negative emotional impact from betting on their favourite to win when they did not win.

[19] We are grateful to Jiaqi Wu and Danny Ng for their help with the data collection, and for comments from the participants at the DR@W forum at the University of Warwick. We received the IRB approval from the Humanities and Social Sciences Research Ethics Committee at the University of Warwick on 15/05/2018 (Ethics Application Reference: 109/17-18). Support from Warwick Business School's personal research fund is gratefully acknowledged. NP got the idea for this paper from a comment he received from Dan Hauser when he was working on the paying for transparently useless advice paper in 2011.

[20] Article received by JEBO on September 8th, 2019; revised on December 2nd, 2019; and accepted on January 9th, 2020.

[21] Warwick Business School, University of Warwick (email: Nattavudh.Powdthavee@wbs.ac.uk)

[22] Department of Economics, University of Oxford (email: donhatai.harris@economics.ox.ac.uk)

[23] Warwick Business School, University of Warwick (email: Nick.Chater@wbs.ac.uk)

3.1 Introduction

In remarkable research, Tang *et al.* (2017) and Morewedge, Tang and Larrick (2018) have shown that people are strongly averse to emotional hedging, i.e., betting against the occurrence of identity-relevant outcomes, despite doing so would minimize both financial and emotional losses for the individuals if the desired outcomes do not occur. For example, when given a choice to bet against the success of their preferred U.S. candidates and sports teams, most people would instead maximize potential gains and losses by betting for their favourite candidates or teams to win, i.e., putting "all their eggs in one basket", than opting to minimize gains and losses by betting against them[24].

The above evidence of disloyalty aversion is important to the economics literature as it suggests a possible violation of the standard utility theories, which predict that people generally have a desire to minimize the risk of potential losses rather than to maximize potential gains (e.g. Bernoulli (1954); Fischer *et al.* (1986)). This counterintuitive behaviour is also not unique to the betting market; evidence of disloyalty aversion reported in Tang *et al.* (2017) and Morewedge, Tang and Larrick (2018) is consistent with the puzzling evidence of equity home bias in finance, in which investors tend to over- invest in domestic assets (e.g., Cooper and Kaplanis (1994); Kang and Stulz (1997); Coval and Moskowitz (1999)) or in their own company stocks (e.g., Benartzi (2001); Meulbroek (2005)) even when it is significantly less risky to diversify.

What explains why many people are reluctant to bet against identity-relevant outcomes? According to Morewedge, Tang and Larrick (2018), emotional hedging induces a motivational conflict between a short-term monetary gain and long-term benefits of staying loyal to one's identity, which has been shown to be a psychologically discomforting experience for the individual to go through (see, e.g., Elliot and Devine (1994); Hogg *et al.* (2004)). Given that people may interpret hedging as a signal that they are not as committed to their identity as they believed they are, the expected diagnostic (or psychic) cost of hedging associated with the negative self-signal regarding their identity may be large enough that it outweighs the expected gains in utility associated with the payout from the hedge. As a result, individuals may want to reduce this psychological discomfort and bet on their preferred outcome instead of hedging.

[24] Morewedge, Tang and Larrick (2018) research showing disloyalty aversion and home bias in sports betting has since been replicated in other studies looking at betting behaviours in other sports such as hockey (Stanek, 2017) and soccer (Na et al., 2019).

From an economic perspective, one could argue that the desire to minimize the expected diagnostic cost of hedging makes the strategy of not hedging a utility-maximizing strategy once we take the expected diagnostic cost into account. Nonetheless, surprisingly very little is known about the success of either strategy –hedging versus not hedging –at maximizing experienced utility during as well as following the realization of the outcome (Daniel Kahneman et al., 1997). Whether or not our decisions produce the kind of experiences that we expect seems like essential information for decision-makers to know, given that the decision of whether to engage in emotional hedging is made mostly on hedonic grounds.

In this study, we argue that people may not be making optimal decisions for their experienced utility by not engaging in more emotional hedging. Specifically, we argue that people's decision utility (or "wantability"), which reflects their reluctance to engage in emotional hedging, is unlikely to be matched by their experienced utility afterwards. Our argument is motivated by many writings in psychology that argue that human beings regularly make prediction errors about their future hedonic experiences from the choices that they make today (Gilbert and Wilson (2000); Wilson and Gilbert (2005)). For example, in a study by Wilson *et al.* (2000), they showed how assistant professors tend to overpredict how happy they would be if they were to receive tenure, whereas former assistant professors who had achieved tenure were no happier than former assistant professors who had not. Additionally, they also showed how voters whose preferred political candidates were victorious in the election were not as happy as they had predicted to be the week after the election, whereas voters whose preferred political candidates had lost the election were nowhere near as unhappy as they had predicted to be.

One possible cause of this prediction error is the focalism bias, which describes people's tendency to focus too much of their attention on the occurrence of the focal event and fail to consider the consequences of other events are likely to occur (Schkade and Kahneman (1998); Wilson *et al.* (2000); Odermatt and Stutzer (2019)). For example, when people reflect on how much happier they would be if they were more affluent, they tend to focus too much of their attention on what money could buy them and little on what they would need to do in order to become richer, e.g., spending more time commuting and working. As a result, they may end up allocating too much of their time engaging in activities that are likely to increase stress and tension in the pursuit of higher incomes, thereby making them not nearly as happy being richer as they had expected to be (Daniel Kahneman et al., 2006). Interrelated to that concept are those of impact bias, or the tendency to overestimate how impactful a future emotional state might be (Gilbert et al., 1998), and projection bias, or the proclivity to exaggerate how much future preferences might resemble present ones (Loewenstein et al., 2003). In short, evidence suggests people tend to misconstrue the future cognitively and affectively.

We hypothesize this also might apply in the case of emotional hedging. Here, the expected diagnostic cost associated with the negative self-signal of hedging is likely to be a focal point (from which to assess affective impact or projection) for many individuals when predicting the effect of betting on or against outcomes that are relevant to their identity on their future happiness. By focusing too much on how bad they expect to feel about betting against identity-relevant outcomes, people may be making little provision for what could happen to their future happiness if they choose not to hedge and then the undesirable outcomes occur. Similarly, they may also be making little provision for how happy they would feel if they decide to hedge, and then the desired outcomes occur. While theories on motivational conflict predict a considerable psychic cost for negative self-signaling against one's identity (Elliot and Devine (1994); Hogg *et al.* (2004)), because of loss aversion (or the human tendency to prefer avoiding loss to acquiring equivalent gains), the effect of a combined loss, i.e., the occurrence of an undesired outcome and losing the bet, on their hedonic experiences may be just as great if not greater (see, e.g., Boyce *et al.*, (2013)) . In addition to this, expected regret may also contribute to a greater feeling of disappointment following a decision not to hedge an unfavourable outcome (Bell, 1982). If the evidence suggests that people do indeed overestimate the negative diagnostic cost on their future hedonic experience, then it may help to explain why people are reluctant to hedge desired outcomes even when it may be more optimal for their future happiness to do so.

Our study attempts to contribute to the emotional hedging literature by investigating whether people make optimal decisions for their experienced utility when deciding to either bet on or against identity-relevant outcomes. Specifically, we conducted a small-scale online study of the betting behaviours and the hedonic experiences of England supporters during one of the most-watched sporting events in the world: the 2018 FIFA World Cup. For the first time in 28 years, England football team reached the semi-final stage in this tournament. Not only their incredible performance surpassed all our expectations, but it also enabled us to repeat our experiment six times, with three wins and three losses in total. We carried out four main hypothesis testing in the current study. First, we examined whether England supporters would be reluctant to bet against England beating their opponent during their time in the World Cup. Second, we investigated whether changes in people's happiness vary significantly across betting decisions –regardless of whether they were endogenously or exogenously determined –following either a desired (England win) or an undesired (England loss) outcome. Third, we tested the accuracy of England supporters' predictions of their future happiness in the case of desired and undesired outcomes, and fourth, whether we could use the extent of people's prediction errors to predict their betting choices.

3.2 Methodology

3.2.1 Overview of the experimental design

To investigate whether people generally engage in emotional hedging during the 2018 FIFA World Cup, we conducted a series of lab-in-the-field experiments on voluntary participants at the University of Warwick between mid-June and early July in 2018. There were three stages to our experimental design. In the first stage, which was conducted two weeks before the World Cup started, we recruited volunteered participants by sending emails to students registered in the SONA system (Sample 1; N = 338) and asked them to fill in an online questionnaire generated by Qualtrics that was designed to elicit their general attitudes towards the World Cup[25]. Using a lottery prize draw as an incentive, we asked our participants to self-report on the extent to which they were looking forward to the World Cup, the team that they usually supported, the team that they thought would win the World Cup, their nationality, and their gender[26]. The logic behind this procedure was to screen out subjects who had little or no interest in, or knowledge about, the upcoming World Cup. This produces a subsample of 94 subjects (Sample 2), who (1) were looking forward to the World Cup, and (2) had explicitly declared England either as their first, second or third favourite team in the event[27].

In the second stage of the experiment, subjects in Sample 2 were sent an online, pre-match questionnaire 24h before the start of each England match. They were first asked the following question about their current happiness, widely used, and validated thoroughly in the literature.

[25] See Appendix for the screening questionnaire.

[26] We also asked participants questions on the strength of their support for their teams on a scale of 1 "Rarely ever follow" to 7 "Die-hard fan"). However, there are a lot of inconsistencies with respect to how participants respond to these questions. For example, we would expect the within-person rating of support for the 1st favorite team to be greater than the rating of support for the 2nd favorite team, and the rating of support for the 2nd favorite team to be greater than the rating of support for the 3rd favorite team. However, we would often find that the rating of support for the 2nd favorite team is greater than the rating of support for the 1st favorite team. This might be due to a variety of reasons, including the way we framed the question. It might be that although participants know the ordinal ranking of team preferences, they may not have the same opportunity to follow the matches of their top favorite team compared to the matches of their less preferred teams. It might also arise from a downwards adjustment of expectations: even though they prefer their own national team to win, they believe it has a low probability of success, and thus they show more support to those teams more likely to win, according to them. Moreover, almost a quarter of England supporters in our sample did not respond to this question. As a result, we have decided not to focus on this variable as a measure of strength of team identity.

[27] 72% of Sample 2 stated England as their top team.

"In general, how happy would you say you are these days? 1. Extremely unhappy, …, 7. Extremely happy."

In this second stage we also elicited each subject's predicted happiness 24h following (i) an England win, (ii) an England loss; and (iii) a draw. We also collected each participants' attitudes towards risk, as well as the reasons for placing the bet they placed.

In the final (third) stage, all subjects were sent a post-match questionnaire[28] to be completed within 24h following the conclusion of the match in each round. Included in the survey were questions about their current happiness in general, current happiness specific to the outcome of the match, whether they watched the match, as well as feelings of regret, their gender, and their nationality. Payment for participation was £2 for the completion of each questionnaire (pre- and post-match). It is important to note here that subjects were only paid participation fees if they had completed both questionnaires (so they either received £4 or nothing).

3.2.2 Experimental conditions

Subjects were randomly assigned to one of three treatment groups in the second stage of the experiment: the "**Free choice**", the forced-choice "**Bet for England**", and the forced-choice "**Bet for England's opponent**". There were two primary reasons for our decision to randomize subjects into either one of the forced-choice options, as well as the free choice group. Firstly, subjects' betting decisions and their experienced happiness in the free choice group are potentially endogenous to different unobserved individual characteristics, i.e., latent individual fandom, for example. And secondly, we had anticipated that the number of people voluntarily choosing to hedge in the free choice group would not have been large enough to make meaningful statistical inferences. Hence, the decision to randomize subjects into betting for England's opponent to win allowed us to estimate the effect of hedging on post-match experience for those who, without the randomization, would not have hedge had they been given a free betting choice. Randomization was stratified: subjects were twice as likely to land on the "**Free choice**" treatment group since there are four possible betting choices that the subjects could make in this experimental condition, and we needed to be able to have sufficient number of observations for each of these four choices in order to conduct any meaningful empirical analysis.

[28] Pre- and post-match questionnaires are found in the Appendix.

In order to minimize learning and streak-related behaviours that might arise from facing the same choice several times in a row, we repeated the stratified random assignment of subjects into different treatments each time before the start of each England match in the World Cup. Hence, depending on the luck of the draw, each subject could have been in one treatment for one of the England matches, and another in another one of the England matches. There were six England matches in total[29]: England vs. Tunisia (1st Group stage); England vs. Belgium (3rd Group stage); England vs. Colombia (2nd round); England vs. Sweden (Quarterfinals); England vs. Croatia (Semi-finals); and England vs. Belgium (3rd/4th playoff).

Following the general happiness question in the pre-match questionnaire, subjects in the "**Free choice**" treatment were given a £3 endowment, which they were asked to decide whether to:

- Keep the £3 endowment;
- Bet for England to win;
- Bet for England's opponent to win; or
- Bet for a draw.

In the betting decisions, the whole amount (£3) was used. However, given that these betting choices are endogenous to unobservable individual effects –for example, people who are not as emotionally attached to their preferred teams are more likely than others to hedge-, we also have two treatments that randomly assigned choices to subjects. In these treatments, subjects were told that they are being given £3 as an endowment. And as a result of a fair coin flip (Head = England, Tail = England's opponent), we have followed the coin's random outcome and put their entire £3 for England to win in the next match in the "**Bet for England**" treatment or for England's next opponent to win in the next match in the "**Bet for England's opponent**" treatment. Hence, we can treat these two treatments as exogenous, especially the "**Bet for England's opponent**" treatment as more subjects are likely to bet for England if they were given the choice.

