Intrahousehold decision-making in coastal Zanzibar: Experiments in behavioural development economics

Shanali Pethiyagoda
M.Sc. B.Sc. (Hons)

Thesis submitted for the degree of Doctor of Philosophy

School of Global Development
University of East Anglia

November 2025

Contents

List of Figures

List of Tables

Acknowledgements

"If you suffer for something that gives delight, soon the suffering itself can give joy."

– Paul Bloom, The Sweet Spot: The Pleasures of Suffering and the Search for Meaning

Although 'doing a PhD' is often framed in terms of hardship, my PhD journey felt markedly different. I attribute this good fortune to the hidden but formidable forces who made developing this thesis both possible and (suspiciously) enjoyable. They each deserve due recognition and forever hold my sincere gratitude and admiration.

I thank my supervisors, Prof. Ben D'Exelle and Prof. Bereket Kebede. This project was an ambitious undertaking, and I am deeply appreciative that they supported it without reservation, affording me the freedom to pursue my passions, whilst challenging my ideas to ensure they reached an elevated standard of enquiry. Their generous intellectual guidance and advice shaped every stage of this work, and cultivated a level of critical thinking and curiosity in me that has been instrumental in my growth as a researcher. I am especially grateful to Ben for his incisive feedback on the numerous drafts of, and discussions on, my chapters, and for being an enduring source of inspiration and encouragement. I am likewise grateful to Bereket, whose expertise in theoretical household models provided essential conceptual grounding for this thesis.

I am deeply grateful to all those who enabled my fieldwork in Zanzibar, specifically Dr. Cecile Brugere, Dr. Narriman Jiddawi and the late Dr. Flower Msuya. Their thought-provoking research on seaweed-farming women in Zanzibar, and groundbreaking efforts to transform Zanzibar's seaweed sector through innovative approaches to farming, sparked the inspiration for this thesis. I am especially indebted to Dr. Narriman who opened the 'Zanzibari door' to my first field visit, guiding my initial steps into fieldwork with coastal communities. Everything that followed rests on this foundation.

I owe exceptional gratitude to my two research implementing partners in Zanzibar, for collaborations that were as meaningful as they were effective. I thank the Panje Project for a remarkable partnership, in particular, Mr Muhammad Said and Mr Bakhtim Marshed, whose infectious passion for drowning prevention provided the impetus for my research focus. I am especially grateful to Muhammad: his unrelenting determination never ceased to amaze me, and his commitment, perseverance, and leadership saw us through the implementation of three field experiments in three months—no small feat. I am equally grateful to the Milele Zanzibar Foundation (MZF) for another generous and enriching collaboration. My deepest appreciation goes to Ms. Khadija Shariff, whose unique blend of thoughtful leadership, warmth, and vision made the task of turning research into practice feel both actionable and purposeful. I am equally grateful to Mr. Mhando Waziri, Mr. Ali Bakar, and Mr. Issa Muhurami, and all other MZF staff in both Pemba and Unguja offices, whose steady support behind the scenes ensured that the many moving parts kept moving.

I also extend my deepest gratitude to my remarkable field team—a constellation of hard-working people whose names deserve special thanks: my enumerators (in alphabetical order) Amina, Asha, Barakat, Bin Boniy, Fatma, Hamid, Hassan, Ismail, Jamila, Lada, Mwanvura, Nasra, Safina, Salha, Salum, Shuweina, and Zuhura; and our dependable drivers Abubakar, Chi Chi, and Omar. Thank you for creating a working environment that was in equal measure (impressively) productive, (unexpectedly) memorable, and (incredibly) supportive. I am also deeply indebted to all the respondents who generously offered their time, trust, and stories—without them this research would not exist. *Asanteni sana na nawashukuru wote!*

I gratefully acknowledge the generous funding provided by WorldFish. I extend a special thanks to Dr. Rahma Adam for enabling the successful completion of this project. I also gratefully acknowledge the pivotal role played by the Norwich Institute for Sustainable Development (NISD), in particular Prof. Arjan Verschoor, for facilitating this fruitful collaboration.

I am sincerely appreciative of the support, encouragement, guidance and opportunities provided by my outstanding mentors—Dr. Lorenzo Lotti, Dr. Silvia Ferrini, and Prof. Andreas Kontoleon. Without them, traversing the minefield of academia would have been far more perilous; their wisdom, patience and occasional reality checks ensured that I always found the right path forward. I am particularly grateful to Prof. Andreas Kontoleon for hosting me as a visiting researcher at the University of Cambridge, a place I have long dreamed about being a part of, and an exciting opportunity that broadened my scholarly horizons.

On a personal note, I acknowledge the jazz piano greats, especially Bill Evans, Chick Corea, Herbie Hancock, and Hiromi Uehara, whose music accompanied countless hours of writing and provided a steady, syncopated rhythm through the more demanding stretches of this work.

Last but not least, I acknowledge the limitless encouragement and love from those I have the privilege of calling friends and family. Special shoutouts to the following. To Rahel, for being a master of all trades in Zanzibar (not just spices): local guide, fixer-upper, matchstick-painter, boda partner, and general emotional support human. To my fellow PhD-ers, especially Heather, Atiya, Amelie, Lucy, and Grania, for their camaraderie, sagacity, and humour; you made the PhD years immeasurably more navigable. To my forever friends (you know who you are): thank you for the regular check-ins, and for patiently awaiting my eventual responses. To my sister, Theruni, who lives a much more glamorous life than I do and reports it back with a *joie de vivre* that has brightened my long days of writing and thinking.

To Joseph, Sylvia, and Andrew—my home away from home. Sylvia and Andrew: I remain profoundly thankful for the sense of belonging you provided me throughout this journey. Your kindness and support will always be cherished. Joe: you were the respite and recovery that kept me going. Your own determination in the face of adversity set the bar for what a person can endure, and your stoic encouragement was a source of confidence and calm when I needed it most. They say family is forever—I'm inclined to agree.

And last, to my parents, Appachchi and Amma. You embody love—selfless, generous, and unconditional—more than anyone I have known. No theoretical household model could begin to capture the environment of care, support, and possibility you created for me. This thesis would not exist without you.

Introduction

"We are never going to have gender equality, … until we have couple's equity."

– Prof. Claudia Goldin, 2023 Nobel Laureate in Economic Sciences

Households have long been of interest to scholars preoccupied with explaining what drives unequal economic outcomes between genders. Ostensibly, there are three main reasons for this interest.

First, the household itself is the result of a (traditionally) gendered partnership—the spousal relationship. This relationship extends beyond a convenient economic arrangement, to a deeply paradoxical social institution characterised by altruism, reciprocity, and respect on the one hand, and control, disagreement, and coercion on the other. As a result, the choices made within the household are rarely simple reflections of individual preferences or rational attempts to achieve allocative efficiency. Instead, inequalities in agency, bargaining power, capabilities, responsibilities, and rights between spouses shape complex decision-making processes, and ultimately, who makes decisions, whose preferences prevail, whose needs are prioritised, and who benefits from economic development.

Second, the household has been, and remains, the cradle of culture and norms. Through the intergenerational transmission of societal values, attitudes and perceptions around gendered roles, mothers and fathers shape the aspirations, constraints, and capabilities of their sons and daughters. This socialisation process embeds preferences and beliefs—sometimes subtly, sometimes forcefully — which often perpetuate the very inequalities between men and women that shape the aforementioned complexities in household decision-making processes (and its resulting development outcomes).

Third, micro decisions made at the household level can aggregate to seismic societal shifts. As a most notable example, the work of Nobel Laureate Claudia Goldin documents how women's participation in the labour force has been shaped by the gradual movement of fault lines within households themselves—from decisions about education and fertility to the renegotiation of who cooks, cleans, or stays home with a sick child (Goldin, 2006; Goldin and Katz, 2002). These intimate choices, made countless times across millions of families, underpin the long-run patterns in female labour supply, women's occupational choices, and the gender wage gap that Goldin traces across two centuries (Goldin, 2014, 2021).

Despite this longstanding interest in household decisions, much of the economics literature has treated the household as a largely coherent decision-making unit, formalised through collective models that assume Pareto efficiency, perfect information, or stable preferences (Browning and Chiappori, 1998; Chiappori, 1992). However, this neoclassical approach to household economics no longer holds water, with mounting empirical evidence of misaligned beliefs, asymmetric information, and hidden actions between spouses leading to inefficient resource allocation (Munro, 2018).

In light of this, scholars are increasingly pursuing a more *behavioural* approach that adds the missing realism to hegemonic household models (Hopfensitz and Munro, 2020). Reflecting this behavioural emphasis, economic experiments have gained a significant foothold in the household economics literature (Munro, 2018, 2025). Experiments are a favoured methodology because they allow researchers to isolate model assumptions and calibrate theory. Crucially, they also provide

a 'sandbox' to test behavioural interventions, offering direct causal evidence for targeted questions that advance our knowledge beyond theoretical conjecture (Munro, 2018).

My thesis contributes to this growing behavioural and experimental tradition in household economics, and the wider discipline of behavioural development economics, using field experiments to lift the veil on intrahousehold decisions. In particular, my three self-contained chapters leverage experimental variation to advance scholarly understanding of three understudied dimensions of intrahousehold decision-making: children's agency, spousal involvement, and spousal effort.

In my first chapter, I use a Randomised Controlled Trial (RCT) to evidence inequalities in outcomes of children's human capital decisions between fathers and mothers as primary decision-makers. I pair this with a choice experiment with children, demonstrating that their agency in involving parents in decisions can fundamentally shape their own human capital development. In my second chapter, I shift the focus to the parents. By endogenising spousal involvement in decision-making, I uncover the intra-couple decision processes that shape observed disparities in children's human capital outcomes. In my third chapter, I examine how spouses influence each other's effort provision. I show that gendered behavioural motivators of effort such as confidence and competition are systematically different across spouses, providing evidence of a new channel through which intrahousehold inequalities can persist between spouses. In my conclusion, I argue that my findings within each dimension—who decides, who is involved, and who exerts effort—contribute to a more nuanced understanding of the decision processes and behavioural drivers that perpetuate intrahousehold inequalities, both intra- and inter-generationally.

This nuanced understanding is of particular relevance in Zanzibar—a semi-autonomous archipelago off the eastern coast of Tanzania, and my research site of choice. There, climate change is exposing rural Zanzibari communities to greater coastal risks, heightening the need for adaptive skills such as swimming and water safety.[1] Still, gender norms in a traditionally Islamic and patriarchal society play a role in whether women and children, especially girls, are permitted, encouraged, or able to acquire these adaptive skills. Where environmental demands intersect with intrahousehold gender inequalities, Zanzibar's context provides a natural setting to examine how child agency, spousal involvement, and spousal effort can influence climate-adaptive livelihood choices within households. Against this contextual backdrop, I now outline my three chapters in more detail.