Subjects were told before making their bet that the potential payments were taken from the odds generated by Bet365.com, which is one of the leading bet providers in the UK, for

[29] It should also be noted here that the other group stage match was England vs. Panama, where England were overwhelmingly clear favourites to win. We decided not to run our experiment using this game as it would have been very unlikely to see any type of insurance in this match. In the end, they ended up winning 6-1.

each match as of the day the questionnaire was sent. In other words, we incentivized our participants to place a bet with real-life odds as payoffs in mind. For example, in the England vs. Tunisia match, participants had the opportunity to win £27 from a £3 bet if they bet for Tunisia to beat England (and Tunisia beat England).

Ninety-four individuals took part in the second and third stage of the experiment in at least one of the England matches. The average participation rate was 4.74 matches (s.d. 1.52), and median participation was 5 matches. There were 349 observations in total: 192 observations in the "**Free choice**" treatment, 74 in the forced-choice "**Bet for England**" treatment, and 83 in the forced-choice "**Bet for England's opponent**" treatment. The average payoff for subjects in the "**Free choice**" treatment was £6.64; £6.14 in the "**Bet for England**" treatment; and £6.65 in the "**Bet for England's opponent**" treatment.

3.2.3 Econometrics specifications

There are three main empirical specifications. The first specification, which focuses on the "**Free choice**" treatment as illustrated in equation (1), estimates the following multinomial logit equation of betting decision with robust standard errors clustering at the individual level in order to determine what predicts betting choices:

(1) $$BD_{it} = \alpha + \beta_1 O_{it} + \beta_2 E_{it} + \beta_3 Y_{it} + \beta_4 X_{it} + \varepsilon_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, 6,$$

where $\boldsymbol{BD_{it}}$ denotes the betting decision for subject ***i*** in period ***t***; $\boldsymbol{O_{it}}$ represents the winning odds provided by Bet365.com for both England and England's opponent (the baseline odds are the odds that the two teams will get a draw within the 90 min of playing time); $\boldsymbol{E_{it}}$ is a dummy variable representing whether England won their previous match; $\boldsymbol{Y_{it}}$ is the accumulated payment for each subject ***i*** up to period ***t***; $\boldsymbol{X_{it}}$ is a vector of other control variables that includes dummy variables representing the stage of the match (group stage vs. knock-out stage), having stated England to be the first favourite team during the screening process, gender, nationality, as well as pre-match happiness, subjective risk profile, accumulated payment, and their reasons behind their betting decision; and $\boldsymbol{\varepsilon_{it}}$ is the error term. Given that tiny number of people in the "**Free choice**" treatment chose the draw option, we have decided to exclude the "Draw" category from the multinomial logit estimation and only focus on the three possible outcomes: keep £3, bet for England, and bet for England's opponent.

Following Morewedge, Tang and Larrick (2018) empirical strategy, we also add as additional controls each subject's reasons behind their betting decision. The variables were derived by asking subjects to state their level of agreement on a 7-point scale (1 = "strongly disagree", …, 7 = "strongly agree") to the following six possible reasons behind their decision to bet the way they did: (i) highest chance of winning; (ii) paid the most money, (iii) want to hedge my chance; (iv) have something to be happy about; (v) want to be loyal, and (vi) will not enjoy money if the other team win.

The second specification estimates the effects of different betting decisions on changes in pre- and post-match general happiness, which we estimate separately for sub-samples of England win and England loss. Our main objectives here are to investigate whether (i) the decision to bet for England to win has a significant psychic benefit following an England win, but a severely adverse psychic effect following an England loss, and (ii) the decision to bet for England's opponent to win has a negative psychic effect following an England win, but buffers any future emotional losses following an England loss. Equation (2), which can be written as follows:

$$(2) \qquad \Delta H_{it} = \theta + \gamma_1 BD'_{it} + \gamma_2 P'_{it} + \gamma_3 Z'_{it} + \mu_i + \epsilon_{it}, \quad i = 1, \dots, N; t = 1, \dots, 6,$$

where $\boldsymbol{\Delta H_{it}}$ denotes the difference between post- and pre-match general happiness for subject $\boldsymbol{i}$ in period **t**; $\boldsymbol{BD'_{it}}$ is a set of dummies of different betting decisions; $\boldsymbol{P'_{it}}t$ represents a vector of accumulated payment, individual payment, and maximum potential payment from betting; $\boldsymbol{\mu_i}$ is the unobserved individual fixed effects; and $\boldsymbol{Z'_{it}}$ is a vector of control variables that includes all of the control variables in Equation (1) (except for subject's reasons behind their betting decision), plus a dummy for whether the subject watched the match in question. Because subjects' happiness is likely to be influenced not only by the outcome of the match but also by their income gains or losses from winning or losing the bet, we first estimate Equation (2) without the payment variables (e.g., accumulated payment, payment received after the match, and maximum potential payment from betting) as controls. This enables us to gage the extent to which the effects of different betting strategies on changes in subjects' happiness are confounded by how much money they won or lost. We then re-estimate Equation (2) with the payment variables as controls in order to see the effects of different betting strategies on changes in subjects' happiness that are independent of how much they earned from the match.

The third and final specification, Equation (3), replaces Equation (2)'s post-match general happiness with post-match happiness that is specific to the match, i.e., "*How happy do you feel about the outcome of the match between England and their opponent?*"

(3) $$\Delta\widehat{H}_{it} = \rho + \delta_1 BD'_{it} + \delta_2 P'_{it} + \delta_3 Z'_{it} + \tau_i + \upsilon_{it}, \quad i = 1, \dots, N; t = 1, \dots, 6,$$

where $\boldsymbol{\Delta\widehat{H}_{it}}$ denotes the difference between specific post- and general pre-match happiness for subject $\boldsymbol{i}$ in period $\boldsymbol{t}$. Following a suggestion by Ferrer-i-Carbonell and Frijters (2004) to always allow for unobserved heterogeneity in happiness regressions whenever possible, all happiness regressions in this study are estimated using both random-effects and fixed-effects estimators with robust standard errors clustered at the individual level. It is worth noting here that the correlation between post-match general happiness and post-match happiness specific to the match is 0.31, which suggests that there is only a moderate correlation between these two subjective variables.

3.3 Results

3.3.1 Do people engage in emotional hedging?

To make the first pass at this question, we present in Figure 3.1 the raw data of betting decisions in the "Free choice" treatment. Consistent with what would have been predicted by the social identity theory and previous findings in the literature (e.g., Stanek, 2017; Morewedge, Tang and Larrick, 2018), we find that more than half of the participants bet for England to win (52.6%) and only a minority (17.2%) of the England supporters in our sample chose to bet for the opponent to win. Qualitatively similar results are obtained in Figure 3.2 A-B when we divide the sample into the group and the knockout stages of the World Cup: in both stages, the highest proportion of subjects continued to bet for England to win (40% and 59% in the group and the knockout stages respectively). There is, however, a small but important difference between these two stages of the tournament. As both figures show, the proportion of subjects who bet for England is notably higher during the knockout stage.

By contrast, the proportion of people who insured themselves against the possibility of England losing (betting for the opposing team to win) is almost the same in the knockout stage as it was in the group stage. One explanation for this is that the participants' social identity might have been further reinforced by an increase in the stakes experienced during the knockout stages.

To understand better what predicts these betting decisions (except for the small number of people betting for a draw), we estimate Equation (1)'s multinomial model and report the results in Table 3.2. Looking across the first panel of results (Model 1), we find little statistical support that the higher the odds of England winning, i.e., the lower the probability that England would win the match, the less likely it is for subjects to bet for England to win compared to the decision of keeping the money (the reference group). The same nonsignificant finding is obtained for the odds of England's opponent winning and pre-matched happiness. There is also little statistical evidence to suggest that subjects would bet for an England to win their next match if England had already won in the previous match. By contrast, there is strong evidence that subjects are significantly more likely to bet for England to win than keeping the money during the knockout stage. Additionally, people who are relatively more risk-loving are also more likely to bet for either England or England's opponent to win than keeping the money, whereas subjects who listed England as their top favourite team were highly unlikely to bet against England winning their next match. On the other hand, the observable characteristics such as gender and nationality, appear to have very little predictive power of subject's betting decision.

In the second panel of Table 3.2 (Model 2), we find that one of the main reasons behind the subjects' decision to bet against their team was because doing so paid the most money. Consistent with the social identity theory and betrayal aversion (Bohnet et al., 2008), the decision to bet for England to win tends to be correlated with the drive of not wanting to hedge, as well as the thought of not enjoying the money if the other team wins. Moreover, we find that the reason to have something to be happy about is positively and statistically significantly correlated with the decision to hedge, thus suggesting that people may have made a systematic calculation before placing a bet that betting against England will help alleviate their future emotional damages should England go on to lose their match. Interestingly, we find that the reason for wanting to hedge are negatively and statistically significantly correlated with both betting for and against England winning. This suggests that subjects may have seen keeping the money as a form of 'soft' hedging as it is the only choice that allows them to keep the relatively small sum of money (£3), while at the same time not having to bet for or against their team identity.

Finally, there is a compelling argument to be made regarding the $3^{rd}/4^{th}$ place match being a consolation prize which teams aren't as eager to win (since they just lost the chance to win the World Cup). For instance, it might be perceived by subjects to be "non-important" in relation to both the tournament and their social identity, and therefore alter their observed patterns of betting decisions. We re-estimate Equation (1)'s multinomial model albeit transforming the dummy variable "Stage: Knockout" into a discrete one with 3 categories: the group stage, the proper knock-out stage, and the $3^{rd}/4^{th}$ place match. We find qualitatively similar results as in the main estimations (subjects are more likely to bet for England to win in the proper knock-out stages and the $3^{rd}/4^{th}$ place match, compared to the group stage), which suggests subjects do not deem this match as "non-important". Indeed, we also replicate the calculations for Figures 3.1 and 3.2 (A-B) without including the $3^{rd}/4^{th}$ place match observations and we find a similar distribution of bets across the board, which reinforces the fact that subjects do not perceive the $3^{rd}/4^{th}$ place match to be qualitatively different than the others.

3.3.2 Can hedging insure against future emotional loss?

We have seen that people typically do not engage in emotional hedging by betting on their non-preferred team. However, if they do, or if hedging is imposed upon them, does it make them unhappy if their team wins, but then reduce the feeling of disappointment if their team loses? In this section, we test whether hedging worked in insuring individuals against future disappointment following an England loss in both the free and forced-choice treatments. We first present in Figure 3.3 A-B raw data evidence of the mean pre-match general happiness, post-match general happiness, and post-match happiness specific to the outcome of the match by betting decisions for the matches that England won and lost.

In the three matches that England won illustrated in Figure 3.3 A, we can see from the overlapping standard errors that the average post-match general happiness levels are statistically insignificantly different from the average pre-match general happiness. On the other hand, with the exception for betting for a draw, the average post-match happiness specific to the outcome of the match is generally higher than (or the same as) the pre-match general happiness, including those who chose to engage in emotional hedging by betting for England's opponent to win. These preliminary results thus suggest that the decision to hedge identity-relevant outcomes did not seem to reduce our subjects' future happiness, both general and specific to the match, even if England went on to win their match and they ended up losing the bet.

What happened when England lost? In the three matches that England lost to their opponent illustrated in Figure 3.3 B, we can see that subjects did not report a drop in their general happiness after the match if they chose to either keep the money, bet for a draw, or bet for England's opponent to win. This suggests that the decision to bet for the opposite team or keep the money and refrain from betting altogether may have helped reduce future disappointment when England lost their match. Similarly, we also observe a significantly smaller drop in the average post-match happiness specific to the outcome of the match among those who chose or were randomly assigned, to betting for England's opponent to win. Not surprisingly, some of the most significant drops in the post-match happiness following an England loss were experienced by those who either chose or were randomly assigned to bet for England to win.

Table 3.3 conducts more systematic testing of the emotional hedging hypothesis by estimating Equation (2) on subsamples by matches won and lost. The dependent variable is each subject's post-match general happiness minus pre-match general happiness so that positive value in the dependent variable denotes a within-person increase in the general happiness following an England match, and vice versa for a negative value. We report both random-effects (RE) and fixed-effects (FE) estimates for the "**England won the match**" subsample in Columns 1–4, and for the "**England lost the match**" subsample in Columns 5 and 6.

Looking across Columns 1–4, there seem to be little differences in terms of changes in the general happiness pre- and post-match across free and forced bets when England won the match; none of the estimated RE and FE coefficients on free and forced bets is statistically significantly different from zero. This nonsignificant finding does not seem to depend on whether the payment variables were included or excluded from the model: subjects who bet for England to win did not become significantly happier than others following an England win, despite having also earned money from the match. By contrast, hedging –either as a choice or randomly assigned –does not correlate with a statistically significant drop in the general happiness following an England win when compared to keeping the money or betting for England to win.

For the "**England lost their match**" subsample, i.e., Columns 5–8 in Table 3.3, we find that betting for England to win is associated with a sizeable and statistically significant drop in the general happiness score of around 0.5-points in the RE model and 0.9-point in the FE model when compared to keeping the money. It is also noteworthy that a similar drop of around 0.7-point in general happiness is also observed in the FE specification for those who were forced to bet for England to win as well; see Column 8. It is also worth noting here that we cannot reject the null hypothesis that the two coefficients are the same in magnitudes. By contrast, there is zero statistical difference in pre- and post-match general happiness between keeping the money and betting for England's opponent to win –e.g., the estimated coefficient on having been forced to bet for England's opponent is −0.1 with robust a standard error of 0.37, which suggests that hedging helps to insure our subjects against the feeling of disappointment when England lost their match. Again, it is worth noting that we can reject the null hypothesis that the estimated coefficients on hedging are statistically significant different from the coefficient on betting for England to win at the 5% level.

There are a few other interesting results in Table 3.3. For example, an England loss in the knockout stages[30] hurts significantly more than a loss in the group stages, while the opposite is true for an England win in the knockout stages. This finding is not surprising, given that the stakes are much higher in the knockout stages than in the group stages. On average, an increase in the maximum potential payment that subjects could have won is associated with a statistically significant drop in happiness following an England loss. Finally, there is some evidence to suggest that women derived greater happiness than do men following an England win but are not significantly unhappier by an England loss compared to men.

Table 3.4 checks whether Table 3.3's results can be replicated with the post-match happiness specific to the match as the dependent variable instead of a general one. Here, we show that subjects who either chose to bet for England to win or had been randomly assigned to bet for England to win became statistically significantly less happy about the outcome of the match following an England loss in the FE regressions, but surprisingly not in the RE regressions; see Columns 7 and 8. By contrast, there is little evidence to suggest that subjects who bet for England to win, either by choice or randomly assigned, were statistically significantly happier with the outcome of the match following an England win. We are not certain why the results are statistically more robust in the FE than the RE regressions, but selection bias and attenuation bias that are associated with the unobserved fixed characteristics such as personality traits might be playing a role here. For example, it might be that people who were born with personality traits that make them happy are more reluctant to hedge. They may also be less likely to drop out from the experiment. Nevertheless, both RE and FE findings are consistent with Table 3.3's results, and we can conclude based on Tables 3.3 and 3.4's results that the decision to hedge –or merely having been randomly made to hedge –works as emotional insurance against the disappointment of a future loss.

We also conduct further robustness checks in Tables 3.5 and 3.6. Given that randomization of treatments occurred at the beginning of each round, it is likely that subjects would have been allocated different treatments throughout their time spent in the experiment. To account for their experience in other treatments, we included previous round's treatment as an additional control variable in Table 3.5. However, including this additional control did

[30] England lost twice in the knockout stages, one at the semi-finals (losing to Croatia) and the other at the 3rd/4th place playoff (losing to Belgium).

little to change the main results and the decision not to hedge continues to hurt psychologically following and England loss. Finally, we tested whether the decision not to hedge is associated positively with subject's level of enjoyment while watching the game and report the estimates in Table 3.6. Here, we find little evidence that people who bet on their favourite to win enjoyed watching the game more than those who engaged in psychological hedging. Hence, this suggests that there may have been little procedural utility to be gained from people's reluctance to hedge while watching their favourite team in a match.

Finally, we explore again the degree to which subjects perceive the 3rd/4th place match to be 'non-important' (in comparison to the rest of the matches of the World Cup). This time, we test whether our main results on the effects of emotional hedging change when making a qualitative differentiation between the group stage, the knock-out stage, and the 3rd/4th place match. The only difference we find is a loss of statistical significance in the coefficient associated with having bet for England to win and England losing the match, in the fixed effects post-match general happiness minus pre-match general happiness regressions without payments as covariates. In general, however, our main results still hold and we can rule out the 3rd/4th place match being perceived as qualitatively different: having bet for England does not make subjects happier following an England win, but it makes them miserable following an England loss.

3.3.3 Affective forecasting and decision errors in hedging

Results in Tables 3.3 and 3.4 suggest that betting for England may be a sub-optimal betting strategy at least in terms of psychological well-being as it made subjects significantly more miserable compared to other betting decisions following an England loss, while at the same time it did not make them significantly happier following an England win. Despite the emotional benefits of hedging, more than half of our observations in the "Free choice" treatment did not, however, choose to bet against their team identity.

What explains this mismatch in the pre-match choice and the post-match experience of our subjects? One hypothesis is that people's decisions are not driven by their desire to maximize hedonic states across time. Instead, people make decisions for a variety of reasons, including the desire to preserve one's identity (Morewedge et al., 2018), that may or may not have a payoff in terms of emotional experiences in the future. An alternative hypothesis is that people are indeed motivated by their emotional consequences from the decisions that they make today, which would be consistent with the recent studies that find people's choices to be primarily driven by how happy they think they will be in the future from making these decisions (see, e.g., Benjamin *et al.* (2012); Adler, Dolan and Kavetsos (2017). However, decision errors occur because people are not very good at predicting which decisions will make them happiest in the future, which would partly explain why people do not engage enough in emotional hedging.

Considering our findings, we propose two additional testable hypotheses to help explain why most people are averse to betting against their team identity. These are:

1) Team identity exacerbates people's inability to accurately forecast their future emotional experiences from betting for and against their team, and
2) Mispredictions of post-match happiness are among the main drivers of people's betting decisions.

We shed some light on these additional two hypotheses below.

3.3.3.1 Evidence of affective forecasting errors

According to studies in the affective forecasting literature, people are prone to making inaccurate predictions of future hedonic experiences of their decisions (Gilbert and Wilson (2000); Timothy D Wilson and Gilbert (2005b)). This inability to accurately forecast our future happiness can be explained in part by focusing illusion –i.e., the tendency to focus too much attention on specific aspects of an event while ignoring other factors (Schkade and Kahneman, 1998), impact bias –i.e., the tendency to overestimate the length or the intensity of future emotional states (Gilbert et al., 1998), and projection bias –i.e., the tendency to exaggerate the degree to which their future preferences resemble their current preferences (Loewenstein et al., 2003). What this implies is that, because of these cognitive biases, an individual's decision utility (or revealed preferences) may not always lead to the same experienced utility (or hedonic experiences) once the choices have already been made (see, e.g., Kahneman and Thaler (2006)).

Building on previous findings in the affective forecasting literature, we hypothesize that having a strong sense of team identity exacerbates our inability to accurately forecast the future emotional consequences of our choices, both positively and negatively. More specifically, because of impact and projection bias, England supporters will tend to overestimate the positive impact of an England win, as well as the negative impact of an England loss, on their future happiness.

Figure 3.4 A-B test these hypotheses by presenting the differences between subjects' pre-match general happiness, predicted general happiness during the 24 h after the match, and post-match general happiness during the 24 h after the match for England win and loss, respectively. Focusing on the "Free bet" sample first, we can see that the decision to bet for England to win is observed with one of the most substantial anticipated increases in their future happiness –from 5.16 in the pre-match happiness to 6.07 in the post-match happiness –if England won compared to other choices. However, the decision to bet for England to win is also observed with one of the most significant anticipated drops in their future happiness – from 4.95 to 3.07 –if England lost. Interestingly, we also find that subjects in the "Forced Bet for England" treatment also anticipated becoming significantly happier from an England win – from 5.35 to 6.22 –and significantly unhappier from an England loss –from 5.11 to 2.95. On the other hand, subjects in the "Forced Bet for England's opponent" even when they did not make the betting decision themselves. For these subjects who were forced to hedge, they anticipated becoming slightly unhappier from an England win –from 5.26 to 4.95 –as well as from an England loss –from 5.66 to 4.89.

Were these predictions accurate? Looking at both figures, we can see that an England loss did not seem to hurt people as much as they thought it would[31]. Also, an England win did not give them nearly as much joy as they had anticipated. What is perhaps most interesting, however, is that the extent of misprediction of future happiness is noticeably smaller among those who hedged when England went on to lose. This can be seen quite clearly in Figure 3.4 B: among those who bet for England's opponent to win, their average pre-match general happiness is 5.06, their average predicted happiness 24 h following an England loss in their next match is 5.00, and their average post-match (or realized) happiness 24 h following the conclusion of the match is 5.33.

In short, Figure 3.4 A-B provide raw data evidence that subjects who bet for England to win tend to make notably more significant affective forecasting errors than those who hedged, bet for a draw, or decided to keep the money. Hence, our results seem to suggest that subjects who strongly identify with their team are among the worst at predicting their future emotional experiences post-match.

3.3.3.2 Measures of affective forecasting as predictors of betting decisions

[31] It is good to recall here that we asked our subjects to make a prediction of their future happiness within 24 h following the conclusion of the match in question.

To what extent can our subjects' predictions of future happiness predict their betting decisions? One hypothesis, which is based on a model of behaviour in which people take into considerations their future emotional experiences before making a choice (Benjamin *et al.* (2012); Adler, Dolan and Kavetsos (2017)), is that people will be more likely to bet for England to win if they anticipate that happiness gain from an England win exceeds that of happiness loss from an England loss. On the other hand, people will be more likely to hedge and bet for England's opponent to win if they anticipate that happiness gain from an England win is smaller than that of happiness loss from an England loss.

We conduct a formal test of our hypothesis by re-estimating Equation (1) with (i) the reported difference between predicted happiness (win minus loss) as an additional explanatory variable in Model 1, and (ii) predictions of post-match happiness for win and loss as two separate additional explanatory variables in Model 2 and report the results in Table 3.7. Consistent with our hypothesis, Model 1's results suggest that the difference in the predicted happiness plays an integral part in determining subjects' betting strategies in the "Free choice" treatment. For instance, compared to keeping the money, the probabilities of betting for England's opponent to win or to bet for a draw is significantly lower for people who predicted to be significantly happier from an England win than unhappier from an England loss, i.e., predicted happiness gain minus predicted happiness loss < 0. The opposite is true for the probability to bet for England to win; the more significant the gap in predicted happiness, the more likely is the bet for England to win in the next game.

Qualitatively similar results are also obtained in Model 2. Here, we can see that predicted happiness from an England win increases the probability of making a bet over keeping the money, although it is more positively correlated with betting for England to win. By contrast, predicted happiness from an England loss is positively and statistically significantly correlated with the decision to bet for England's opponent, but not the decision to bet for England. This finding implies that subjects who anticipated to become significantly unhappy by an England loss were also significantly less likely to engage in emotional hedging, holding other things constant.

Tables 3.7's results also confirm that the odds related to England winning is not one of the main determinants of our subjects' betting strategy in any of our models. What seems to matter much more to the way we make our decisions is the predicted (or expected) happiness from different possible scenarios, as well as the desire to make the most money and to avoid betraying their team identity. Moreover, because the intense loyalty to their team identity heavily drives most of our subjects, the majority (approximately 72%) ended up overestimating the extent of happiness of an England win compared to the extent of the unhappiness of an England loss. This, we believe, helps to explain why significantly more people chose to bet for an England to win even when, given the risk and uncertainty of an eventual outcome, it is not emotionally beneficial to do so.

3.3.4 Constraints on generality

In this subsection, we take the opportunity to express what we believe to be the constraints on the generality of our findings (Simons et al., 2007). We have shown that most of our subjects do not engage in emotional hedging during the 2018 FIFA World Cup even when it may be emotionally beneficial to do so. Thus, we expect the results to generalize to other major sporting events in which subjects have a strong identity with the team(s) involved in the competition. While we do not have evidence that the findings will be reproducible outside the domain of sporting competition, we believe that our results would have been generalizable across non-sporting events where people have an identity-relevant outcome in mind, e.g., the E.U. Referendum or the U.S. Presidential Election. We also believe the results will be reproducible with students from similar subject pools serving as participants, although we have no evidence to suggest that the results will be reproducible for the general population. Finally, we have no reason to believe that the results depend on other characteristics of the subjects, materials, or context.

3.4 Discussions and limitations

There are a number of potential issues and limitations with regards to our findings. One objection is that our sample size may be too small to detect effects. Using a power analysis in STATA, the sample size required to obtain 0.80 power to detect a medium effect size of 0.75 at the standard 0.05 alpha error probability is approximately 47 observations in each group. With a repeated sampling design, we began with 192 observations in the "**Free choice**" group, 74 in the "**Bet for England**" group, and 83 in the "**Bet for England's Opponent**" group. However, given that we had no control over the outcomes of all England's matches, we were not able to predetermine the number of observations we would end up with in each group following an England win and an England loss, as well as the number of observations in each choice set in the "Free choice" group. As a result, the power to detect a medium effect size in our test ranges from 0.38 for the smallest group (bet for a draw) to 0.89 (bet for England to win), depending on the number of observations in each cell[32]. Nevertheless, we have not

[32] For instance, we simply did not have enough observations of people who bet for a draw in the "Free choice" group to detect any effect.

drawn any conclusions from these low powered choices. In addition to this, we are confident that we have sufficient number of subjects who chose to bet for England to win in the "**Free choice**" treatment, as well as those who were forced to bet for either England or England's opponent to win, to obtain between 0.7 and 0.89 power to detect a medium effect size. We also acknowledge that a better design would have been to conduct a series of survey experiments in which a large N of both England's supporters and the other team's supporters are surveyed. This would have guaranteed us a large number of observations of subjects who had experienced a win and a loss per England match. Unfortunately, it was impossible for us to find sufficient number of participants who supported each of England's opponents at the time to execute the above design.

A further concern is the possibility that our subjects modified their behaviours/responses to comply with what they think are appropriate behaviours/responses demanded by the experimenters (Zizzo, 2010). However, we believe that we had managed to circumvent or at least minimize the experimenter demand effect for the following reasons. First, given that participants were randomised into three different treatments in each round, it is unlikely that they would have figured out what the aim of our research was and, consequently, what the appropriate responses or behaviours were. Second, subjects had no way of knowing the outcome of each England's match before it was played. Since they were engaging in a real-life gamble with potential winnings at stake, it seems natural to assume that their incentives were to maximize potential earnings and not attempting to figure out the behaviours/responses that are demanded by the experimenters. Finally, since this was an online experiment, the subjects and the experiments had never met face-to-face, which in turn should reduce the possibility of an experimenter demand effect.