In the first chapter of my thesis (**Chapter 1:** *Sink or swim: An experiment on child agency in parental decisions*), I posit that although parents ultimately decide on their child's behalf, children as active economic agents possess meaningful agency in household decisions relating to their own development, specifically in preferring *which* of two parents to involve in the decision. I further hypothesise that child preferences over parental involvement (i) are malleable to context, and thus

[1]This is especially pressing for Zanzibar's seaweed farmers, 90% of whom are women (Msuya et al., 2022). As Africa's largest producer of seaweed, the sector has been a key driver of Zanzibari women's decision-making autonomy, agency, economic empowerment, and intrahousehold bargaining power in Tanzania's conservative Islamic patriarchal society (Msuya and Hurtado, 2017). However, rising sea temperatures and rapidly declining farm productivity have spurred development policy concerns amongst the Revolutionary Government of Zanzibar. Although innovative deep-water farming (DWF) technology holds promise as an alternative production method that is resilient to temperature shocks (Brugere et al., 2020), adoption remains critically low amongst women as the technology requires them to have swimming skills, which the majority of women in Zanzibar do not possess given that it has been traditionally viewed as a male-dominant labour skill (de Jong Cleyndert et al., 2021).

depend on child perceptions of parental attitudes and behaviours, and (ii) can influence differences in decision outcomes between fathers and mothers as primary decision-makers.

To test these hypotheses, I leverage a natural setting in which children decide which parent to approach for consent regarding a human capital development opportunity—swimming lessons. The analysis uses data from a Randomised Controlled Trial (RCT) and a choice experiment (CE) conducted with 1,272 children in Zanzibar. The RCT experimentally varied whether the mother or the father was required to provide consent for the child to begin swimming lessons. By tracking consent-form returns between mother-form and father-form treatments, I test whether maternal or paternal involvement matters for decisions made on the child's behalf. The CE, administered prior to the RCT to a subsample of children, elicited children's preferences for parental involvement by asking them if they would rather receive a mother-form or father-form. Before eliciting preferences, the framing of swimming lessons was experimentally varied using a 2×2 factorial design resulting in four different framings: neutral, high-risk (where the dangers of swimming lessons taking place in open ocean are highlighted), high-conflict (which highlights that involving just one parent has the potential to cause parental disagreement), or high-risk and high-conflict combined. This design is intended to shift children's perceived costs and benefits of involving each parent, allowing identification of the key determinants of their parental-involvement choices. Combining the CE with the RCT allows me to examine whether child agency matters for differences in maternal and paternal decision outcomes.

My study has three main results. From the RCT, I find that girls are 38% less likely to secure consent when required from their father, with no significant difference for boys. From the CE, I find that girls are more likely to prefer involving mothers when parental conflict is made salient, whereas boys shift towards a preference for involving fathers when risk or conflict is emphasised. Combining both experiments, I find that RCT treatment effects are driven by girls who did *not* receive the form they expressed a preference for in the CE. Moreover, making parental conflict salient in the CE reduces girls' success in obtaining *paternal* consent but increases it when *maternal* consent is required. Taken together, my results show that children have systematic preferences over parental involvement, that this is responsive to anticipated parental disagreement and relative risk attitudes of parents, and that child agency in parental involvement shapes the outcome of parental decisions.

This study contributes to three main strands of the intrahousehold literature. First, I extend the conceptual framework of studies linking spouses' involvement in decision processes to key development outcomes, by incorporating children as actors who can influence which parent participates in decision-making (Ashraf, 2009; D'Exelle and Ringdal, 2022, 2025; D'Exelle et al., 2023; McCarthy, 2019). Second, I contribute to the empirical intrahousehold literature which documents gender differences in parental resource allocation, by showing experimentally that children's involvement preferences shape spouses' allocation outcomes (Armand et al., 2020; Dizon-Ross and Jayachandran, 2023; Duflo, 2003; Duflo and Udry, 2004; Lundberg, Pollak and Wales, 1997; Thomas, 1990, 1994). Third, I contribute to the human capital formation literature (Cunha and Heckman, 2007; Cunha, Heckman and Schennach, 2010; Del Boca, Flinn and Wiswall, 2014) by introducing the child's own preference for parental involvement as a novel input into skill development. I also offer causal evidence which responds to Heckman and Mosso (2014)'s call for a 'next wave' of family studies that explicitly account for children's own actions on their own human capital formation.

The second chapter of my thesis (**Chapter 2:** *Does autonomy matter? A deep dive into spousal involvement in decisions*) draws on a field experiment with 664 spouses in rural Zanzibar to address three central research questions: (1) how much autonomy do spouses retain and relinquish in household decisions? (2) what influences whether individuals retain or relinquish autonomy? and (3) does retaining or relinquishing autonomy matter for the final uptake of the services offered? To answer these, I depart from previous experimental studies that exogenously assign spouses to make choices either alone or jointly and compare decision outcomes. Instead, I *endogenise* the spousal involvement process.

To do so, I use the following experimental approach. In private, I offer each spouse in a married or co-habiting couple a choice between two goods, where one option is always swimming lessons and the other is a consumption good (Decision Block 1), then a choice to make the decision themselves or defer decision-making to their spouse (Decision Block 2), and finally a choice to consult the decision with their spouse where, if consultation is chosen, the chance to switch to the spouse's choice is offered (Decision Block 3). I then use a 2×2 factorial treatment design which varies (i) whether swimming lessons are for the wife or the child and hence whether the decision falls in the decision-making domain of the wife or a joint domain that explicitly requires the involvement of the husband, and (ii) whether the paired consumption good benefits an individual or the household.

Endogenising spousal involvement in this way allows me to provide specific insights into my three research questions. For example, tracing the choices in Decision Blocks 2 and 3 allows me to examine how much autonomy spouses choose to retain, relinquish, and reconsider in the decision-making process. Similarly, experimentally varying who is approached first, who benefits, and whose domain the decision falls into, allows me to identify the factors that influence the preference for autonomy and spousal involvement, with a particular emphasis placed on the role of spousal beliefs about choices and expected spousal preference differences. To examine whether autonomy matters, I follow the evolution of decision outcomes i.e., the uptake of swimming lessons, from initial reflection to final resolution. Rather than inferring allocation from a single endpoint, my design allows me to dynamically evaluate whether autonomy exercised at different points in the spousal involvement process ultimately matters for the final decision (and development) outcome.

The study has three main sets of findings. The first set of results relates to Decision Block 1, where individual preferences are elicited. Here, there are significant gender differences in choices across treatments, with more husbands than wives who prefer swimming lessons, whether for the wife or child. Treatment comparisons show that both husbands and wives shift preferences in the same direction, with both preferring swimming lessons more, and believing their spouse prefers swimming lessons more, when it is for the child than the wife. A closer look at the possible mechanisms driving these results suggests that spouses consider what the other spouse might prefer (e.g., those who believe their spouse chose swimming are more likely than those who do not, to choose swimming lessons in all treatments.). Correlations of individual choices with their spouse's beliefs suggest that respondents may also consider what their spouse *thinks* the respondent would choose.

The second set of results relates to Decision Blocks 2 and 3. I find that women are more likely to defer decision-making to their husband, and are more willing to consult their husband, when the swimming lessons are for the child than for the wife. On the contrary, husbands are less likely to

defer when swimming is for the child than the wife, although they are more likely to consult when swimming is for the wife and the paired good benefits the wife individually than the household. As potential mechanisms driving spousal involvement, I find that expected preference differences do not substantially account for shifts in spousal involvement behaviour. Instead, whether a decision is made for the child versus the wife appears to activate distinct domain-specific norms that govern wives' deferral and consultation. A striking finding, however, is the low proportion of spouses who expect to have different preferences, despite *actual* within-couple preference differences being relatively high. That is, spouses systematically underestimate how much their preferences diverge.

The third set of results examines whether retaining or relinquishing autonomy matters for the final uptake of swimming lessons. I find that when spouses retain decision-making control, gendered preference differences manifest strongly: husbands select swimming lessons at higher rates than wives in every treatment, with the highest uptake (57%) and the largest spousal gap (13 percentage points) both recorded where child swimming lessons are paired with a household good. Once the spousal involvement process is endogenised, however, these differences largely dissipate—deferral, and consultation with switching, both produce convergence in final uptake to the point of statistical insignificance across treatments. This convergence arises partly because wives more frequently defer and switch to their husbands' preferences for swimming. Autonomy therefore matters: gender gaps persist when it is retained, but the spousal involvement process reconciles these differences when it is relinquished, ultimately mitigating the influence of which spouse moved first.

Taken together, these findings contribute to a few different strands of literature. As mentioned, by endogenising spousal involvement and comparing decision outcomes once spouses decide whether to involve their partner or not, I offer a process-level perspective to spousal involvement. This adds further nuance to studies that look at spousal involvement as a static snapshot and compare outcomes between exogenously varied decisions made by an individual versus jointly by both spouses (Ashraf, 2009; D'Exelle and Ringdal, 2022; D'Exelle et al., 2023; McCarthy, 2019).[2] Moreover, by experimentally varying the decision-making domains and beneficiaries of the paired goods, I contribute to recent efforts to unpack the determinants of spousal involvement (Abbink, Islam and Nguyen, 2020; Afzal et al., 2022; D'Exelle and Ringdal, 2025; Hillesland et al., 2025). Whilst prior work documents the role of cognitive load, perceived ability differences, and anticipated conflict in shaping involvement decisions, this study is, to my knowledge, the first to exogenously vary preference heterogeneity, decision beneficiaries, and the decision-making domain simultaneously to establish causal effects. In doing so, I also advance the emerging literature documenting that individuals attach an intrinsic value to holding decision-making authority, independent of the final allocation itself (Ertac, Gumren and Gurdal, 2020; Falk and Kosfeld, 2006; Ferreira, Hanaki and Tarroux, 2020; Ferreira et al., 2025; Freundt and Herz, 2025). By endogenising the involvement decision across varied domains, I disentangle demand for autonomy (the preference for holding the right to decide) from spousal involvement (the behavioural choice of how to exercise that right)—showing that the willingness to involve a partner is not always a uniform reflection of autonomy demand but might also be fundamentally bounded by domain-based norms. Last, by tracking decision outcomes on the

[2]Whilst the approach of using decision blocks is established by Bakhtiar et al. (2024), my design varies from theirs by experimentally varying paired goods along dimensions of who benefits and whose decision domain it is.

uptake of swimming lessons where both spouses are first movers, I also add to studies that show that the identity of the decision-maker matters and that, in joint decisions, husbands' preferences tend to carry greater weight (Anuwong and Munro, 2025; Ashraf, Bandiera and Lee, 2014; Bateman and Munro, 2005; Carlsson et al., 2012, 2013; D'Exelle and Ringdal, 2025).

My third chapter (**Chapter 3**: *Spousal effort: the role of confidence and competition*) examines the behavioural motivators of spousal effort within the paradox of the spousal relationship, where cooperation and competition coexist. Despite extensive work on how spouses influence each other's decisions, far less is known about how they influence one another's *effort provision*—a research gap that becomes increasingly important to address as dual-career households grow and relative spousal effort plays a more direct role in shaping gender gaps in wages and labour supply.

I hypothesise that, given the close proximity which characterises the spousal relationship, spouses regularly exchange information that facilitates spousal comparison and evaluation, which influences effort through competition and confidence. To test this hypothesis, I generate experimental evidence from a field experiment with 1,328 spouses (664 couples) in Tanzania.

In the experiment, couples were required to perform an incentivised real-effort task which involved separating coloured matchsticks in the quickest time possible, with payments deducted for longer timings. The same task was done twice. In the first round, spouses did not know that the other spouse was performing the same task in the adjacent room, and in the second round, this was made known to spouses. Whilst the first round was the same for all participants, I varied the second round using a 2×2 factorial treatment design, with treatments randomly assigned at the couple level.