Another conceptual concern is that the affective forecasting results were not the outcome of some cognitive biases such as the impact bias or durability bias as suggested by Gilbert and Wilson (2000). Rather, it may be the case that affective forecasters strategically overestimate the hedonic impact of future events as a means to motivate them to obtain or avoid the forecasted experience; a strategy known as "motivated reasoning" (Morewedge & Buechel, 2013). For example, students may strategically overpredict the negative effect of failing an exam a motivation to work harder to avoid it. However, it is unlikely that motivated reasoning applies in the case of predicting future emotional impacts of a World Cup match. This is because, according to Morewedge and Buechel (2013), forecasters only exhibit motivated reasoning if they believe that their behaviours can influence the outcome; for example, like studying harder for an exam or training harder for a sports competition. As far as we know, there are no good reasons to believe that any of our subjects' forecasted affects have had any influence on England's fortunes in the competition[33].

3.5 Conclusion

Many of us have a strong preference to stay loyal to people whom we identify with socially. There are potentially many reasons for this, but one of the main reasons stated in the literature is that most of us expect that there would be a sizeable diagnostic cost on our utility that comes from betraying our social identity. Hence, our expectation of a negative psychological impact from the negative self-signaling explains why we tend to bet on our favourites to win in a competition even when it appears to be more rational to bet against rather than on desired outcomes (Morewedge et al., 2018; Tang et al., 2017).

In this chapter, we conducted a lab-in-the-field experiment to test whether we can justify people's decision not to hedge by studying their experienced utility (or ex post happiness). Using the 2018 FIFA World Cup as our case study, we first showed that most England supporters bet on England to win in each of the England matches, which is consistent with the evidence in the disloyalty aversion literature. However, from our analysis of the England supporters' pre- and post-match happiness, we found little evidence of any emotional benefits to betting on England to win when England went on to win the game. By contrast, we showed that people who bet on England to win tend to report a significant drop in happiness following an England loss. Additionally, we did not find evidence of a significant drop in happiness following an England win or a significant increase in happiness following an England loss among England supporters who chose to hedge or having been forced to hedge. We then demonstrated that this mismatch between people's decision utility and experienced utility is due partly to the fact that people often overestimate the size of the expected diagnostic cost of the negative self-signal associated with emotional hedging, while underestimate the negative emotional impact of betting on their favourite team to win when they did not win.

Like all studies in social sciences, our work is not without limitations. As previously mentioned, one of the biggest short-comings of our study is the relatively low power to detect effects. However, we believe that our findings have offered new insights into the puzzling findings of many home bias behaviours in the sporting and financial markets, and that betting on our favourites to win may not be the optimal strategy, at least in terms of happiness, even

[33] Although if there is provable evidence of any subjects exhibiting motivated reasoning here, we would love to buy them a drink as a thank you for whatever they did to get England to the semi-final of the World Cup for the first time in nearly thirty years.

when we get the results that we desired. Future research should therefore return with a larger sample size –at least 60 observations per group –and retest our hypotheses across different settings like the U.S. Presidential Election, for example.

Table 3.1: Summary statistics

	Treatment 1: Free choice	Treatment 2: Forced bet for England	Treatment 3: Forced bet for England's opponent
Pre-match happiness	5.13	5.22	5.46
	(1.24)	(1.11)	(1.14)
England as 1st favourite team	0.76	0.79	0.53
	(0.42)	(0.40)	(0.50)
Gender: Female	0.60	0.58	0.51
	(0.49)	(0.49)	(0.50)
Nationality: Other	0.17	0.17	0.34
	(0.37)	(0.38)	(0.47)
No. of unique individuals	76	48	57
N	192	74	83

Note: We cannot reject t-tests of equality (with standard errors clustered at the individual level) for any pair of characteristics across different treatments.

Table 3.2: Multinomial Logit Betting Decisions – All stages

	Model 1		Model 2	
	Bet for England to Win	**Bet for Other to Win**	**Bet for England to Win**	**Bet for Other to Win**
England won in the previous match	3.498	2.454	6.223	2.319
	(2.534)	(3.285)	(3.848)	(3.333)
Odds England Winning	2.064	1.177	4.035*	1.345
	(1.358)	(1.747)	(2.120)	(1.901)
Odds Other Winning	0.203	0.020	0.489	0.282
	(0.138)	(0.185)	(0.364)	(0.417)
Stage: Knockout	5.346**	3.269	9.609**	1.818
	(2.490)	(3.368)	(4.277)	(3.635)
Subjective risk profile	0.612***	0.629***	0.984***	0.845***
	(0.161)	(0.180)	(0.268)	(0.286)
England as a 1st favourite team	-0.863	-2.069***	-1.907	-3.415**
	(0.737)	(0.695)	(1.486)	(1.462)
Gender: Female	-0.690	-1.072*	-1.525	-2.673***
	(0.506)	(0.597)	(0.934)	(0.895)
Nationality: Other	-0.943	-0.991	-0.970	-0.604
	(0.706)	(0.691)	(0.615)	(0.822)
Pre-Match Happiness	-0.330	-0.408	-0.278	-0.438
	(0.245)	(0.265)	(0.493)	(0.505)
Accumulated Payment	-0.054*	-0.065	-0.051	0.010
	(0.029)	(0.043)	(0.044)	(0.059)
How strongly agree/disagree				
Highest Chance of Winning			0.478**	0.235
			(0.207)	(0.145)
Paid Most Money			0.210	0.819***
			(0.212)	(0.194)
Wanted to Hedge			-1.534***	-0.844**
			(0.404)	(0.417)
Have Something to be Happy about			0.328	0.508**
			(0.236)	(0.296)
Wanted to be Loyal			0.435	-0.143
			(0.283)	(0.296)
Won't Enjoy Money if Other Team Wins			0.447**	0.004
			(0.206)	(0.202)
Constant	-16.87	-7.413	-36.31*	-11.25
	(11.96)	(15.16)	(18.53)	(17.50)
Observations	180		143	
Pseudo R^2	0.138		0.483	
Log Pseudo Likelihood	-152.60		-73.18	

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Robust standard errors clustered at the individual level are in parentheses. Outcome variables are 0 = keep £3 (reference group; N = 46); 1 = bet for England to win (N = 101); 2 = bet for England's opponent to win (N = 33). Given that there are only 12 observations who bet for a draw, we have decided to leave this category out from our multinomial logit equation.

Table 3.3: First-differenced general post- and pre-match happiness following a win and a loss: random-effects and fixed-effects regressions

	Random effects regressions				Fixed effects regressions			
	England Won Match	**England Won Match**	**England Lost Match**	**England Lost Match**	**England Won Match**	**England Won Match**	**England Lost Match**	**England Lost Match**
Bet England	-0.001	1.437	-0.445*	-0.523**	-0.036	-1.484	-0.659*	-0.915**
	(0.252)	(2.982)	(0.228)	(0.263)	(0.330)	(3.404)	(0.375)	(0.387)
Bet Other	-0.039	-3.533	0.111	0.337	-0.002	3.114	-0.591	0.127
	(0.294)	(6.738)	(0.295)	(0.345)	(0.381)	(7.856)	(0.484)	(0.488)
Bet Draw	0.018	-3.546	-0.426	-0.609**	0.203	3.316	0.035	-0.533
	(0.195)	(6.833)	(0.267)	(0.308)	(0.326)	(8.121)	(0.680)	(0.719)
Forced bet England	0.204	1.680	-0.335	-0.470	0.251	-1.140	-0.272	-0.696*
	(0.271)	(2.977)	(0.235)	(0.286)	(0.377)	(3.384)	(0.341)	(0.403)
Forced bet Other	0.091	-3.431	-0.474**	-0.332	0.055	3.164	-0.494	-0.110
	(0.242)	(6.773)	(0.220)	(0.249)	(0.327)	(7.951)	(0.349)	(0.370)
Stage: Knockout	-0.034	2.755*	-0.449***	-2.380**	-0.017	0.601	-0.080	-2.524**
	(0.129)	(1.667)	(0.156)	(0.960)	(0.145)	(1.704)	(0.163)	(1.043)
Subjective risk profile	0.115**	0.113**	0.005	-0.001	0.148	0.156	0.013	-0.054
	(0.053)	(0.053)	(0.042)	(0.043)	(0.114)	(0.107)	(0.150)	(0.149)
Watch match: no	-0.098	-0.124	0.173	0.112	0.176	0.203	-0.030	-0.255
	(0.168)	(0.168)	(0.138)	(0.137)	(0.285)	(0.288)	(0.295)	(0.325)
England as 1st favourite	-0.142	-0.115	-0.329**	-0.280				
	(0.166)	(0.162)	(0.165)	(0.179)				
Gender: female	0.025	0.026	-0.135	-0.109				
	(0.155)	(0.150)	(0.125)	(0.132)				
Nationality: Other	0.084	0.087	-0.148	-0.098				
	(0.182)	(0.171)	(0.220)	(0.243)				
Accumulated payment		0.009		0.005		-0.0115		-0.012
		(0.015)		(0.010)		(0.031)		(0.036)
Individual payment		-1.180		-0.056		1.053		-0.161**
		(2.267)		(0.051)		(2.625)		(0.064)
Maximum potential payment		0.176*		-1.275*		0.023		-1.777**
		(0.107)		(0.676)		(0.119)		(0.736)
Constant	-0.281	2.492	0.681***	16.45**	-0.553	-8.599	0.221	23.40**
	(0.277)	(14.30)	(0.235)	(8.367)	(0.430)	(17.86)	(0.630)	(9.419)
Observations	179	179	170	170	179	179	170	170
Number of groups	85	85	77	77	85	85	77	77
R^2	0.052	0.075	0.133	0.162	0.022	0.024	0.013	0.018

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Robust standard errors clustered at the individual level are in parentheses. Dependent variable = post-match general happiness minus pre-match general happiness. RE = random-effects model. FE = fixed-effects model. Time- invariant characteristics, e.g., England as 1st favourite team, gender, and nationality are naturally dropped in the FE regressions. Number of observations (England win): Keep £3 (N = 26), Bet for England (N = 58), Bet for Other (N = 15), Bet for Draw (N = 4), Forced bet for England (N = 37), and Forced bet for other (N = 39). Number of observations (England loss): Keep £3 (N = 20), Bet for England (N = 43), Bet for Other (N = 18), Bet for Draw (N = 8), Forced bet for England (N = 37), and Forced bet for other (N = 44).

Table 3.4: First-differenced happiness regressions (specific post-pre)

	Random effects regressions				Fixed effects regressions			
	England Won Match	England Lost Match	England Lost Match	England Lost Match	England Won Match	England Lost Match	England Lost Match	England Lost Match
Bet England	0.098	-0.704	-0.396	-0.534	-0.107	-0.022	-1.179***	-1.148**
	(0.255)	(3.273)	(0.379)	(0.438)	(0.306)	(3.770)	(0.385)	(0.477)
Bet Other	0.010	2.069	0.680	0.917	0.117	0.241	-0.453	0.156
	(0.370)	(7.469)	(0.521)	(0.665)	(0.367)	(8.673)	(0.564)	(0.699)
Bet Draw	-0.996*	1.059	-0.090	-0.370	-0.911	-0.881	-0.306	-0.556
	(0.572)	(7.605)	(0.806)	(0.887)	(0.726)	(8.942)	(1.043)	(1.132)
Forced bet England	0.349	-0.508	-0.408	-0.667	0.265	0.288	-0.932**	-1.129**
	(0.284)	(3.173)	(0.365)	(0.435)	(0.323)	(3.690)	(0.375)	(0.478)
Forced be Other	-0.410	1.673	-0.231	-0.108	-0.567*	-0.395	-0.702*	-0.332
	(0.296)	(7.539)	(0.382)	(0.409)	(0.335)	(8.708)	(0.378)	(0.399)
Stage: Knockout	0.183	-2.760	-0.994***	-4.382***	0.065	-2.068	-1.031***	-4.465***
	(0.152)	(2.014)	(0.274)	(1.125)	(0.151)	(2.037)	(0.300)	(1.089)
Subjective risk profile	0.002	0.008	-0.058	-0.059	-0.015	-0.022	0.321	0.312
	(0.068)	(0.066)	(0.117)	(0.119)	(0.130)	(0.122)	(0.319)	(0.335)
Watch match: no	-0.821***	-0.797***	0.131	-0.049	-0.719**	-0.702**	-1.157***	-1.521***
	(0.256)	(0.250)	(0.367)	(0.372)	(0.338)	(0.325)	(0.428)	(0.431)
England as 1st favourite	0.537*	0.511*	-0.647**	-0.564				
	(0.275)	(0.274)	(0.329)	(0.347)				
Gender: female	0.590**	0.544**	-0.042	-0.070				
	(0.266)	(0.256)	(0.354)	(0.367)				
Nationality: Other	-0.182	-0.228	0.311	0.290				
	(0.324)	(0.321)	(0.378)	(0.373)				
Accumulated payment		-0.030		-0.013		-0.032		0.042
		(0.019)		(0.017)		(0.037)		(0.054)
Individual payment		0.718		-0.061		0.057		-0.113
		(2.496)		(0.099)		(2.879)		(0.095)
Maximum potential payment		-0.203		-2.591***		-0.157		-1.744
		(0.130)		(0.846)		(0.139)		(1.144)
Constant	0.074	1.493	-0.660	31.38***	1.045**	5.651	-1.636	19.91
	(0.347)	(15.73)	(0.562)	(10.36)	(0.464)	(19.58)	(1.097)	(14.74)
Observations	179	179	170	170	179	179	170	170
Number of groups	85	85	77	77	85	85	77	77
R^2	0.247	0.265	0.164	0.186	0.145	0.170	0.006	0.008

Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Robust standard errors clustered at the individual level are in parentheses. Dependent variable = post-match happiness specifically related to the outcome of the match minus pre-match general happiness. RE = random-effects model. FE = fixed-effects model. Time-invariant characteristics, e.g., England as 1st favourite team, gender, and nationality are naturally dropped in the FE regressions.