The four between-subject treatments varied the type of feedback provided to each respondent before round two as being the respondent's spouse's actual performance in round 1 ('spousal comparison' treatment), the respondent's spouse's belief about the respondent's performance in round 1 ('spousal evaluation' treatment), a combination of both, or neither. I hypothesised that the spousal comparison feedback would influence a respondent's *competitiveness* through direct spousal comparisons, and that the spousal evaluation feedback would influence a respondent's *confidence* through a positive or negative spousal evaluation (defined as the relative difference between the individual's self-confidence and their 'spouse-confidence' i.e., what their spouse thinks about their ability). The latter is motivated by a rich body of psychological theory: the 'looking-glass self' (Cooley, 1922) and 'reflected appraisal' (Felson, 1985) theories postulate that self-image is shaped by how others perceive us. Relatedly, empirical research shows that others' expectations can influence performance both positively—through the 'Pygmalion effect' (Eden, 1990; Rosenthal and Jacobson, 1968) and the 'underdog effect' (Xue et al., 2022)—and negatively, through the 'Golem effect' (Babad, Inbar and Rosenthal, 1982; Davidson and Eden, 2000).

My results find striking gender differences in the behavioural motivators of spousal effort. Across treatments, I find that wives have higher relative timing improvements. I also find that wives in the sample are less (more) over(under)confident about their own performance and their spouse's performance than husbands are. Between treatments, when spousal comparisons are prompted, husbands react more strongly than wives to information that they are 'behind' their spouse (i.e., their performance is worse than their spouse's), suggesting they are more behind-averse than wives. When spousal evaluations are provided alone, I find no significant effect on effort. However, when

spousal evaluations are combined with spousal comparisons, wives and husbands respond to different motivators. Specifically, wives are more responsive than husbands to positive spousal evaluations, reducing (increasing) effort if they are ahead (behind). On the contrary, husbands are more sensitive than wives to negative spousal evaluations, reducing effort whether they are ahead or behind—noteworthy as husbands were four times less likely than wives to receive a negative spousal evaluation.

This study advances several strands of the literature on the determinants of effort. Extensive research highlights both monetary and non-monetary motivators of effort (DellaVigna and Pope, 2018), including self-confidence (Bénabou and Tirole, 2002) and social comparisons (Charness and Kuhn, 2007; Charness et al., 2016; D'Exelle, Habraken and Verschoor, 2024; Frey and Meier, 2004; Gächter and Thöni, 2010; Huet-Vaughn, 2025), with inequality aversion representing one specific social preference arising from comparison (Bolton and Ockenfels, 2000; D'Exelle and Riedl, 2018; Fehr and Schmidt, 1999; Friedl, Pondorfer and Schmidt, 2020; Kamas and Preston, 2015). I extend this work by studying these behavioural drivers in a spousal setting where emotional and economic linkages might amplify their relevance.

Taking competition and confidence as specific effort motivators, I contribute to the substantive literature on their gender differences and implications for inequality in effort-dependent outcomes. By examining effort allocation within the uniquely interdependent context of spouses, I build on the seminal work of Gneezy and Rustichini (2004); Gneezy, Niederle and Rustichini (2003); Niederle (2017); Niederle and Vesterlund (2007, 2011) on gender differences in competitiveness, and recent empirical findings on effort in mixed-sex competitions (Backus et al., 2023; Booth and Yamamura, 2018; Booth, Hayashi and Yamamura, 2022; Buser, van den Assem and van Dolder, 2023; Datta Gupta, Poulsen and Villeval, 2013). Additionally, by introducing a novel method for influencing confidence—namely, spousal evaluations, a mechanism previously unexplored in both psychology and economics—I complement recent work on gendered confidence belief updating (Adamecz-Völgyi and Shure, 2022; Bastani, Giebe and Gürtler, 2023; Buser et al., 2020; Möbius et al., 2022).

Most directly related to this study, a small experimental literature has begun examining spousal effort in controlled settings. Notably, Masekesa and Munro (2020) use a real-effort task to show that intrahousehold wage inequality reduces spousal effort through fairness concerns, whilst Verschoor et al. (2019) examine how cooperation shapes spousal effort provision. My study complements this work: rather than varying the payment scheme or the cooperative structure, I vary the informational feedback spouses receive about each other's performance and beliefs. This allows me to isolate confidence and competitiveness—two motivators with well-documented gender differences—as drivers of spousal effort, a channel that, to my knowledge, has not been examined in prior experimental work.

Taken together, these three chapters offer a granular, behaviourally-grounded account of the micro processes through which intrahousehold inequalities are produced and reproduced i.e., through whose preferences shape children's development, how autonomy is retained or relinquished between spouses, and how gendered motivators of effort persist even within the household's most intimate economic relationship. Each chapter isolates a distinct mechanism, yet together they advance a broader argument: achieving couple's equity—which Goldin identifies as the precondition for gender equality—demands a deeper understanding of the hidden behavioural processes shaping household decision outcomes.

References

Abbink, Klaus, Asad Islam, and Chau Nguyen. 2020. "Whose voice matters? An experimental examination of gender bias in intra-household decision-making." *Journal of Economic Behavior & Organization*, 176: 337–352.

Adamecz-Völgyi, Anna, and Nikki Shure. 2022. "The gender gap in top jobs–the role of overconfidence." *Labour Economics*, 79: 102283.

Afzal, Uzma, Giovanna d'Adda, Marcel Fafchamps, and Farah Said. 2022. "Intrahousehold Consumption Allocation and Demand for Agency: A Triple Experimental Investigation." *American Economic Journal: Applied Economics*, 14(3): 400–444.

Anuwong, Anusara, and Alistair Munro. 2025. "Better Together? Evidence On Information Flows and the Impact of Debt Management Training from A Lab-In-The-Field Experiment with Married Couples." Available at SSRN: https://ssrn.com/abstract=5462114 or http://dx.doi.org/10.2139/ssrn.5462114.

Armand, Alex, Orazio Attanasio, Pedro Carneiro, and Valérie Lechene. 2020. "The effect of gender-targeted conditional cash transfers on household expenditures: Evidence from a randomized experiment." *The Economic Journal*, 130(631): 1875–1897.

Ashraf, Nava. 2009. "Spousal control and intra-household decision making: An experimental study in the Philippines." *American Economic Review*, 99(4): 1245–1277.

Ashraf, Nava, Oriana Bandiera, and Scott S Lee. 2014. "Awards unbundled: Evidence from a natural field experiment." *Journal of Economic Behavior & Organization*, 100: 44–63.

Babad, Elisha Y, Jacinto Inbar, and Robert Rosenthal. 1982. "Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers." *Journal of educational psychology*, 74(4): 459.

Backus, Peter, Maria Cubel, Matej Guid, Santiago Sánchez-Pagés, and Enrique López Mañas. 2023. "Gender, competition, and performance: Evidence from chess players." *Quantitative Economics*, 14(1): 349–380.

Bakhtiar, M Mehrab, Marcel Fafchamps, Markus Goldstein, Kenneth L Leonard, and Sreelakshmi Papineni. 2024. "To Defer or to Differ: Experimental Evidence on the Role of Cash Transfers in Nigerian Couples' Decision–Making." *The Economic Journal*.

Bastani, Spencer, Thomas Giebe, and Oliver Gürtler. 2023. "Overconfidence and gender equality in the labor market." CESifo Working Paper No. 10339.

Bateman, Ian, and Alistair Munro. 2005. "An experiment on risky choice amongst households." *The Economic Journal*, 115(502): C176–C189.

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 91(1): 166–193.

Booth, Alison, and Eiji Yamamura. 2018. "Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in Japan." *Review of Economics and Statistics*, 100(4): 581–593.

Booth, Alison, Ryohei Hayashi, and Eiji Yamamura. 2022. "Gender differences in tournament-performance over time in single-sex and mixed-sex environments." *Labour Economics*, 76: 102173.

Browning, Martin, and Pierre-André Chiappori. 1998. "Efficient intra-household allocations: A general characterization and empirical tests." *Econometrica*, 1241–1278.

Brugere, Cecile, Flower E. Msuya, Narriman Jiddawi, Betty Nyonje, and Ritha Maly. 2020. "Can innovation empower? Reflections on introducing tubular nets to women seaweed farmers in Zanzibar." *Gender, Technology and Development*, 24(1): 89–109.

Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. 2020. "Overconfidence and gender gaps in redistributive preferences: Cross-Country experimental evidence." *Journal of Economic Behavior & Organization*, 178: 267–286.

Buser, Thomas, Martijn J. van den Assem, and Dennie van Dolder. 2023. "Gender and willingness to compete for high stakes." *Journal of Economic Behavior & Organization*, 206: 350–370.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics*, 117(3): 871–915.

Carlsson, Fredrik, Haoran He, Peter Martinsson, Ping Qin, and Matthias Sutter. 2012. "Household decision making in rural China: Using experiments to estimate the influences of spouses." *Journal of Economic Behavior & Organization*, 84(2): 525–536.

Carlsson, Fredrik, Peter Martinsson, Ping Qin, and Matthias Sutter. 2013. "The influence of spouses on household decision making under risk: an experiment in rural China." *Experimental Economics*, 16(3): 383–401.

Charness, Gary, and Peter Kuhn. 2007. "Does Pay Inequality Affect Worker Effort? Experimental Evidence." *Journal of Labor Economics*, 25(4): 693–723.

Charness, Gary, Ramón Cobo-Reyes, Juan A. Lacomba, Francisco Lagos, and Jose Maria Perez. 2016. "Social comparisons in wage delegation: experimental evidence." *Experimental Economics*, 19(2): 433–459.

Chiappori, Pierre-André. 1992. "Collective labor supply and welfare." *Journal of Political Economy*, 100(3): 437–467.

Cooley, Charles Horton. 1922. *Human nature and the social order.* New York: Scribner.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97(2): 31–47.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78(3): 883–931.

Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval. 2013. "Gender matching and competitiveness: Experimental evidence." *Economic Inquiry*, 51(1): 816–835.

Davidson, Oranit B, and Dov Eden. 2000. "Remedial self-fulfilling prophecy: Two field experiments to prevent Golem effects among disadvantaged women." *Journal of Applied Psychology*, 85(3): 386.

de Jong Cleyndert, Georgia, Rebecca Newman, Cecile Brugere, Aida Cuni-Sanchez, and Robert Marchant. 2021. "Adaptation of Seaweed Farmers in Zanzibar to the Impacts of Climate Change." In *African Handbook of Climate Change Adaptation.* , ed. Nicholas Oguge, Desalegn Ayal, Lydia Adeleke and Izael da Silva, 3–28. Cham:Springer International Publishing.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household Choices and Child Development." *The Review of Economic Studies*, 81(1 (286)): 137–185.

DellaVigna, Stefano, and Devin Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts." *The Review of Economic Studies*, 85(2): 1029–1069.

Dizon-Ross, Rebecca, and Seema Jayachandran. 2023. "Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation." *American Economic Review: Insights*, 5(4): 445–459.

Duflo, Esther. 2003. "Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa." *The World Bank Economic Review*, 17(1): 1–25.