Table 3.5: First-differenced post- and pre-match general happiness following a win and a loss with prior treatment as a control variable: random-effects and fixed-effects regressions

	(1) RE	(2) RE	(3) FE	(4) FE
	England Won Match	**England Lost Match**	**England Won Match**	**England Lost Match**
Bet for England to win	2.654	-0.547**	-0.453	-0.953**
	(3.024)	(0.274)	(3.294)	(0.385)
Bet for the other team to win	-6.296	0.332	0.642	0.074
	(6.742)	(0.339)	(7.394)	(0.454)
Bet for Draw	-6.357	-0.692**	0.908	-0.664
	(6.888)	(0.313)	(7.759)	(0.709)
Forced Bet for England to win	2.855	-0.510*	-0.235	-0.680*
	(3.029)	(0.299)	(3.273)	(0.398)
Forced Bet for the other to win	-6.197	-0.344	0.699	-0.155
	(6.806)	(0.235)	(7.542)	(0.327)
Stage: Knockout	2.325	-2.319**	0.796	-2.466**
	(1.672)	(0.960)	(1.697)	(1.058)
Subjective risk profile	0.106*	-0.002	0.141	-0.038
	(0.054)	(0.043)	(0.103)	(0.153)
Watch match: No	-0.116	0.119	0.192	-0.252
	(0.170)	(0.139)	(0.315)	(0.335)
England as 1st favorite team	-0.090	-0.323*		
	(0.168)	(0.171)		
Gender: Female	0.038	-0.142		
	(0.149)	(0.140)		
Nationality: Other	0.128	-0.110		
	(0.181)	(0.239)		
Accumulated Payment	-0.005	0.002	-0.010	-0.015
	(0.020)	(0.010)	(0.032)	(0.038)
Individual Payment	-2.091	-0.059	0.249	-0.155**
	(2.281)	(0.053)	(2.502)	(0.064)
Maximum Potential Payment	0.167	-1.269*	0.037	-1.779**
	(0.105)	(0.657)	(0.129)	(0.742)
Bet for England treatment in t-1	-0.223	0.049	-0.387	0.177
	(0.342)	(0.217)	(0.361)	(0.339)
Bet for opposition treatment in t-1	0.104	-0.197	0.188	-0.028
	(0.221)	(0.203)	(0.343)	(0.299)
No prior treatment, i.e., 1st time participating	-0.529	-0.039	-0.015	0.020
	(0.449)	(0.243)	(0.389)	(0.333)
Constant	9.798	16.55**	-3.264	23.35**
	(14.65)	(8.113)	(16.51)	(9.473)
Observations	179	170	179	170
Number of groups	85	77	85	77
R^2	0.089	0.168	0.103	0.172

Note: Robust standard errors clustered at the individual level are in parentheses. Dependent variable = post-match general happiness *minus* pre-match general happiness. RE = random-effects model. FE = fixed-effects model. Time-invariant characteristics, e.g., England as 1st favorite team, gender, and nationality are naturally dropped in the FE regressions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Enjoyment while watching the game as dependent variable

	(1) RE	(2) RE	(3) FE	(4) FE
	England Won Match	England Lost Match	England Won Match	England Lost Match
Bet for England to win	-1.378	-0.517	-0.170	-0.383
	(3.865)	(0.424)	(5.051)	(0.520)
Bet for the other team to win	2.552	0.197	1.229	0.305
	(9.240)	(0.622)	(12.02)	(0.602)
Bet for Draw	2.590	-1.090*	-0.218	-0.592
	(9.335)	(0.634)	(12.17)	(0.759)
Forced Bet for England to win	-1.198	-0.559	0.108	-0.411
	(3.868)	(0.471)	(5.069)	(0.627)
Forced Bet for the other to win	2.756	-0.311	0.723	-0.404
	(9.179)	(0.425)	(11.88)	(0.476)
Stage: Knockout	-4.859**	2.256*	-6.868**	2.138*
	(2.296)	(1.212)	(2.775)	(1.210)
Subjective risk profile	0.048	0.106	0.044	0.082
	(0.074)	(0.106)	(0.134)	(0.259)
Watch match: No	0.067	-0.607*		
	(0.407)	(0.359)		
England as 1st favorite team	0.141	0.003		
	(0.306)	(0.329)		
Gender: Female	0.141	0.271		
	(0.438)	(0.354)		
Nationality: Other	-0.001	0.001	-0.083**	-0.035
	(0.017)	(0.016)	(0.037)	(0.034)
Accumulated Payment	1.029	-0.095	0.350	-0.136*
	(3.052)	(0.088)	(3.948)	(0.078)
Individual Payment	-0.317**	2.059**	-0.516***	1.379
	(0.141)	(0.877)	(0.182)	(1.145)
Maximum Potential Payment	7.489	-19.75*	18.88	-11.33
	(18.94)	(10.62)	(25.31)	(14.03)
Constant	-1.378	-0.517	-0.170	-0.383
	(3.865)	(0.424)	(5.051)	(0.520)
Observations	131	123	131	123
Number of groups	73	60	73	60
R^2	0.107	0.158	0.294	0.257

Note: Robust standard errors clustered at the individual level are in parentheses. Dependent variable = level of enjoyment while watching the game, with 1 = not enjoyed at all, …, 7 = completely enjoyed watching the game. RE = random-effects model. FE = fixed-effects model. Time-invariant characteristics, e.g., England as 1st favorite team, gender, and nationality are naturally dropped in the FE regressions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Logit Betting Decisions Including Differences in Predicted Happiness

	Model 1		Model 2	
	Bet for England to Win	Bet for Other to Win	Bet for England to Win	Bet for Other to Win
Difference in the predicted general happiness (win minus loss)	0.473*	-0.808**		
	(0.262)	(0.325)		
Predicted post-match general happiness if England win			1.643***	1.001**
			(0.508)	(0.442)
Predicted post-match general happiness if England lose			0.304	2.239***
			(0.510)	(0.480)
England won in the previous match	6.609	6.720	5.550	7.661
	(4.412)	(4.641)	(5.401)	(5.831)
Odds England Winning	4.497*	3.998	4.182	4.585
	(2.502)	(2.474)	(3.089)	(3.133)
Odds Other Winning	0.471	0.784	0.425	0.761
	(0.361)	(0.553)	(0.403)	(0.633)
Stage: Knockout	9.855*	6.915	9.425	8.426
	(5.188)	(4.566)	(6.300)	(6.006)
Subjective risk profile	1.020***	0.578	1.067***	0.503
	(0.288)	(0.354)	(0.341)	(0.437)
England as a 1st favourite team	-2.714*	-3.454**	-2.567	-2.767
	(1.605)	(1.706)	(1.834)	(1.887)
Gender: Female	-1.187	-3.493***	-1.307	-3.831***
	(1.040)	(1.224)	(0.929)	(1.238)
Nationality: Other	-0.874	-0.884	-1.041	-0.661
	(0.607)	(1.250)	(0.786)	(1.192)
Pre-Match Happiness	-0.494	-0.081	-0.996	-1.091**
	(0.560)	(0.553)	(0.613)	(0.490)
Accumulated Payment	-0.066	0.064	-0.065	0.033
	(0.048)	(0.069)	(0.072)	(0.072)
How strongly agree/disagree				
Highest Chance of Winning	0.498**	0.370**	0.437	0.550**
	(0.244)	(0.189)	(0.290)	(0.241)
Paid Most Money	0.239	0.880***	0.378	1.236***
	(0.216)	(0.199)	(0.250)	(0.258)
Wanted to Hedge	-1.557***	-0.695	-1.803***	-1.189***
	(0.397)	(0.498)	(0.385)	(0.411)
Have Something to be Happy about	0.332	0.896***	0.181	0.989***
	(0.278)	(0.340)	(0.301)	(0.317)
Wanted to be Loyal	0.309	0.045	0.164	-0.053
	(0.272)	(0.366)	(0.276)	(0.361)
Won't Enjoy Money if Other Team Wins	0.388*	0.116	0.649**	0.584**
	(0.208)	(0.231)	(0.280)	(0.287)
Constant	-38.17*	-41.66*	-40.30	-57.60*
	(21.57)	(24.56)	(26.29)	(30.97)
Observations	143		143	
Pseudo R^2	0.583		0.641	
Log Pseudo Likelihood	-59.03		-50.84	

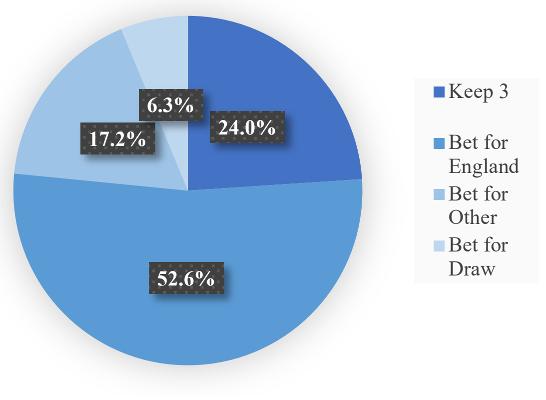
Note: *, **, *** denote statistical significance at the 0.1, 0.05 and 0.01 levels, respectively. Robust standard errors clustered at the individual level are in parentheses.

Table 3.8: Reasons behind betting choices in the "Free Choice" group

Reasons behind choice: 1="Strongly disagree", …, 7="Strongly agree"	**Keep £3**	**Bet for England**	**Bet for Other**	**Bet for Draw**
A: "Highest chance of winning"	4.47	4.51	4.82	5.75
	(1.76)	(1.76)	(1.69)	(0.88)
B: "Paid the most money"	3.32	3.35	5.00	4.62
	(1.34)	(1.63)	(1.69)	(1.40)
C: "Want to hedge my chance"	5.02	3.31	4.62	3.37
	(1.44)	(1.36)	(1.26)	(1.30)
D: "Have something to be happy about"	4.35	3.92	4.65	4.75
	(1.63)	(1.67)	(1.71)	(1.66)
E: "Want to be loyal"	3.52	4.81	3.31	3.00
	(1.70)	(1.51)	(1.49)	(1.41)
F: "Will not enjoy money if the other team win"	2.47	3.61	2.34	3.37
	(1.61)	(1.72)	(1.42)	(1.84)
Observations	34	84	29	8

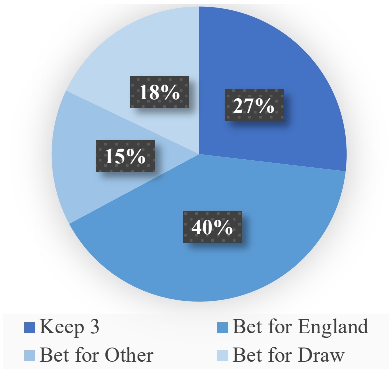
Note: Standard deviations are in parentheses.

Figure 3.1: Proportions of different betting decisions

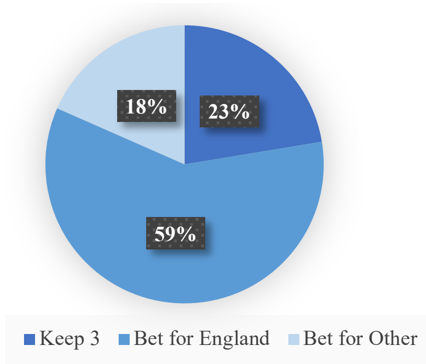


Note: Subjects are taken from the "Free choice" treatment (N = 192). Number of observations: Keep three pounds (N = 46), Bet for England (N = 101), Bet for Other (N = 33), and Bet for Draw (N = 12).

Figure 3.2: A-B Proportions of different betting decisions by World Cup Stage



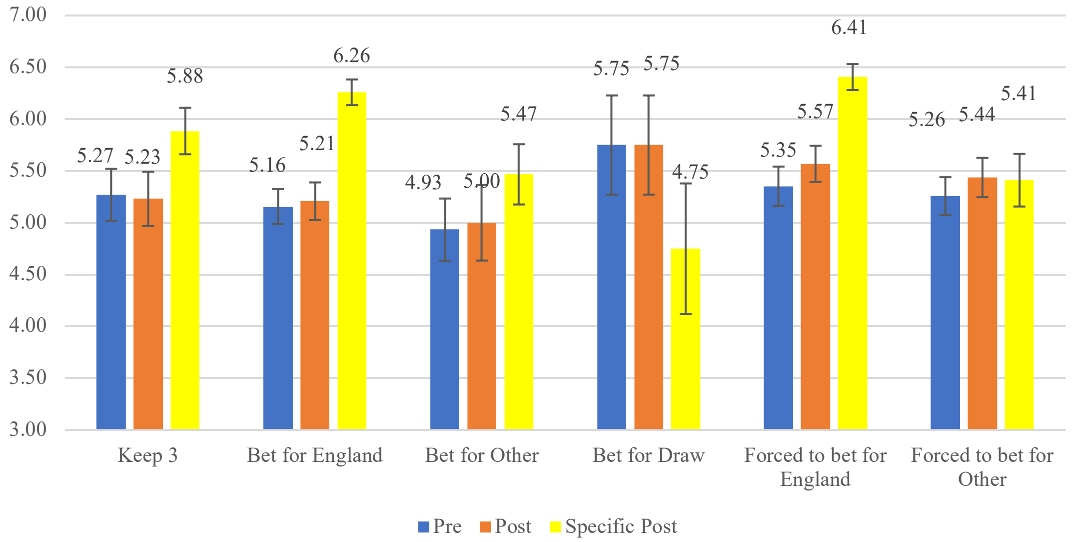
***A**: Group stage (N=67)*



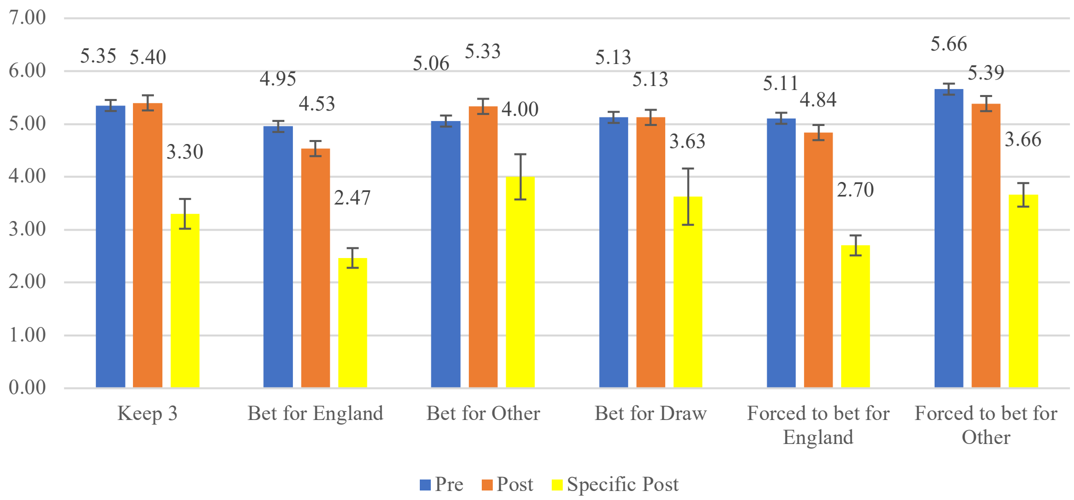
***B**: Knockout stage (N=125)*

<u>*Note*</u>**:** Subjects are taken from the "Free bet" treatment (*N* = 192). Number of observations (group stage): Keep £3 (*N* = 18), Bet for England (*N* = 27), Bet for Other (*N* = 10), and Bet for Draw (*N* = 12). Number of observations (knockout stage): Keep £3 (*N* = 28), Bet for England (*N* = 74), and Bet for Other (*N* = 23).

Figure 3.3: A-B Average happiness pre- and post-match (general and specific) by betting decisions

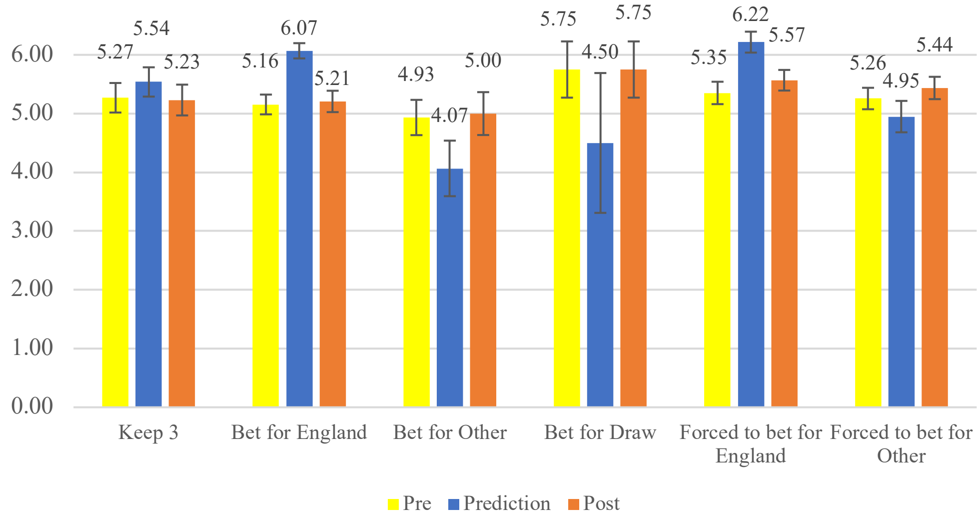


***A**: England won the match*

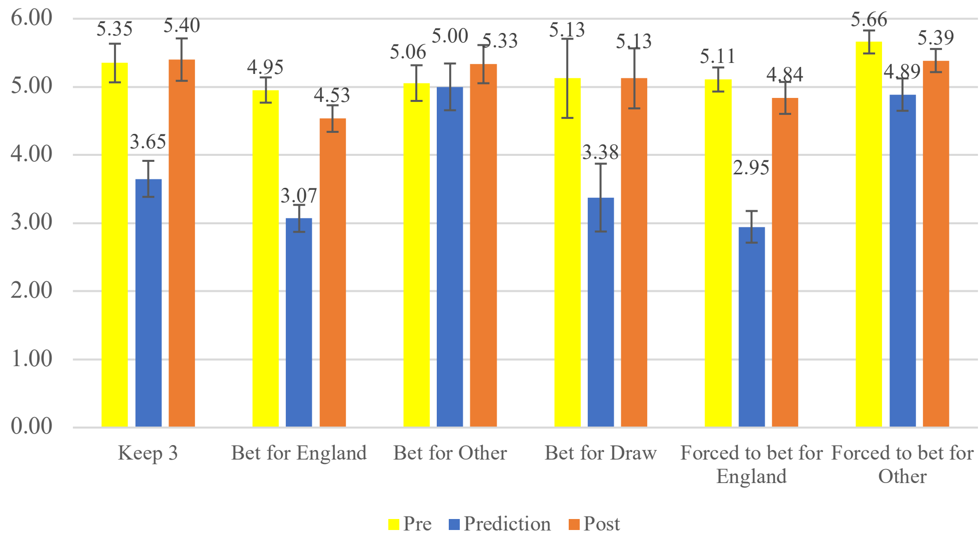


***B**: England lost the match*

Figure 3.4: A-B Average happiness post-match and happiness predictions if England win by betting decisions – England won/lost match



***A**: England won the match*



***B**: England lost the match*

Appendix

Screening instructions and questionnaire

Welcome to our online survey on people's attitudes towards the 2018 FIFA World Cup. This survey consists of 5 questions and should only take you less than 3 minutes to complete. Please make sure that you answer all of the questions in order to be eligible for the price draw, which is £5 Eating at Warwick credits. Your responses are anonymous. Thank you very much for your participation!

Q1: How much are you looking forward to the upcoming 2018 FIFA World Cup in Russia?
(Measured on a scale from 1-7, in which 1: not at all, and 7: the most)

	Not At All					The Most	
	1	2	3	4	5	6	7
()							

Q2: Which team(s) do you support in the 2018 FIFA World Cup?
(If there are more than one, please list the top three teams in the order of preference)

	Rarely Ever Follow			**How strongly do you support the above team?**		Die Hard Fan	
	1	2	3	4	5	6	7
First Choice ()							
Second Choice ()							
Third Choice ()							

Q3: Which team do you think will win the world cup?
(You can pick up to three teams and rank your choices in the order)

○ Most Likely (1) ____

○ The Second Most likely (2) ____

○ The Third Most Likely (3) ____

Q4: Gender

○ Male (1)

○ Female (2)

○ Others (3)

Q5: Nationality

____

You have now completed the survey. Thank you for your participation!

Main survey

Pre-match instructions and questionnaire (Treatment 1: Free Choice)

Thank you very much for taking part in this two-stage survey. In this first survey, there are 12 questions that require your responses in total. It should only take you less than 6 minutes to complete. Please make sure that you answer all of the questions in order to be eligible for the participation fees of £2 plus a potential winning that depends on the bet outcomes. You will then be sent the second survey after the England vs Belgium match, which you will then have 48 hours to complete before you can collect your winning plus another £2 participation fee (The Potential winning could be up to £9.60 + £2 + £2= £13.60). Your responses are anonymous. Thank you very much again for your participation!

Q1 In general, how happy would you say you are these days?

	Extremely unhappy				Extremely happy	
	1	2	3	4	5	6	7
Happiness ()							

Q2 We are now giving you **£3**. You now have a decision to make. Imagine now you have the following four options:(all the payoffs will be credited to you once you completed the follow-up survey that will be sent to you within 24 hours following the match between England and Belgium)

1st option: Keep the entire 3 pounds
2nd option: Use 3 pound to bet England to win, if England win, you will get £8.10
3rd option: Use 3 pound to bet Belgium to win, if Belgium win, you will get £7.87
4th option: Use 3 pound to bet on a draw, if the outcome is a draw, you will get £9.60

Please note that this bet is for the match between England **and** Belgium**, which is scheduled to be played on Thursday 28th June at 7 pm (British Time).**
(The betting odds is taken from Bet365.com on Sunday 24th June)

○ **1st option** (1)

○ **2nd option** (2)

○ **3rd option** (3)

○ **4th option** (4)

Q2a How happy would you be 24 hours following the match between England and Belgium if England win?

	Extremely unhappy				Extremely happy	
	1	2	3	4	5	6	7
Happiness ()							

Q2b How happy would you be 24 hours following the match between England and Belgium if Belgium win?

	Extremely unhappy				Extremely happy	
	1	2	3	4	5	6	7
Happiness ()							

Q2c How happy would you be 24 hours following the match between England and Belgium **if the result is a draw**?

	Extremely unhappy				Extremely happy	
	1	2	3	4	5	6	7
Happiness ()							

Display This Question:

If Q2 We are now giving you £3. You now have a decision to make. Imagine now you have the following... = 2nd option

Or Q2 We are now giving you £3. You now have a decision to make. Imagine now you have the following... = 3rd option

Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support?

- ○ Yes (1)
- ○ No (2)

Display This Question:

If Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support? = Yes

Q4 Imagine that you were told to place a bet for **the other team** to win instead of the team you support. How happy would you be in 24 hours following the match between your team and the other team if **the team you support** win?

	Extremely unhappy				Extremely happy	
	1	2	3	4	5	6	7
Happiness ()							

Display This Question:

If Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support? = Yes

Q5 Imagine that you were told to place a bet for **the other team** to win instead of the team you support. How happy would you be in 24 hours following the match between your team and the other team if **the other team** win?

Extremely unhappy | Extremely happy

	1	2	3	4	5	6	7
Happiness ()							

Display This Question:

If Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support? = Yes

Q6 Imagine that you were told to place a bet for **the other team** to win instead of the team you support. How happy would you be in 24 hours following the match between your team and the other team **if the result is a draw**?

	Extremely Unhappy				Extremely Happy		
	1	2	3	4	5	6	7
Happiness ()							

Display This Question:

If Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support? = No

Q4 Imagine that you were told to place a bet for **the team you support** to win instead of the other team. How happy would you be in 24 hours following the match between your team and the other team if **the team you support** win?

	Extremely unhappy				Extremely happy		
	1	2	3	4	5	6	7
Happiness ()							

Display This Question:

If Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support? = No

Q5 Imagine that you were told to place a bet for **the team you support** to win instead of the other team. How happy would you be in 24 hours following the match between your team and the other team if **the other team** win?

	Extremely unhappy				Extremely happy		
	1	2	3	4	5	6	7
Happiness ()							

Display This Question:

If Q3 If in Q.2 you chose to bet for a team to win, did you choose the team that you support? = No

Q6 Imagine that you were told to place a bet for **the team you support** to win instead of the other team. How happy would you be in 24 hours following the match between your team and the other team **if the result is a draw**?

	Extremely unhappy				Extremely happy		
	1	2	3	4	5	6	7

Happiness ()	

Q7: In general, please rate your willingness to take risks in betting.

	Completely Unwilling				Completely Willing	
	1	2	3	4	5	6	7
Willingness ()							

Q24 **Q8: Please indicate how strongly you agree or disagree with all the following statements which apply to you by selecting a number from 1 (Completely disagree) to 7 (Completely agree).**

	Strongly Disagree (1)	Disagree (8)	Somewhat Disagree (2)	Neither Agree or Disagree (3)	Somewhat Agree (9)	Agree (10)	Strongly Agree (11)
1. I wanted to choose the option with the highest chances of winning money. (1)	○	○	○	○	○	○	○
2. I wanted to choose the option that paid the most money. (2)	○	○	○	○	○	○	○
3. I wanted to insure myself against a bad match result. (3)	○	○	○	○	○	○	○
4. I wanted to be sure to have something to be happy about -- (either winning money or having my supported team win). (8)	○	○	○	○	○	○	○
5. I wanted to be loyal. (4)	○	○	○	○	○	○	○
6. I would not enjoy the money that I received if my opposing team won. (5)	○	○	○	○	○	○	○

Q9: Gender

○ **Male** (1)

○ **Female** (2)

○ **Others** (3)

Q10: Nationality

____

You have now completed the survey. Thank you for your participation!