Duflo, Esther, and Christopher Udry. 2004. "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices." NBER Working Paper No. w10498, Available at SSRN: https://ssrn.com/abstract=552103.

D'Exelle, Ben, and Arno Riedl. 2018. "Gender inequality and resource sharing: Evidence from rural Nicaragua." *Journal of Development Perspectives*, 2(1-2): 62–87.

D'Exelle, Ben, and Charlotte Ringdal. 2022. "Women's use of family planning services: An experiment on the husband's involvement." *Journal of Development Economics*, 158: 102915.

D'Exelle, Ben, and Charlotte Ringdal. 2025. "Who decides in the household when risk preferences conflict?" *Review of Economics of the Household.*

D'Exelle, Ben, Aurélia Lépine, Richard Bakyono, and Ludovic D. G. Tapsoba. 2023. "Fertility and polygyny: Experimental evidence from Burkina Faso." *Journal of Development Economics*, 164: 103134.

D'Exelle, Ben, Rik Habraken, and Arjan Verschoor. 2024. "Effort and Social Comparison: Experimental Evidence from Uganda." *Economic Development and Cultural Change*, 72(4).

Eden, Dov. 1990. *Pygmalion in management: Productivity as a self-fulfilling prophecy.* Lexington Books/DC Heath and Com.

Ertac, Seda, Mert Gumren, and Mehmet Y Gurdal. 2020. "Demand for decision autonomy and the desire to avoid responsibility in risky environments: Experimental evidence." *Journal of Economic Psychology*, 77: 102200.

Falk, Armin, and Michael Kosfeld. 2006. "The hidden costs of control." *American Economic Review*, 96(5): 1611–1630.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." *The Quarterly Journal of Economics*, 114(3): 817–868.

Felson, Richard B. 1985. "Reflected appraisal and the development of self." *Social Psychology Quarterly*, 71–78.

Ferreira, João V, Nobuyuki Hanaki, and Benoît Tarroux. 2020. "On the roots of the intrinsic value of decision rights: Experimental evidence." *Games and Economic Behavior*, 119: 110–122.

Ferreira, João V, Nobuyuki Hanaki, Fabrice Le Lec, Erik Schokkaert, and Benoît Tarroux. 2025. "Freedom counts: Cross-country empirical evidence." *European Economic Review*, 176: 105022.

Freundt, Jana, and Holger Herz. 2025. "A Cross-Country Study of Preferences for Choice Autonomy." CESifo Working Paper No. 12238.

Frey, Bruno S., and Stephan Meier. 2004. "Social Comparisons and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment." *American Economic Review*, 94(5): 1717–1722.

Friedl, Andreas, Andreas Pondorfer, and Ulrich Schmidt. 2020. "Gender differences in social risk taking." *Journal of Economic Psychology*, 77: 102182.

Gneezy, Uri, and Aldo Rustichini. 2004. "Gender and competition at a young age." *American Economic Review*, 94(2): 377–381.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics*, 118(3): 1049–1074.

Goldin, Claudia. 2006. "The quiet revolution that transformed women's employment, education, and family." *American economic review*, 96(2): 1–21.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American economic review*, 104(4): 1091–1119.

Goldin, Claudia. 2021. "Career and family: Women's century-long journey toward equity."

Goldin, Claudia, and Lawrence F Katz. 2002. "The power of the pill: Oral contraceptives and women's career and marriage decisions." *Journal of political Economy*, 110(4): 730–770.

Gächter, Simon, and Christian Thöni. 2010. "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis." *Journal of Economic Behavior & Organization*, 76(3): 531–543.

Heckman, James J, and Stefano Mosso. 2014. "The economics of human development and social mobility." *Annual Review of Economics*, 6(1): 689–733.

Hillesland, Marya, Cheryl R Doss, Serena Masino, Martina Querejeta, Aletheia Amalia Donald, Greg Seymour, and Clare Clingain. 2025. "Deciding Not to Decide: When Is There Power in Not Deciding?" The World Bank. Policy Research Working Paper. Report Number WPS11065. Washington, D.C. : World Bank Group.

Hopfensitz, Astrid, and Alistair Munro. 2020. "Behavioral household economics." *Handbook of Labor, Human Resources and Population Economics*, 1–21.

Huet-Vaughn, Emiliano. 2025. "Do social comparisons motivate workers? A field experiment on relative earnings, labor supply and the inhibitory effect of pay inequality." *Experimental Economics*, 28(1): 155–180.

Kamas, Linda, and Anne Preston. 2015. "Can social preferences explain gender differences in economic behavior?" *Journal of Economic Behavior & Organization*, 116: 525–539.

Lundberg, Shelly J, Robert A Pollak, and Terence J Wales. 1997. "Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit." *Journal of Human Resources*, 32(3): 463–480.

Masekesa, Faith, and Alistair Munro. 2020. "Intra-household inequality, fairness and productivity. Evidence from a real effort experiment." *World Development*, 127: 104763.

McCarthy, Aine Seitz. 2019. "Intimate partner violence and family planning decisions: Experimental evidence from rural Tanzania." *World Development*, 114: 156–174.

Msuya, Flower E., and Anicia Q. Hurtado. 2017. "The role of women in seaweed aquaculture in the Western Indian Ocean and South-East Asia." *European Journal of Phycology*, 52(4): 482–494.

Msuya, Flower E., John Bolton, Fred Pascal, Koushul Narrain, Betty Nyonje, and Elizabeth J. Cottier-Cook. 2022. "Seaweed farming in Africa: current status and future potential." *Journal of Applied Phycology*, 34(2): 985–1005.

Munro, Alistair. 2018. "Intra-household experiments: A survey." *Journal of Economic Surveys*, 32(1): 134–175.

Munro, Alistair. 2025. "Intra-household experiments." In *Handbook of Experimental Development Economics*. 136–166. Edward Elgar Publishing.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2022. "Managing self-confidence: Theory and experimental evidence." *Management Science*, 68(11): 7793–7817.

Niederle, Muriel. 2017. "A gender agenda: a progress report on competitiveness." *American Economic Review*, 107(5): 115–119.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122(3): 1067–1101.

Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and competition." *Annual Review of Economics*, 3(1): 601–630.

Rosenthal, Robert, and Lenore Jacobson. 1968. "Pygmalion in the classroom." *The Urban Review*, 3(1): 16–20.

Thomas, Duncan. 1990. "Intra-Household Resource Allocation: An Inferential Approach." *The Journal of Human Resources*, 25(4): 635–664.

Thomas, Duncan. 1994. "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height." *The Journal of Human Resources*, 29(4): 950–988.

Verschoor, Arjan, Bereket Kebede, Alistair Munro, and Marcela Tarazona. 2019. "Spousal control and efficiency of intra-household decision-making: Experiments among married couples in India, Ethiopia and Nigeria." *The European Journal of Development Research*, 31(4): 1171–1196.

Xue, Jinli, Mengting Zhu, Yanan Guo, and Demin Kong. 2022. "The Double-Edged Sword of Underdog Expectations in Organizations in Shanghai, China: The Mediating Role of Feedback-Avoiding Behaviors and Proving Others Wrong and the Moderating Role of Regulatory Focus." *Psychology Research and Behavior Management*, 15: 2205–2218.

Chapter 1

Sink or Swim: An Experiment on Child Agency in Parental Decisions

Abstract

Standard economic models and policy commonly treat children as passive actors in parental decisions made on developing the child's human capital. Yet, children often have an influence in parental decision-making. As the first point of contact for schools or providers of extra-curricular activities, one sphere of influence children have is in choosing *which* parent to involve in the decision-making process. To examine this choice, its determinants, and its eventual effect on parental decisions, we use two experiments with 1,272 schoolchildren in Tanzania. For the first experiment, we use a Randomised Controlled Trial (RCT) conducted with the full sample. In the RCT, children were offered free swimming lessons that required one parent's consent. To test whether parental decision outcomes vary by which parent is involved in decision-making, we randomised whether children had to obtain maternal consent or paternal consent. For the second experiment, we use a choice experiment (CE) run on a subsample of 848 children. We offer the child a choice between a maternal and paternal consent form, allowing us to measure child preferences over parental involvement in decision-making. A 2 $\times$ 2 treatment design which varies the framing of swimming lessons further allows us to test whether a child's preference to involve the mother or father is determined by their perceptions of risk and inter-parental disagreement. From the RCT, we find that girls are 38% *less* likely to secure *paternal* consent than maternal consent, with no significant difference for boys. From the CE, we find that girls are more likely to prefer involving mothers when inter-parental disagreement is made salient, whereas boys shift toward a preference for involving fathers when risk or disagreement is emphasised. Combining both experiments, we find that RCT treatment effects are driven by girls who did *not* receive the form they expressed a preference for in the CE. Moreover, making inter-parental disagreement salient in the CE reduces girls' success in obtaining *paternal* consent but increases it when *maternal* consent is required. In sum, our results show that children have systematic preferences over which parent to involve to decide on their behalf, that this is responsive to anticipated parental disagreement and perceived risk attitudes of parents, and that child agency in parental involvement shapes the outcome of parental decisions.

1. Introduction

Parents, as guardians, must routinely make decisions on behalf of their children. This is particularly true for household decisions related to children's human capital development and skills formation (Cunha and Heckman, 2007). Yet, children themselves are rational economic agents with a high degree of autonomy and influence over their own development outcomes (Heckman and Mosso, 2014; Lundberg, Romich and Tsang, 2009). One definitive opportunity children have to exert this agency is through *parental involvement* in decisions made on their behalf, the focus of this paper.

More specifically, a child can often choose *which* parent to involve in the decision-making process. This opportunity for children to exploit liberties in parental involvement can manifest in a variety of natural settings. In many human capital investment contexts, it is the child—rather than the parent—who is the first point of contact regarding a development opportunity. The school, for example, might ask the child to secure permission from one parent to participate in a school activity (e.g., a specialised STEM training course introduced outside of the regular curriculum), to make a significant educational purchase (e.g., a new tablet, textbook, or specialised software), to join an extracurricular activity (e.g., sports or travel), or to support participation in a career-building programme. In such cases, children have agency in choosing which parent to involve to decide on their behalf, with several influences determining this choice.

To see this, consider a framework where parental preferences over the development opportunity diverge. In this case, the child has an incentive to approach the parent whose preference is more aligned with their own preference. This perceived alignment can be influenced by a myriad of factors such as parents' perceived risk attitudes, the strength of child-parent relations, or nature of inter-parental consultation and disagreement. In essence, the child must solve a constrained optimisation problem when choosing to involve one parent over the other, taking into account the (perceived) costs and benefits of parental involvement. Such a scenario creates a field rich for analysis. Still, no empirical research exists on what determines child preferences around parental involvement, nor how children's strategic behaviour might influence parental decision-making and eventually, the child's own development outcomes.

In this paper, we provide novel insights into children's agency around parental involvement. Using a sample of 1,272 children and their parents in Zanzibar, Tanzania, we exploit a natural setting where children have agency to involve one parent in a decision on a human capital development opportunity: the chance to learn to swim. Given that most livelihoods in the Zanzibar archipelago rely on good swimming skills, and the majority of children do not know how to swim (Royal National Lifeboat Institution, 2019), this opportunity—through free swimming lessons provided by a local non-profit known as 'The Panje Project' (Panje henceforth)—represents a critical investment decision in their long-term economic prospects. In this setting, children are afforded agency in parental involvement as they must secure parental consent from one parent to begin swimming lessons with Panje. Within this context, we provide evidence on (1) whether paternal or maternal involvement matters for the child's development outcome, (2) child preferences over which parent to involve and its key determinants, and (3) how children's choices over parental involvement influence eventual parental decision-making outcomes.