Pre-match instruction and questionnaire (Treatment 2a: Forced bet for England)

All questions are the same as in B.1 except for Q2:

Q2 Thank you very much for taking part in this two-stage survey. In this first survey, there are 10 questions that require your responses in total. It should only take you less than 5 minutes to complete. Please make sure that you answer all of the questions in order to be eligible for the participation fees of £2 plus a potential winning that depends on the bet outcomes. You will then be sent the second survey after the England vs Belgium match, which you will then have 48 hours to complete before you can collect your winning plus another £2 participation fee (The Potential winning could be up to £6.60 + £2 + £2= £10.60). Your responses are anonymous. Thank you very much again for your participation!

Pre-match instruction and questionnaire (Treatment 2a: Forced bet for England's Opposition)

All questions are the same as in B.1 except for Q2:

Q2 Thank you very much for taking part in this two-stage survey. In this first survey, there are 10 questions that require your responses in total. It should only take you less than 5 minutes to complete. Please make sure that you answer all of the questions in order to be eligible for the participation fees of £2 plus a potential winning that depends on the bet outcomes. You will then be sent the second survey after the England vs Belgium match, which you will then have 24 hours to complete before you can collect your winning plus another £2 participation fee (The Potential winning could be up to £5 + £2 + £2= £9). Your responses are anonymous. Thank you very much again for your participation!

Post-match instruction and questionnaire

Thank you very much for taking part in this two-stage surveys. In this second survey, there are 4 questions that require your responses in total. It should only take you less than 2 minutes to complete. Please make sure that you answer all of the questions within 48 hours, in order to be eligible to claim your previous bet winning and another £2 participation fee. Your responses are anonymous. Thank you very much again for your participation!

Q1: In general, how happy would you say you are these days?

	Extremely unhappy				Extremely happy		
	1	2	3	4	5	6	7
()							

Q2: Did you watch the match between **England** and Belgium ?

- Yes (1)
- No (2)

Display This Question:

If Q2: Did you watch the match between England and Belgium ? = Yes

Q2a: If you watched the match, how would you rate your level of enjoyment while you were watching it?

	Did not enjoy at all				Completely enjoyable	
1	2	3	4	5	6	7

()

Q3: How happy do you feel about the outcome of the match between **England** and Belgium?
(The score was: **England** 0: 2 Belgium)

	Extremely unhappy				Extremely happy	
1	2	3	4	5	6	7

()

Q4: Knowing the outcome of the match, do you regret the betting decision?

- Yes (1)
- No (2)

You have now completed the survey. Thank you for your participation!

Bibliography

Adler, M. D., Dolan, P., & Kavetsos, G. (2017). Would you choose to be happy? Tradeoffs between happiness and the other dimensions of life in a large population survey. *Journal of Economic Behavior and Organization*, *139*, 60–73. https://doi.org/10.1016/j.jebo.2017.05.006

Akay, A., & Martinsson, P. (2011). Does relative income matter for the very poor? Evidence from rural Ethiopia. *Economics Letters*, *110*(3), 213–215. https://doi.org/10.1016/j.econlet.2010.11.046

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *Quarterly Journal of Economics*, *133*(4), 1885–1932. https://doi.org/10.1093/qje/qjy012

Asadullah, M. N., & Chaudhury, N. (2012). Subjective well-being and relative poverty in rural Bangladesh. *Journal of Economic Psychology*, *33*(5), 940–950. https://doi.org/10.1016/j.joep.2012.05.003

Ash, E., Chen, D. L., & Ornaghi, A. (2021). Gender Attitudes in the Judiciary: Evidence from U.S. Circuit Courts. *NBER Working Paper*.

Bagues, M. F., & Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *Review of Economic Studies*, *77*(4), 1301–1328. https://doi.org/10.1111/j.1467-937X.2009.00601.x

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017). Does the gender composition of scientific committees matter? *American Economic Review*, *107*(4), 1207–1238. https://doi.org/10.1257/aer.20151211

Balafoutas, L., & Sutter, M. (2019). How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior. *European Economic Review*, *118*, 1–13. https://doi.org/10.1016/j.euroecorev.2019.05.005

Bell, D. E. (1982). Regret in Decision Making under Uncertainty. *Operations Research*, *30*(5), 961–981.

Benartzi, S. (2001). Excessive extrapolation and the allocation of 401(k) accounts to company stock. *Journal of Finance*, *56*(5), 1747–1764. https://doi.org/10.1111/0022-1082.00388

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2012). What do you think would make you happier? What do you think you would choose? *American Economic Review*, *102*(5), 2083–2110. https://doi.org/10.1257/aer.102.5.2083

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23. https://doi.org/10.2307/1909829

Björklund, A., & Sundström, M. (2006). Parental separation and children's educational attainment: A siblings analysis on Swedish register data. *Economica*, *73*(292), 605–624. https://doi.org/10.1111/j.1468-0335.2006.00529.x

Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, *88*(7–8), 1359–1386. https://doi.org/10.1016/S0047-2727(02)00168-8

Bohnet, I., Greig, F., Hermann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310. https://doi.org/10.1257/aer.100.1.628

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, *109*(3), 739–773. https://doi.org/10.1257/aer.20170007

Boyce, C. J., Brown, G. D. A., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471–475. https://doi.org/10.1177/0956797610362671

Boyce, C. J., Wood, A. M., Banks, J., Clark, A. E., & Brown, G. D. A. (2013). Money, Well-Being, and Loss Aversion: Does an Income Loss Have a Greater Effect on Well-Being Than an Equivalent Income Gain? *Psychological Science*, *24*(12),

2557–2562. https://doi.org/10.1177/0956797613496436
Boyd, C. L. (2016). Representation on the Courts? The Effects of Trial Judges' Sex and Race. *Political Research Quarterly*, *69*(4), 788–799. https://doi.org/10.1177/1065912916663653
Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science*, *54*(2), 389–411.
Brown, G. D. A., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does wage rank affect employees' well-being? *Industrial Relations*, *47*(3), 355–389. https://doi.org/10.1111/j.1468-232X.2008.00525.x
Brown, G. D. A., Gathergood, J., & Weber, J. (2017). Relative Rank and Life Satisfaction: Evidence from US Households. *SSRN Electronic Journal*, *September*, 1–39. https://doi.org/10.2139/ssrn.2912892
Byrne, K. A., & Worthy, D. A. (2016). Toward a mechanistic account of gender differences in reward-based decision-making. *Journal of Neuroscience, Psychology, and Economics*, *9*(3–4), 157–168. https://doi.org/10.1037/npe0000059
Cahyadi, N., Hanna, R., Olken, B. A., Prima, R. A., Satriawan, E., & Syamsulhakim, E. (2020). Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia. *American Economic Journal: Economic Policy*, *12*(4), 88–110. https://doi.org/10.1257/pol.20190245
Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001
Chiappori, P.-A., & Weiss, Y. (2007). Divorce, Remarriage, and Child Support. *Journal of Labor Economics*, *1*, 37–74.
Clark, A. E. (2017). Happiness, income and poverty. *International Review of Economics*, *64*(2), 145–158. https://doi.org/10.1007/s12232-017-0274-7
Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, *46*(1), 95–144. https://doi.org/10.1257/jel.46.1.95
Clark, A. E., Kristensen, N., & Westergård-Nielsen, N. (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association*, *7*(2–3), 519–527. https://doi.org/10.1162/JEEA.2009.7.2-3.519
Clark, A. E., & Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, *61*(3), 359–381. https://doi.org/10.1016/0047-2727(95)01564-7
Clark, A. E., & Senik, C. (2010). Who compares to Whom? The anatomy of income comparisons in Europe. *Economic Journal*, *120*(544), 573–594. https://doi.org/10.1111/j.1468-0297.2010.02359.x
Clark, A. E., Senik, C., & Yamada, K. (2017). When experienced and decision utility concur: The case of income comparisons. *Journal of Behavioral and Experimental Economics*, *70*, 1–9. https://doi.org/10.1016/j.socec.2017.07.002
Collins, T., & Moyer, L. (2008). Gender, race, and intersectionality on the federal appellate bench. *Political Research Quarterly*, *61*(2), 219–227. https://doi.org/10.1177/1065912907306467
Coontz, P. (2000). Gender and judicial decisions: Do female judges decide cases differently than male judges? *Gender Issues*, *18*(4), 59–73. https://doi.org/10.1007/s12147-001-0024-7
Cooper, I., & Kaplanis, E. (1994). Home Bias in Equity Portfolios, Inflation Hedging, and International Capital Market Equilibrium. *The Review of Financial Studies*, *7*(1), 45–60. https://doi.org/10.15713/ins.mmj.3
Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance*, *54*(6), 2045–2073. https://doi.org/10.1111/0022-1082.00181
Crawford Solberg, E., Diener, E., Wirtz, D., Lucas, R. E., & Oishi, S. (2002). Wanting, having, and satisfaction: Examining the role of desire discrepancies in satisfaction with income. *Journal of Personality and Social Psychology*, *83*(3), 725–734.

https://doi.org/10.1037/0022-3514.83.3.725
Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474. https://doi.org/10.1257/jel.47.2.448
de Chaisemartin, C., & d'Haultfoeuille, X. (2021). Difference-in-Differences Estimators of Intertemporal Treatment Effects. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3731856
de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169
de Neve, J. E., Ward, G., de Keulenaer, F., van Landeghem, B., Kavetsos, G., & Norton, M. I. (2018). The asymmetric experience of positive and negative economic growth: Global evidence using subjective well-being data. *Review of Economics and Statistics*, *100*(2), 362–375. https://doi.org/10.1162/rest_a_00697
Derks, B., Ellemers, N., van Laar, C., & de Groot, K. (2011). Do sexist organizational cultures create the Queen Bee? *British Journal of Social Psychology*, *50*(3), 519–535. https://doi.org/10.1348/014466610X525280
Diener, E., & Biswas-Diener, R. (2002). Will money increase subjective well-being? A literature review and guide to needed research. *Social Indicators Research*, *57*(2), 119–169. https://doi.org/10.1023/A:1014411319119
Diener, E., & Fujita, F. (1997). Social Comparisons and Subjective Well-Being. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–357). Lawrence Erlbaum Associates.
Diener, E., Sandvik, E., Diener, M., & Seidlitz, L. (1993). The Relationship between Income and Subjective Well-Being : Relative or Absolute? *Social Indicators Research*, *28*(3), 195–223.
Dolan, P., Foy, C., Kavetsos, G., & Kudrna, L. (2021). Faster, higher, stronger… and happier? Relative achievement and marginal rank effects. *Journal of Behavioral and Experimental Economics*, *95*, 101761. https://doi.org/10.1016/j.socec.2021.101761
Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz* (pp. 89–125). https://doi.org/10.1007/BF00286477
Easterlin, R. A. (2001). Income and Happiness: Towards a Unified Theory. *The Economic Journal*, *111*(473), 465–484. https://doi.org/10.1111/1468-0297.00646
Easterlin, R., & Sawangfa, O. (2007). Happiness and Domain Satisfaction: Theory and Evidence. *IZA Discussion Paper*, *2584*.
Elliot, A. J., & Devine, P. G. (1994). On the Motivational Nature. In *Journal of Personality and Social Psychology* (Vol. 67, Issue 3, pp. 382–394).
Falk, A., & Knell, M. (2004). Choosing the Joneses: Enodgenous Goals and Reference Standards. *Scandanavian Journal of Economics*, *106*(3), 417–435.
Farhang, S., & Wawro, G. (2004). Institutional dynamics on the U.S. court of appeals: Minority representation under panel decision making. *Journal of Law, Economics, and Organization*, *20*(2), 299–330. https://doi.org/10.1093/jleo/ewh035
Ferrer-i-Carbonell, A. (2003). Income satisfaction inequality and its causes. *Journal of Economic Inequality*, *1*, 107–127. https://doi.org/10.1023/A:1026109706836
Ferrer-i-Carbonell, Ada. (2005). Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, *89*(5–6), 997–1019. https://doi.org/10.1016/j.jpubeco.2004.06.003
Ferrer-i-Carbonell, Ada, & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *Economic Journal*, *114*(497), 641–659. https://doi.org/10.1111/j.1468-0297.2004.00235.x
Fischer, G. W., Kamlet, M. S., Fienberg, S. E., & Schkade, D. (1986). Risk Preferences for Gains and Losses in Multiple Objective Decision Making. *Management Science*, *32*(9), 1065–1086.
FitzRoy, F. R., & Nolan, M. A. (2021). Income Status and Life Satisfaction. *Journal of*