To do so, we constructed a unique dataset that combines results from a Randomised Controlled Trial (RCT) and a choice experiment (CE), with household data collected from the perspective of the child. The RCT experimentally varied whether the child was instructed to approach the mother (treatment 1) or the father (treatment 2) for parental consent to begin swimming lessons. By tracking how many consent forms were returned in each treatment, we were able to test whether maternal or paternal involvement in decisions made on the child's behalf matters for the eventual development outcome.

The CE, administered prior to the RCT on a subsample of 848 children, elicited child preferences to involve either the father or the mother, by asking them which parental consent form they would prefer, the mother-form or father-form. Crucially, before preferences were elicited, we used a 2×2 experimental design to vary how swimming lessons were framed, as neutral, high-risk (as swimming lessons take place in open ocean), high-conflict (i.e., with the possibility that seeking one parent's consent might cause parental disagreement), or a combined high-risk and high-conflict framing. In essence, we varied how children perceive the costs and benefits of parental involvement, which allowed us to identify key determinants of child preferences over which parent to involve.

Combining the RCT with the CE allows us to make a direct link between child preferences around parental involvement, and the influence of child agency on parental decision outcomes. Specifically, by comparing final parental decision outcomes between children who were given the form they expressed a preference for, and children given the form they did not prefer, we observe real-world implications of child agency in parental decision-making.

Our study has three main results. First, mothers play an important role in their daughters' human capital development: girls in the RCT who are instructed to seek paternal consent are 38% less likely to return the forms than girls instructed to seek maternal consent. Second, children display strategic preferences over parental involvement, and this varies between boys and girls. Specifically, compared to the neutral framing of swimming lessons in the CE, the conflict and combined framings create a stronger preference amongst girls for the mother-form. The opposite is true for boys, who have a stronger preference for father-forms under risk and conflict framings. Third, by linking the CE with the RCT, we find that making parental conflict salient in the CE reduces girls' success in obtaining *paternal* consent but increases it when *maternal* consent is required. Moreover, we find that RCT treatment differences are driven by girls who did *not* receive the form they expressed a preference for in the CE. The latter finding offers suggestive evidence that (when child-parent incentives are aligned) affording the child agency around parental involvement can have a positive effect on human capital development outcomes.

Broadly, we contribute to the intrahousehold decision-making literature that takes into account the preference differences between household members, by moving beyond the unitary model to collective and non-cooperative models (Becker, 1981; Becker and Tomes, 1976; Browning, Chiappori and Weiss, 2014; Chiappori, 1992; Lundberg and Pollak, 1993). While much of this work models parents as the only relevant decision makers, we offer empirical evidence of the merits in modelling the child as an active economic agent in household decisions. We also make the following specific contributions to different strands of the intrahousehold literature.

First, we contribute to studies examining differences in how mothers and fathers allocate resources. A large body of work has documented that women often display a stronger preference than men towards transferring resources to children than to their own private consumption (Armand et al., 2020; Dizon-Ross and Jayachandran, 2023; Duflo, 2003; Duflo and Udry, 2004; Haddad, Hoddinott and Alderman, 1997; Lundberg, Pollak and Wales, 1997; Thomas, 1990, 1994). This has led many programs targeting children's development to target mothers specifically, leading to notable improvements for children (Attanasio, Meghir and Santiago, 2012; Hall, 2008). For example, Armand et al. (2020) show that in Burkina Faso, cash transfers targeted to mothers rather than fathers led to larger improvements in children's nutrition and clothing expenditures. In this study, we provide further experimental evidence on gender differences in spousal allocation by focusing on investment decisions around children's human capital development—in this case, the decision to enrol the child in swimming lessons.

Second, we make a conceptual and empirical contribution to the intrahousehold literature that examines how development outcomes are affected by the level of spousal involvement in the household decision-making process. This literature finds that whether husbands or wives are involved in the decision process can significantly alter choices in domains such as health, education, and fertility (Ashraf, 2009; D'Exelle and Ringdal, 2022, 2025; D'Exelle et al., 2023; McCarthy, 2019). For example, D'Exelle and Ringdal (2022) demonstrate that involving the husband in the decision process can drastically shift women's use of family planning services. However, existing studies in this domain define their enquiry within a framework that views spouses as the only relevant decision-makers who can influence spousal involvement. We extend this conceptual framework by broadening the set of household actors who influence spousal involvement to include children. Then, by experimental design, we are able to manipulate spousal involvement through the agency of the child. Empirically, we also offer novel evidence to demonstrate that, in a setting where *children* have agency in deciding which spouse is involved in the decision-making process, their preferences can influence the eventual decision outcome of spouses.

Third, we contribute to the large body of work on human capital production and parental investments in children (Cunha and Heckman, 2007; Cunha, Heckman and Schennach, 2010; Del Boca, Flinn and Wiswall, 2014). This literature identifies a slew of parental behaviours that are critical inputs in their children's skills formation. For instance, Doepke and Zilibotti (2017) identify parenting styles as an important investment in children's development, and Tungodden and Willén (2023) show that parents' competitiveness choices for their children can be key inputs in the child's long-term education outcomes. Still, a literature review on this body of work by Heckman and Mosso (2014) notes the absence of studies which consider the role of a child's own actions on his or her human capital accumulation, calling for a 'next wave' of family studies that explicitly incorporate the active role of the autonomous child. This study responds to this call with two contributions. First, we specify a new input into the capital development decision: the child's own preference for parental involvement. Second, by experimentally varying children's risk perceptions and perceptions around inter-parental disagreement, we offer causal evidence that child perceptions of intrahousehold dynamics can become a critical factor in their own human capital formation.

The rest of this paper is organised as follows. In Section 2, we provide a brief overview of the cultural and local context for children learning to swim in Zanzibar. In Section 3, we present the experimental design of the RCT and CE, followed by a simple theoretical model, a series of hypotheses based on this model, an empirical strategy for testing these hypotheses, a brief description of sampling and implementation, and a discussion of ethical considerations. Sections 4 through 6 present our empirical results: Section 4 reports summary statistics of the sample, Section 5 presents average treatment effects for both the RCT and the CE, and Section 6 examines possible mechanisms driving these. A discussion of our results ensues in Section 7. Section 8 concludes.

2. Context

Zanzibar is a semi-autonomous archipelago off the East coast of Tanzania, comprising the islands of Unguja and Pemba (Figure 1.3, panel **(a)**). Yet, despite being surrounded by water, an estimated 60% of children lack basic aquatic skills (Royal National Lifeboat Institution, 2019). This strikingly low swimming proficiency among Zanzibari children can be explained by two primary barriers to swimming: limited opportunities for formal skill acquisition, and restrictive social norms.

The first barrier is structural. Historically, the absence of formal aquatic training meant swimming was learned informally through repeated exposure to the sea. In Zanzibar's patriarchal context, this pathway was largely reserved for boys, who are often afforded greater unsupervised mobility. In addition to structural barriers, girls also face certain normative barriers. In a predominantly Islamic society, strict expectations regarding modesty, gender segregation, and female mobility—compounded by a lack of modest swimwear and female instructors—rendered swimming socially 'inappropriate' for women (Brugere et al., 2020; de la Torre-Castro et al., 2017; Felter, 2015; Makame et al., 2021). This disparity mirrors similar contexts in Tanzania such as the Songo Songo islands (Besta, 2013), and reflects a broader global pattern in which swimming proficiency remains a starkly gendered dimension of human capital: in low-income countries, only 15% of women can swim without assistance, compared to 39% of men (Borgonovi, Seitz and Vogel, 2022).

Against this contextual backdrop, the Panje Project—established in 2013 with technical guidance and financial support from the Royal National Lifeboat Institution (RNLI) of the UK—has made substantial progress in addressing both constraints. By providing structured swimming instruction programmes and adapting delivery to local norms, most notably through female instructors and modest swimwear known as *burkinis*, Panje has contributed to a marked shift in community attitudes toward swimming (Felter, 2015; Stacke, 2017). As of 2025, Panje has trained over 15,000 children—half of them girls—in basic aquatic survival skills. Importantly, community acceptance has increased over time through Panje's sustained engagement with village leaders, parents, and schools, alongside the integration of water safety into the national curriculum. Indeed, scoping interviews conducted during preliminary visits confirmed broad parental support for swimming lessons when framed as essential survival skills rather than leisure. Consequently, remaining reluctance amongst parents appears to reflect the perceived opportunity cost of time, specifically the trade-off with formal

education, and safety concerns stemming from a fear of the ocean, rather than ingrained cultural opposition.[3]

3. Design

The study design entails two components. The main component is a Randomised Controlled Trial (RCT), which experimentally varied whether the mother or the father needs to be involved in a human capital development decision on whether the child should take up swimming lessons. This allows us to test whether the gender of the parent involved matters for the child's final development outcome. The second component of the study is a choice experiment (CE), which elicited children's preferences to involve the father or the mother, by asking them to express which parental consent form they would prefer, the mother-form or father-form. Before preferences were elicited, we experimentally varied how swimming lessons were framed, as neutral, high-risk (as swimming lessons take place in open ocean), high-conflict (i.e., with the possibility that seeking one parent's consent might cause parental disagreement), or a combined high-risk and high-conflict framing. This allows us to test whether and how child preferences around parental involvement might be influenced by how they perceive their parents in relation to the decision.

The study procedures involved three stages as outlined in Figure 1.1: an initial interview, the CE, and the RCT. The interview was conducted with the full sample of 1,272 children. After a brief welcome, each child was randomly assigned a unique ID number. Following this, a private interview was conducted by same-gender enumerators, where assent was obtained from each child, and eligibility checks were carried out to ensure that they were between the age of 8 and 18 years, lived with both parents, and did not know how to swim. Data was then collected on a range of variables from socio-demographics to children's perceptions around parenting styles.[4]

Following the interview, the sample was split into two subsamples based on their randomly assigned ID number: 424 children who only participated in the RCT ('RCT-only' subsample), and 848 children who did both the CE and the RCT ('CE&RCT' subsample). This approach allows us to distinguish between development outcomes from parental decision-making without a child's influence (using the 'RCT-only' subsample), and development outcomes that are influenced by children's agency (using the 'CE&RCT' subsample).

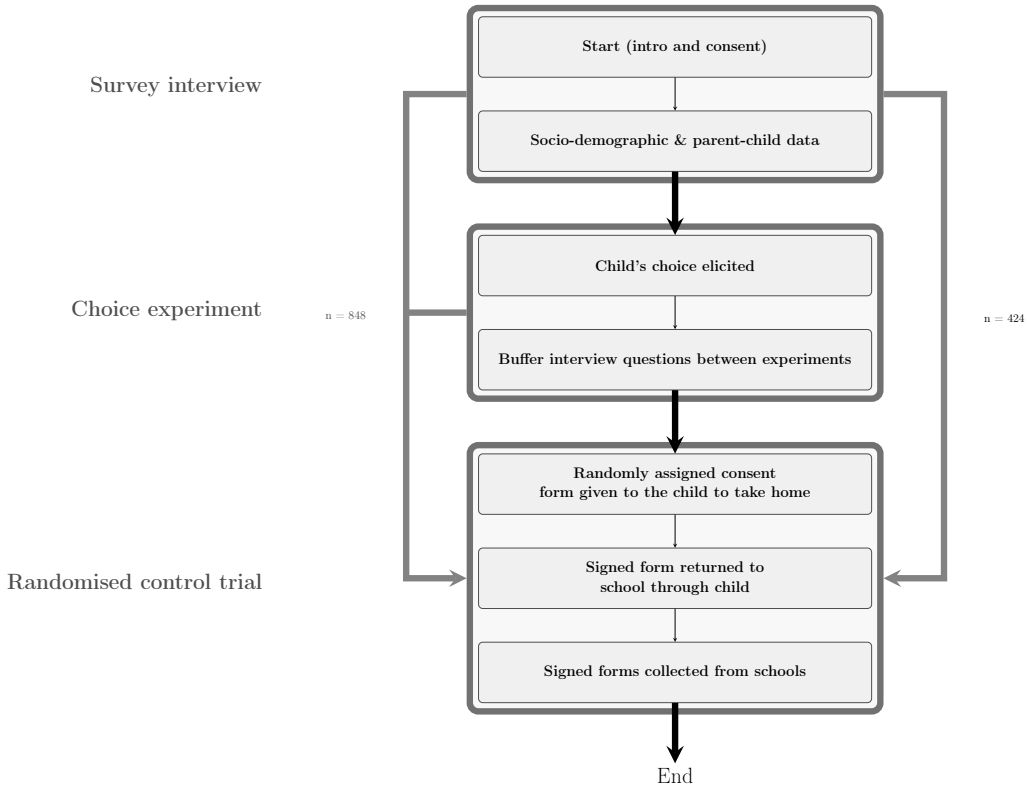
3.1. Randomised Controlled Trial (RCT)

To test whether the gender of the parent signing the consent form for swimming lessons influences the likelihood that the form is returned, the instructions given to the child were experimentally varied on which parent the consent form needed to be given to and signed by. Prior to this, the following script was read out.

[3]This observation is based on consultations with management staff and swim trainers from Panje, conducted by me during fieldwork in June 2024.

[4]This is not an exhaustive list of variables. For a full list, please refer to the study pre-analysis plan registered in the AEA RCT Registry with the unique identifying number AEARCTR-0015247 and accessible here: https://www.socialscienceregistry.org/trials/15247.

Figure 1.1: Experimental procedures for children's experiment



"We are offering you the opportunity to participate in free swimming lessons. These lessons are offered by the Panje Project, a local organisation that provides swimming lessons to children in Zanzibar. You may or may not be selected to participate. If you are selected, you will be expected to spend 3-6 hours each week for 2 weeks attending lessons. To be considered for a place to participate in these swimming lessons, you will need permission from one of your parents."

In the first treatment (T1), the child was given a 'mother-form', requiring the signature of the mother on the consent form. In the second treatment (T2), the child was given a 'father-form', requiring the signature of the father. To implement this, enumerators were required to read out the following text to the child:

"We will now randomly give you one of the two consent forms. This might be different from the one you prefer".

Children were randomised into two treatment groups.[5] Depending on the treatment, the enumerator then continued with the following script for T1 (T2):

"This form is for your MOTHER (FATHER). If you want swimming lessons, you must take it home and give it directly to HER (HIM). SHE (HE) needs to sign it. You

[5]Randomisation was achieved through the XLSform used by enumerators to conduct the interview (deployed through KoboToolbox on handheld tablets used by enumerators), which automatically randomised children into treatments by their ID number. ID numbers were randomly assigned to children before the start of the interview.

do not need the signature of your father (mother). You must return the signed form to your teacher as soon as possible, ideally within one week."

After these instructions, the enumerator recorded the child's name, gender and ID number on the form, and handed it to them. One month after the interview, Panje revisited the schools and collected all forms that were signed and returned (full experimental script in Appendix C).

3.2. Choice experiment (CE)

Prior to the RCT, a simple choice experiment was conducted with a subsample of 848 children. Each child was required to make a choice. At the end of an individual interview with a same-gender enumerator, the child was asked which one of two options they preferred to take home: a consent form for free swimming lessons that needs to be signed by the mother ('mother-form') or a consent form that needs to be signed by the father ('father-form'). This allowed us to capture which parent they preferred the form to be signed by.[6] Before this choice was made, all children received a further explanation of swimming lessons using the following script:

"Before we proceed, here's some information about these swimming lessons. Your swimming trainer will be someone of your same gender, and the lessons will take place in the open ocean."

We experimentally varied how swimming lessons were presented to children along two dimensions—**Risk** and **Conflict**—using a 2 × 2 factorial design (Figure 1.2).

Figure 1.2: Experimental treatments for choice experiment

		Risk framing	
		Neutral risk framing	Risk framing
Conflict framing	Neutral conflict framing	T_1	T_2
	Conflict framing	T_3	T_4

Along the **risk dimension**, swimming lessons were presented using two different framings. The first was the following neutral risk framing, which was read out after the general introduction:

"While the ocean is different from playing on school grounds because it involves water, it's just another environment where you can have fun and learn as long as you pay attention to simple safety rules".

[6] In a pilot, we included a third option (not choosing a form). As none of the children opted for this form, we did not include this third option in the final design. Still, to control for swimming preferences, children were asked in the interview whether they wanted swimming lessons. Over 95% confirmed that they wanted them, as reported in Table 1.1.

In the second framing, the wording of the explanation was adjusted as follows, to highlight the potential dangers and hazards to be aware of:

> *"The ocean is not like a swimming pool, and can be a dangerous place with strong currents, sudden waves, and deep water. It's crucial to follow all safety instructions from your swimming trainer. Not doing so could lead to serious consequences."*

Along the **conflict dimension**, the neutral risk framing statement was followed by a conflict framing through additional text that also got children to consider inter-parental conflict in decision-making processes around swimming. We used the following text for the neutral conflict framing:

> *"As you are still a minor, we require the consent of one of your parents. We understand that it is sometimes difficult for children to get signatures from both parents, and so we are happy to accept just one signature"*

For the conflict framing, the wording of the explanation was adjusted:

> *"As you are still a minor, we require the consent of one of your parents. We understand that this might create the possibility for disagreements between your parents since we do not ask for both signatures, but we are happy to accept just one signature."*

After the choice was made, and following some buffer interview questions, children moved to the RCT portion of the experiment, where it was explained to them that they would receive one of the two consent forms, and that this might be different from the one they preferred.

3.3. Theoretical model and hypotheses

In this section, we will present a basic model to illustrate how parent's and children's preferences influence the child's decision (which parent to involve) and the parent's decision (whether to provide consent). We will use this model to develop some hypotheses around treatment effects. We end by laying out the empirical strategy to test these hypotheses.

Model

We start with the parent's decision made in the RCT. Here, we assume that any child $c \in \{S, D\}$—son or daughter—wants swimming lessons. In order to register interest for swimming lessons, the child must secure consent from one parent, and return the signed form. The child's parent $p \in \{M, F\}$—either mother or father—must decide on behalf of the child c, by choosing $x \in \{A_c, B_c\}$, with A_c being consenting to swimming lessons for child c and B_c being not consenting. Parent p will choose x that maximises their utility U. Thus, they will consent to swimming lessons if $\Delta U_p = U_p(A_c) - U_p(B_c) > 0$.

Prior to the parent deciding in the RCT, a child participating in the CE must choose between the mother-form or the father-form, which requires the signature of either the mother or the father, respectively. As we assume a child does not have complete information about each parent's

preference, each parent's preference is 'perceived' by the child as $\widehat{U}$. Thus, where a child is given the option to choose a parental consent form, and given their preference for swimming lessons, a child will choose the parent with a positive utility difference $\Delta\widehat{U}_p$. So, if $\Delta\widehat{U}_M > 0$ and $\Delta\widehat{U}_P < 0$, the child's preference ordering for the consent form would be $Mother \succ Father$, and vice versa. If the child perceives both parents as having the same preference for consenting to swimming lessons (i.e., both parents have a positive utility difference), or the same preference for not consenting to swimming lessons (i.e., both have a negative utility difference), the child is indifferent between the two.

Under the risk framing in T2, we assume that the child's perception of each parent's utility function changes depending on how sensitive to risk they expect the parent to be. That is, the child's preference ordering between both parents would change if the child believes that one parent is more sensitive to risk than the other. We illustrate this with an example. Let's assume that the child believes the mother is sensitive to risks while the father is not. If both parents preferred to sign the form under a neutral framing, the father would still do so under the risk-framing, i.e., $\Delta\widehat{U}_F^{T2} > 0$, but it is likely that the mother would not sign the form anymore, i.e., $\Delta\widehat{U}_M^{T2} < 0$, since she is sensitive to the risks of swimming lessons. Therefore, the child's preference ordering between the parental consent forms would now be $Father \succ Mother$, switched from $Father \sim Mother$ in the neutral framing.

We apply the same logic for the conflict framing in T3 where we now assume the child considers the increased likelihood of spousal conflict. Again, the likelihood of a parent signing the consent form decreases if they perceive that doing so would cause some conflict between that parent and their spouse. Thus, the child's preference ordering would be affected by whether they perceive one parent as being more sensitive to conflict than the other. For example, assume that the child believes the mother is more conflict-averse than the father. If both parents prefer to sign the form under the neutral framing, the child would perceive that the father would still do so under the conflict-framing, i.e., $\Delta\widehat{U}_F^{T3} > 0$, but that the mother would not any more, i.e., $\Delta\widehat{U}_M^{T3} < 0$, because of her conflict aversion. Once again, the child's preference ordering between the parental consent forms becomes $Father \succ Mother$, whereas it was $Father \sim Mother$ in the neutral framing.

Hypotheses

Based on our model, we make the following predictions for the treatment effects in RCT and CE.

Form returns under the RCT: Our main outcome variable of interest is whether the consent form is returned to the school. To develop hypotheses about treatment differences, we assume that (1) the child follows the instructions given to them, and (2) the parent who receives the form does not consult their spouse. In Section 7, we discuss the implications of these assumptions not holding for the interpretation of our results.

Hypothesis 1.1: *Uptake of swimming lessons is higher in T1 than T2 for girls.*

We expect that, on average, mothers will have a stronger preference than fathers for their daughters to learn to swim. That is, we expect more mothers to have $U_M(A_D) > 0$ than

fathers. Put differently, we expect more fathers to have $U_F(A_D) < 0$ than mothers. This is consistent with evidence provided by Dizon-Ross and Jayachandran (2023) that fathers have a lower willingness-to-pay (WTP) for their daughters' human capital than mothers. Thus in the RCT, we expect girls in T1 (mother-form) to have higher return rates than girls in T2 (father-form).