Happiness Studies, *0123456789*. https://doi.org/10.1007/s10902-021-00397-y
Frank, R. H. (2006). Does Absolute Income Matter? In *Economics and Happiness: Framing the Analysis* (Issue February 2016). Oxford Scholarship Online. https://doi.org/10.1093/0199286280.001.0001
Gilbert, D. T., & Wilson, T. D. (2000). Miswanting: Some Problems in the Forecasting of Future Affective States. In J. P. Forgas (Ed.), *Thinking and feeling: The role of affect in social cognition* (pp. 178–197). Cambridge University Press. https://doi.org/10.1017/cbo9780511618031.031
Gilbert, D. T., Wilson, T. D., Pinel, E. C., Blumberg, S. J., & Wheatley, T. P. (1998). Immune Neglect: A Source of Durability Bias in Affective Forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638. https://doi.org/10.1037/0022-3514.75.3.617
Gill, R. D., Kagan, M., & Marouf, F. E. (2015). Chivalry, Masculinity, and the Importance of Maleness to Judicial Decision Making. *UNLV William S. Boyd School of Law Legal Studies Research Paper*. https://doi.org/10.2139/ssrn.2616502
Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2021.03.014
Graham, C., & Felton, A. (2006). Inequality and happiness: Insights from Latin America. *Journal of Economic Inequality*, *4*(1), 107–122. https://doi.org/10.1007/s10888-005-9009-1
Gruhl, J., Spohn, C., & Welch, S. (1981). Women as Policymakers : The Case of Trial Judges. *American Journal of Political Science*, *25*(2), 308–322.
Guillen-Royo, M. (2011). Reference group consumption and the subjective wellbeing of the poor in Peru. *Journal of Economic Psychology*, *32*(2), 259–272. https://doi.org/10.1016/j.joep.2009.12.001
Haushofer, J., Mudida, R., & Shapiro, J. P. (2021). The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3759722
Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360
Hidrobo, M., Hoddinott, J., Peterman, A., Margolies, A., & Moreira, V. (2014). Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador. *Journal of Development Economics*, *107*, 144–156. https://doi.org/10.1016/j.jdeveco.2013.11.009
Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The social identity perspective: Intergroup relations, self-conception, and small groups. *Small Group Research*, *35*(3), 246–276. https://doi.org/10.1177/1046496404263424
Ibarrarán, P., Medellín, N., Regalia, F., & Stampini, M. (2017). *How Conditional Cash Transfers Work: Good Practices after 20 Years of Implementation*. https://publications.iadb.org/bitstream/handle/11319/8159/How-conditional-cash-transfers-work.PDF?sequence=9
Jeandidier, B., Bourreau-Dubois, C., Ray, J.-C., & Doriat-Duban, M. (2016). Does Gender Matter in the Civil Law Judiciary? Evidence from French Child Support Court Decisions. In *Document de Travail* (Issue 55).
Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489–16493. https://doi.org/10.1073/pnas.1011492107
Kahneman, Daniel, Alan, K., Schkade, D., Schwarz, N., & Stone, A. (2006). Would You Be Happier If You Were Richer? A Focusing Illusion. *Science*, *1908*(2006), 1908–1910.
Kahneman, Daniel, & Thaler, R. H. (2006). Anomalies: Utility Maximization and Experienced Utility. *Journal of Economic Perspectives*, *20*(1), 221–234. https://doi.org/10.1257/089533006776526076
Kahneman, Daniel, Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations

of experienced utility. *Quarterly Journal of Economics*, *112*(2), 375–405. https://doi.org/10.1162/003355397555235
Kang, J. K., & Stulz, R. M. (1997). Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. *Journal of Financial Economics*, *46*(1), 3–28. https://doi.org/10.1016/S0304-405X(97)00023-8
Kingdon, G. G., & Knight, J. (2007a). Community, comparisons and subjective well-being in a divided society. *Journal of Economic Behavior and Organization*, *64*(1), 69–90. https://doi.org/10.1016/j.jebo.2007.03.004
Kingdon, G. G., & Knight, J. (2007b). Community, comparisons and subjective well-being in a divided society. *Journal of Economic Behavior and Organization*, *64*(1), 69–90. https://doi.org/10.1016/j.jebo.2007.03.004
Kiser, R. L., Asher, M. A., & McShane, B. B. (2008). Let's Not Make a Deal: An Empirical Study of Decision Making in Unsuccessful Settlement Negotiations. *Journal of Empirical Legal Studies*, *5*(3), 551–591. https://doi.org/10.1111/j.1740-1461.2008.00133.x
Knepper, M. (2018). When the shadow is the substance: judge gender and the outcomes of workplace sex discrimination cases. *Journal of Labor Economics*, *36*(3), 623–664. https://doi.org/10.1086/696150
Kuegler, A. (2009). A Curse of Comparison? Evidence on Reference Groups for Relative Income Concerns. In *Policy Research Working Paper. The World Bank.* (Vol. 4820, Issue January).
Loewenstein, G., O'Donnoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics*, *November*.
Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. *Quarterly Journal of Economics*, *120*(3), 963–1002. https://doi.org/10.1162/003355305774268255
Macchia, L., Plagnol, A. C., & Powdthavee, N. (2019). Buying Happiness in an Unequal World: Rank of Income More Strongly Predicts Well-Being in More Unequal Countries. *Personality and Social Psychology Bulletin*, 1–32. https://doi.org/10.1177/0146167219877413
Martin, F. E. (1989). Differences in Men and Women Judges: Perspectives on Gender. *Journal of Political Science*, *17*(1).
Maslow, A. H. (1954). *Motivation and personality*. Harper.
Mavin, S. (2008). Queen bees, wannabees and afraid to bees: No more "best enemies" for women in management? *British Journal of Management*, *19*(SUPPL. 1). https://doi.org/10.1111/j.1467-8551.2008.00573.x
Mavin, S. A., Williams, J., & Grandy, G. (2014). *Negative intra-gender relations between women: friendship, competition and female misogyny*. http://hdl.handle.net/10454/10249
Mavin, S., & Grandy, G. (2012). Doing gender well and differently in management. *Gender in Management: An International Journal*, *27*(4), 218–231. https://doi.org/10.1108/17542411211244768
McGuire, J., Kaiser, C., & Bach-Mortensen, A. (2020). The impact of cash transfers on subjective well-being and mental health in low- and middle- income countries: A systematic review and meta-analysis. *SocArXiv*, *November 2020*. https://osf.io/preprints/socarxiv/ydr54/
Meulbroek, L. (2005). Company Stock in Pension Plans: How Costly Is It? *Journal of Law and Economics*, *48*(2), 443–474.
Michalos, A. C. (1985). Multiple Discrepancies Theory (MDT). *Social Indicators Research*, *16*(4), 347–413.
Miething, A. (2013). A matter of perception: Exploring the role of income satisfaction in the income-mortality relationship in German survey data 1995-2010. *Social Science and Medicine*, *99*, 72–79. https://doi.org/10.1016/j.socscimed.2013.10.017
Millán, T. M., Barham, T., Macours, K., Maluccio, J. A., & Stampini, M. (2019). Long-term impacts of conditional cash transfers: Review of the evidence. *World Bank*

Research Observer, *34*(1), 119–159. https://doi.org/10.1093/wbro/lky005
Morewedge, C. K., & Buechel, E. C. (2013). Motivated underpinnings of the impact bias in affective forecasts. *Emotion*, *13*(6), 1023–1029. https://doi.org/10.1037/a0033797
Morewedge, C. K., Tang, S., & Larrick, R. P. (2018). Betting Your Favorite to Win: Costly Reluctance to Hedge Desired Outcomes. *Management Science*, *64*(3), 997–1014. https://doi.org/10.1287/mnsc.2016.2656
Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, *46*(1), 69–85.
Na, S., Su, Y., & Kunkel, T. (2019). Do not bet on your favourite football team: the influence of fan identity-based biases and sport context knowledge on game prediction accuracy. *European Sport Management Quarterly*, *19*(3), 396–418. https://doi.org/10.1080/16184742.2018.1530689
Niederle, M. (2016). Gender. In J. Kagel & A. Roth (Eds.), *Handbook of Experimental Economics* (pp. 481–553). Princeton University Press. https://doi.org/10.1017/9781316576755.030
Odermatt, R., & Stutzer, A. (2019). (Mis-)predicted subjective well-being following life events. *Journal of the European Economic Association*, *17*(1), 245–283. https://doi.org/10.1093/jeea/jvy005
Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Eribaum.
Peresie, J. L. (2005). Female judges matter: Gender and collegial decision making in the federal appellate courts. *Yale Law Journal*, *114*(7), 1759–1790.
Perova, E., & Vakis, R. (2011). *Más Tiempo en el Programa, Mejores Resultados: Duración e Impactos del Programa JUNTOS en el Perú*.
Pettigrew, T. F. (1967). Social Evaluation Theory: Convergences and Application. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (pp. 1–335). University of Nebraska Press.
Powdthavee, N. (2009). How important is rank to individual perception of economic standing? A within-community analysis. *Journal of Economic Inequality*, *7*(3), 225–248. https://doi.org/10.1007/s10888-008-9076-1
Praag, B. van, & Ferrer-i-Carbonell, A. (2005). Happiness Quantified: A Satisfaction Calculus Approach. In *Happiness Quantified: A Satisfaction Calculus Approach* (Issue September 2021). https://doi.org/10.1093/0198286546.001.0001
Ravallion, M., & Lokshin, M. (2002). Self-rated economic welfare in Russia. *European Economic Review*, *46*(8), 1453–1473. https://doi.org/10.1016/S0014-2921(01)00151-9
Ravallion, M., & Lokshin, M. (2010). Who cares about relative deprivation? *Journal of Economic Behavior and Organization*, *73*(2), 171–185. https://doi.org/10.1016/j.jebo.2009.08.008
Ritter Burga, P. (2014). *¿Más te quiero, más te pego? El efecto del Programa Juntos en el empoderamiento de las mujeres dentro del hogar*.
Rojas, M. (2006). Life satisfaction and satisfaction in domains of life: Is it a simple relationship? *Journal of Happiness Studies*, *7*(4), 467–497. https://doi.org/10.1007/s10902-006-9009-2
Sadka, J., Seira, E., & Woodruff, C. (2020). Information and Bargaining Through Agents: Experimental Evidence from Mexico's Labor Courts. *NBER Working Paper Series*.
Saldarriaga, V., & Diaz, J. J. (2021). (Un) Conditional Love in the Time of Conditional Cash Transfers : The Effect of the Peruvian JUNTOS Program on Spousal Abuse. *Economic Development and Cultural Change*.
Samuelson, L. (2004). Information-Based Relative Consumption Effects. *Econometrica*, *72*(1), 93–118.
Sapolsky, R. M. (2004). Social Status and Health in Humans and Other Animals. *Annual Review of Anthropology*, *33*(1), 393–418. https://doi.org/10.1146/annurev.anthro.33.070203.144000

Scheurer, K. F. (2014). Gender and Voting Decisions in the US Court of Appeals: Testing Critical Mass Theory. *Journal of Women, Politics and Policy*, *35*(1), 31–54. https://doi.org/10.1080/1554477X.2014.863696

Schkade, D. A., & Kahneman, D. (1998). Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction. *Psychological Science*, *9*(5), 340–346. https://doi.org/10.1111/1467-9280.00066

Simons, D., Shoda, Y., & Lindsay, S. (2007). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science*.

Singh, V., Schiebener, J., Müller, S. M., Liebherr, M., Brand, M., & Buelow, M. T. (2020). Country and Sex Differences in Decision Making Under Uncertainty and Risk. *Frontiers in Psychology*, *11*(March), 1–7. https://doi.org/10.3389/fpsyg.2020.00486

Smith, R. H., Diener, E., & Wedell, D. H. (1989). Intrapersonal and social comparison determinants of happiness: A range-frequency analysis. *Journal of Personality and Social Psychology*, *56*(3), 317–325. https://doi.org/10.1037//0022-3514.56.3.317

Stanek, R. (2017). Home bias in sport betting: Evidence from Czech betting market. *Judgment and Decision Making*, *12*(2), 168–172.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Stutzer, A. (2004). The role of income aspirations in individual happiness. *Journal of Economic Behavior and Organization*, *54*(1), 89–109. https://doi.org/10.1016/j.jebo.2003.04.003

Sun, L., & Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.09.006

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

Tang, S., Morewedge, C. M., Larrick, R. P., & Klein, J. G. (2017). Disloyalty aversion: Greater reluctance to bet against close others than the self. *Organizational Behavior and Human Decision Processes*, *140*, 1–13. https://doi.org/10.1016/j.obhdp.2017.02.001

Valenzuela, I. (2015). Focalización de los programas sociales en el Perú: 2007 - 2011. *Moneda: Politica Social*, 22. https://www.bcrp.gob.pe/docs/Publicaciones/Revista-Moneda/moneda-155/moneda-155-03.pdf

van den Bos, R., Homberg, J., & de Visser, L. (2013). A critical review of sex differences in decision-making tasks: Focus on the Iowa Gambling Task. *Behavioural Brain Research*, *238*(1), 95–108. https://doi.org/10.1016/j.bbr.2012.10.002

Walasek, L., & Brown, G. D. A. (2015). Income Inequality and Status Seeking: Searching for Positional Goods in Unequal U.S. States. *Psychological Science*, *26*(4), 527–533. https://doi.org/10.1177/0956797614567511

Wilson, T D, Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axson, D. (2000). Focalism: A souce of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821–836.

Wilson, Timothy D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131–134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wilson, Timothy D, & Gilbert, D. T. (2005). Affective Forecasting. *Psychological Bulletin*, *14*(3), 1–4. https://doi.org/10.1016/S0065-2601(03)01006-2

Wooldridge, J. M. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *SSRN Electronic Journal*, *August*. https://doi.org/10.2139/ssrn.3906345

Zimmerman, A., Garman, E., Avendano-Pabon, M., Araya, R., Evans-Lacko, S., McDaid, D., Park, A. La, Hessel, P., Diaz, Y., Matijasevich, A., Ziebold, C., Bauer, A., Paula, C. S., & Lund, C. (2021). The impact of cash transfers on mental health in children and young people in low-income and middle-income countries: A systematic review and meta-analysis. *BMJ Global Health*, *6*(4), 1–16. https://doi.org/10.1136/bmjgh-2020-004661

Zizzo, D. J. (2002). Between utility and cognition: The neurobiology of relative position. *Journal of Economic Behavior and Organization*, *48*(1), 71–91. https://doi.org/10.1016/S0167-2681(01)00227-X

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z