Hypothesis 1.2: *Uptake of swimming lessons is higher in T2 than T1 for boys.*

We expect that, on average, fathers will have a stronger preference than mothers for their sons to learn to swim. That is, we expect more fathers to show $U_F(A_S) > 0$ compared to mothers, and more mothers to show $U_M(A_S) < 0$ compared to fathers. This too is in line with empirical evidence across a range of contexts that fathers have a stronger preference for sons' development than mothers. (Godoy et al., 2006; Lundberg and Rose, 2002; Thomas, 1994). Therefore, boys in T2 are expected to have higher return rates than in T1.

Form choices under the CE: Our main outcome variable of interest is the child's form choice between a mother-form and a father-form. We expect the following outcomes.

Hypothesis 1.3: *Children choose father-forms more often in T2 than T1.*

When the framing changes from neutral (T1) to high-risk (T2), we expect children to adjust their perceptions of their parents' preferences for consenting to swimming. Specifically, assuming that most children perceive their mother as more risk-averse than their father[7], we expect the increase in the proportion of children who believe $\Delta\widehat{U}_M < 0$ to be higher than the increase in the proportion of children who believe $\Delta\widehat{U}_F < 0$. Thus, we expect children to choose the father-form more often in T2 than T1.

Hypothesis 1.4: *Children choose father-forms more often in T3 than T1.*

When the framing changes from neutral (T1) to high-conflict (T3), we expect children to adjust their perceptions of their parents' preferences for consenting to swimming. Specifically, assuming that most children perceive their mother as more conflict-averse than their father[8], we expect the increase in the proportion of children who believe $\Delta\widehat{U}_M < 0$ to be higher than the increase in the proportion of children who believe $\Delta\widehat{U}_F < 0$. Thus, we expect children to choose the father-form more often in T3 than T1.

Hypothesis 1.5: *Children choose father-forms more often in T4 than in T1, T2, or T3.*

Combining the risk and conflict framings, given the described rationales under hypotheses 1.3 and 1.4, we expect a reinforcing effect to lead to a higher preference for father-forms in T4 compared to T1, T2 and T3.

[7]This assumption is based on evidence from risk elicitation experiments which find that, on average, women are more risk-averse than men (Croson and Gneezy, 2009; Falk et al., 2018).

[8]There is some empirical evidence to suggest that wives are the more conflict-averse spouse (e.g., see D'Exelle and Ringdal (2025)).

3.4. Empirical strategy

To test our hypotheses, we will use a series of regressions. In the RCT, the main outcome variable of interest is whether the consent form is signed and returned to the school, while the treatment variable records whether the child is instructed to approach the mother or the father for a signature on the consent form. To test for treatment differences as specified by hypotheses 1.1 and 1.2, we will estimate the following regression equation with OLS, separately for boys and girls:

$$Y_i = \beta_0 + \beta_1 \mathrm{T}_i + \beta_2 \mathbf{X}_i + \mu_e + \sigma_s + \varepsilon_i \tag{1.1}$$

where for each child i, Y_i is equal to one if the form is signed and returned, zero otherwise, and T_i is equal to one if the child received a father-form and zero if they received a mother-form. Coefficient β_1 estimates the difference between both treatments of the probability that the form is signed and returned. To increase efficiency of the estimates, we include fixed effects for the enumerator who conducted the interview (μ_e for enumerator e), the school the interview was conducted at (σ_s for school s), and a vector of control variables $\mathbf{X}_i$. As controls, we include the child's age, their attitude to risk, their preference for competition, their preference for swimming lessons, both parents' education and occupations, how often they see both parents at home, their perceptions of both parents' abilities to swim, their perceptions of both parents' attitudes to risk preferences, their perceptions on whether their parents argue or not, and their perceptions of relative parenting styles, including which parent they think is more likely to scold, punish, help, worry, be more cautious, or be more likely to grant permission for swimming lessons. ε_i is the error term. Given the multi-stage sampling approach we used, the error term might be correlated within schools. Thus, to handle non-independencies between respondents of the same school, which could underestimate standard errors, we cluster standard errors at the school level.

To investigate the role of child agency in parental decisions, we will conduct a heterogeneity analysis of the RCT treatment differences which links the RCT with the CE. Specifically, we will run:

$$Y_i = \beta_0 + \beta_1 \mathrm{T}_i + \beta_2 \mathrm{Var}_i + \beta_3 (\mathrm{T}_i \times \mathrm{Var}_i) + \mu_e + \sigma_s + \varepsilon_i \tag{1.2}$$

where Var_i is one of two variables of interest: (1) a binary variable indicating whether the child received the form they chose in the CE, and (2) a set of treatment indicators for the CE treatment the child was assigned to. We run separate regressions for each variable, where $\mathrm{T}_i \times \mathrm{Var}_i$ is an interaction term between each variable and the RCT treatment indicator. Again, we estimate separate regressions for girls and boys, using OLS with standard errors clustered at the school level.

To test for treatment effects in the CE, as specified by hypotheses 1.3, 1.4 and 1.5, we will estimate the following equation using OLS:

$$Y_i = \beta_1 + \beta_2 \mathrm{T2} + \beta_3 \mathrm{T3}_i + \beta_4 \mathrm{T4}_i + \beta_5 \mathbf{X}_i + \mu_e + \sigma_s + \varepsilon_i \tag{1.3}$$

where for each child i, Y_i is equal to one if the father-form is chosen, zero if the mother-form is chosen. In this model, T1 is the reference group. T2, T3, and T4 are dummy variables for the respective treatment groups, where coefficients β_2, β_3 and β_4 estimate the difference in the probability of the child choosing the father-form between T2, T3, and T4 respectively, and T1. To increase efficiency of the estimates, we include fixed effects for the enumerator who conducted the interview (μ_e for enumerator e), the school the interview was conducted at (σ_s for school s), and a vector of control variables $\mathbf{X}_i$, which includes the same list of controls specified above for the RCT regression models. ε_i is the error term. We cluster standard errors at the school level, and estimate separate regressions for girls and boys.

To examine heterogeneity of treatment effects in the CE, we estimate the following regression:

$$Y_c = \beta_1 + \sum_{j=2}^{4} \beta_j \mathrm{T}_j + \mathrm{Var} + \sum_{j=2}^{4} (\gamma_j \mathrm{T}_j \times \mathrm{Var}) + \varepsilon_i \tag{1.4}$$

which is the same regression as equation 1.3, but with the inclusion of Var_i, a variable that relates to child perceptions around the following dimensions: parents' relative attitudes to risk, how parents interact, household norms around permission-seeking, and parenting styles. $T_j \times \mathrm{Var}$ is an interaction term between each of the treatment indicator variables (where $j = 2, 3, 4$) and the heterogeneity variable of interest. Again, ε_i is the error term and we estimate the regression using OLS with standard errors clustered at the school level. To interpret heterogeneity in CE treatment effects, we focus on the coefficients of the interaction terms γ_j.

Note that all CE regressions have binary dependent variables. The use of OLS in this case differs from the commonly used logit regression approach. We do this for several reasons. It accommodates the inclusion of school fixed effects, allows for an easier interpretation of the magnitude of treatment effects, and enables an examination of heterogeneity across treatments that does not depend on the specific values assigned to the covariates in the model. However, we are aware that this approach might have its own limitations, and will therefore check the robustness of the results by comparing them with those derived from logit regressions.

3.5. Sampling and implementation

We recruited a sample of 1,272 children through the following sampling approach. We partnered with a local non-profit organisation in Zanzibar known as 'The Panje Project' (Panje henceforth), who offer free aquatic survival training to Zanzibari children. Panje provided us with a list of all their target schools for 2025, which included 62 primary schools and secondary schools covering all districts in Unguja. Next, we randomly selected 13 primary and 13 secondary schools. Panje then asked each school to provide them with a list of 100 eligible students. Eligibility criteria were being aged 8-18, not knowing how to swim, and living with both parents. We randomly selected between 30-50 children from each school, selecting a higher number of children in secondary schools to ensure a larger number of mature respondents, and stratifying the sample by gender.

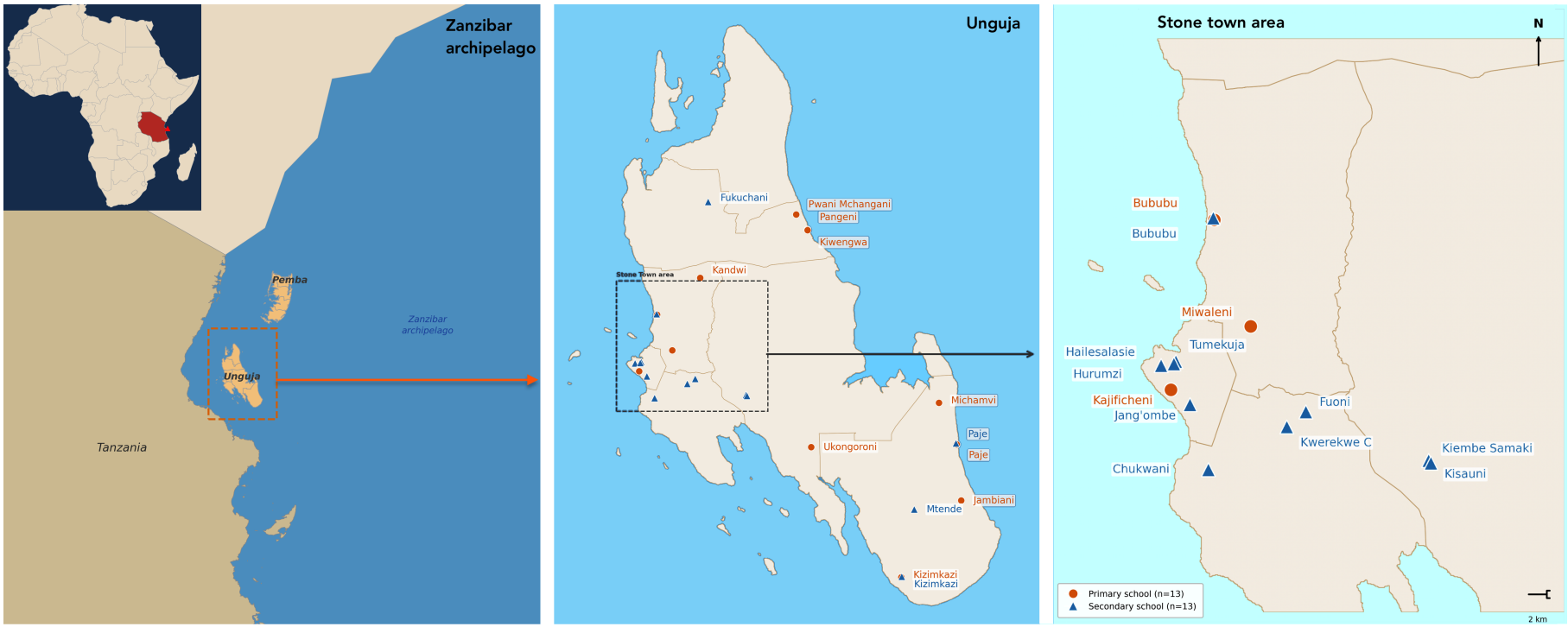


Figure 1.3: Map of Zanzibar

(a) *Zanzibar archipelago,* ***(b)*** *sampled districts and school locations on Unguja island, and* ***(c)*** *detailed school locations within the Stone Town area. Orange circles denote primary schools; blue triangles denote secondary schools. Unguja inset in* ***(a)*** *shows that the northern Zanzibar island Pemba was not included in the study.*

Data collection for this study was conducted in February and March, 2025. Prior to data collection, a team of enumerators were trained on conducting data collection using Computer-Assisted Personal Interviewing (CAPI), with emphasis on the need for child-friendly interviewing approaches that prioritised seeking informed assent, ensuring confidentiality, establishing rapport, and maintaining a non-judgmental atmosphere to ensure the emotional well-being of the child. Interviews began in the second week of February and were completed by the end of February. Forms were collected by Panje from the schools in the last week of March, which allowed schools a month to gather the returned forms.

On the day of the interview, the school arranged for all selected students to assemble in a large classroom. Upon arrival, each student was asked to sign a registry of attendance after which they were given a sticker with their ID number for them to wear on their chest. After this, each child was directed to, and interviewed individually by, a same-gender trained enumerator in a private space. After the interview and choice experiment was complete, a consent form for swimming lessons was provided as part of the RCT, along with a compensation gift for their time. A month after the interview had been conducted, Panje visited the school to collect the returned consent forms. After evaluating all forms, Panje used a lottery system to grant swimming places, then contacted the selected children through the school for final confirmation of swimming lessons. All children contacted then participated in an introductory 2-week swimming training course.

3.6. Ethics

Ethical approval and research permission process

Ethical approval for this study was granted on 9 January 2025 by the School of Global Development Research Ethics Subcommittee (DEV S-REC) at the University of East Anglia (UEA) (ID: ETH2425-0204). The full ethics application and supporting documentation can be made available upon request. Research permits were further secured from the Office of the Second Vice President of Zanzibar and district-level offices of the Ministry of Education. The study adheres to established principles on research with children, including non-maleficence (doing no harm), beneficence (research that improves wellbeing), respect for autonomy, justice, and prioritising the child's best interests (King and Churchill, 2000). Key safeguards based on the Ethics Research Involving Children (ERIC) guidance of Graham et al. (2013) included 1) ensuring voluntary participation, confidentiality and appropriate compensation, and 2) minimising psychological distress from the experimental design. We elaborate on these in more detail below.

Consent, assent, confidentiality, and compensation

Given the experimental design, direct prior parental consent for the child to be interviewed was not feasible as it would have compromised the randomisation of parental involvement. Consequently, an institutional consent model was adopted. Precisely, in lieu of parental consent, permission was obtained through formal approval from District Education Officers, headteachers, and classroom teachers, consistent with standard practice for school-based research in Zanzibar and globally (Robinson,

2025). Individual child assent was then obtained through age-appropriate, non-coercive explanations. Enumerators were trained to monitor for distress and to terminate interviews immediately without penalty to the participant if they showed any signs of unwillingness to participate. To maintain confidentiality, data were anonymised using unique IDs; no names were recorded on survey instruments. The study was incentivised, as swimming lessons constituted a real and valuable intervention. However, no financial payments were made. Still, to acknowledge children's time in a manner consistent with ethical guidelines for research with minors, participants received a small non-monetary token (a school notebook and pen).

Mitigating experimental risks

The behavioural experiments involved framing effects that required specific safeguards to minimise psychological or social risk to the child:

(i) Parental preferences: To prevent emotional distress when asking children which parent they 'prefer', the script emphasised that the choice was hypothetical. Moreover, children were informed from the outset that they might not receive the form they selected, a procedure common in school environments to manage expectations and reduce potential disappointment.

(ii) Experimental framings: In the CE 'high-conflict' framing, by making salient that the requirement to involve only one parent originated from the *research protocol* rather than the child's own actions, children were explicitly absolved of responsibility for potential parental disagreement. Additionally, the 'high-risk' framing was intended to make risk salient for the experiment. It was not seen as inherently alarmist, rather, it aligned with standard water safety pedagogy and reflected the real-world gravity of aquatic survival training without introducing disproportionate psychological distress.

(iii) Sensitivity of survey instruments: The survey included a question regarding whether parents argue about the child to examine if perceived household conflict moderated experimental responses. While sensitive, such questions are standard in child wellbeing research to capture lived experiences that other methods might obscure. Enumerators were specifically trained to approach this question with caution, monitoring for signs of discomfort and maintaining the authority to terminate the interview immediately if the child appeared distressed.

(iv) Sensitivity of swimming as activity: While swimming carries gendered norms in Zanzibar, scoping work and consultations with the Panje Project confirmed that swimming is not widely framed as morally prohibited for children in Zanzibar, even if it has historically been uncommon. Rather, as with any extracurricular activity, parental resistance is primarily economic (opportunity cost of time) rather than cultural. Still, this was the first time swimming lessons had been formally offered to participants. Thus, to alleviate concerns children might have had regarding breaching the topic of swimming lessons with parents, the intervention was presented through the school as a signal of institutional legitimacy.

(v) Intervention safety: All swimming lessons were conducted by certified Panje Project instructors in accordance with RNLI-partnered safety protocols. Exclusions from the lessons (approx. 10%)

were made solely by instructors based on health and safety criteria disclosed in the parental consent forms, independent of research treatment status.

4. Summary statistics

We now turn to our empirical results. This section reports summary statistics of the sample, Section 5 presents average treatment effects, and Section 6 examines mechanisms.

We collected data from a sample of 1,272 children. Table 1.1 presents some gender-disaggregated summary statistics of key socio-demographic indicators, including two-sided p-values of t-tests that compare mean outcomes between boys and girls. Given that the sample was stratified by gender, we have the same number of boys and girls (636 children each). As we oversampled older children, the proportion of students sampled from secondary schools (55%) is higher than the proportion of students sampled from primary schools (45%).

The average age for boys and girls in this sample is 13 years, and they both have over five siblings. Crucially, over 95% of girls and boys have a preference to learn to swim, and this is statistically significantly different between boys and girls, with more boys than girls expressing a preference to learn to swim. We elicited children's competitiveness, their risk preferences, as well as what they think their father's and mother's risk preferences are, using a simple continuous scale from 0-10. On average, boys and girls score the same on competitiveness (7.3), but boys view themselves as more risk-seeking than girls do (risk-seeking score of 6.4 vs 5.1). Boys also rate their fathers' and mothers' risk-seeking scores lower than girls do (6.1 vs 5.1 for boys, 7.4 vs 5.7 for girls). Comparing these absolute risk scores, we see that on average, boys and girls view their mothers as more risk-averse than their fathers. This is in line with the assumption made under hypothesis 1.3. We also construct three binary variables to capture children's perceptions of their parents' relative attitudes to risk: 'Mother more risk-averse', 'Mother more cautious', and 'Father more cautious' (the latter two use a more general measure of risk aversion—please refer to notes under Table 1.1 for variable definitions). Using these relative indicators, we find more girls than boys with 'Mother more risk-averse' scores (58% vs 51%). Using the general questions, we find that boys are more likely than girls to report that their mother is more cautious (56% vs 51%), although this difference is not statistically significant. Where children view the father as more cautious than the mother, boys are more likely than girls to think so (16% vs 9%).

Next, we report on binary variables related to child perceptions around parental interactions. ∼76% of boys and girls think that in decisions made about the child, their 'father involves mother'. Interestingly, boys and girls perceive conflict between parents differently, with girls more likely than boys to perceive that their parents argue (39% vs 34%), but boys more likely than girls to perceive that their parents argue about the child themself (19% vs 11%). Girls are more likely than boys to report that the parent they go to for help is the mother (72% vs 58%), but boys are more likely than girls to seek permission from the father than mother (31% vs 26%). Taken together, the latter two findings suggest that girls are closer to their mother and boys closer to their father. Still, we find girls more likely than boys to see the father as the final decision-maker (52% vs 44%), although

Table 1.1: Descriptive statistics

Variable	(1) Boys Mean/(SD)	(2) Girls Mean/(SD)	(1)-(2) Pairwise t-test P-value
Education	0.450	0.453	0.910
	(0.498)	(0.498)	
Age	13.583	13.374	0.067*
	(2.107)	(1.957)	
Siblings	5.432	5.818	0.038**
	(3.174)	(3.430)	
Preference for swimming lessons	0.975	0.953	0.036**
	(0.157)	(0.212)	
Child's self-score of competition	7.289	7.297	0.954
	(2.466)	(2.430)	
Child's self-score of risk	6.355	5.063	0.000***
	(2.703)	(2.614)	
Child's risk score of father	6.099	7.373	0.000***
	(2.471)	(2.506)	
Child's risk score of mother	5.080	5.657	0.000***
	(2.687)	(2.587)	
Mother more risk-averse	0.513	0.577	0.021**
	(0.500)	(0.494)	
Father more cautious	0.157	0.088	0.000***
	(0.364)	(0.284)	
Mother more cautious	0.557	0.513	0.116
	(0.497)	(0.500)	
Father involves mother	0.758	0.770	0.598
	(0.429)	(0.421)	
Parents argue	0.340	0.393	0.048**
	(0.474)	(0.489)	
Parents argue about child	0.195	0.112	0.000***
	(0.396)	(0.315)	
Seek help from mother	0.577	0.715	0.000***
	(0.494)	(0.452)	
Seek permission from father	0.314	0.259	0.030**
	(0.465)	(0.439)	
Father final decision-maker	0.436	0.519	0.003***
	(0.496)	(0.500)	
Own	0.605	0.586	0.493
	(0.489)	(0.493)	
Reason	0.884	0.876	0.667
	(0.321)	(0.330)	
Views	0.560	0.505	0.049**
	(0.497)	(0.500)	
Receive preferred form	0.346	0.343	0.906
	(0.476)	(0.475)	
N	636	636	1272

Notes: Summary statistics of the following variables: 'Education' which is a dummy taking the value one if they are in primary school, zero if they are not; 'Age' which captures the child's age; 'Siblings' which is the total number of siblings the child has; 'Preference for swimming lessons (y/n)' which is a dummy taking the value one if the child answers yes to the question 'Would you like to learn to swim?', zero if the answer is no; Child's self-score of competition' which is a continuous variable that asks the child to consider how willing they are to compete on a scale of 0-10, where zero means 'not willing to compete', and 10 means 'very willing to compete'; 'Child's self-score of risk' which is a continuous variable that asks the child to consider how they view their willingness to take risk on a scale of 0-10, where zero means 'not willing to take risk at all', and 10 means 'very willing to take risk'; 'Child's risk score of father/mother' which offers the same scale, but to capture how the child views their father's/mother's willingness to take risk; 'Mother more risk-averse' which is a binary variable with value one if the mother's risk score is lower than the father's risk score, zero otherwise; 'Father/Mother more cautious' which are two binary variables with value one if the child answers with 'Father/Mother' to the question 'Which parent is more cautious in life?', zero otherwise; 'Father involves mother' which is a binary variable taking the value one if the child answers yes to the question 'For any decision your father takes, does he always involve your mother?', and zero otherwise; 'Parents argue' which is a binary variable with value one if the child answers 'false' to the statement 'My parents never argue', and zero otherwise; 'Parents argue about child' which is a binary variable taking the value one if the child answers 'true' to the statement 'My parents usually argue or disagree because of things that I do', zero otherwise; 'Seek help from mother' which is a binary variable with value one if the child answers 'mother' to the question 'Which parent do you prefer to go to when you need help with a problem?', zero otherwise; 'Seek permission from father' which is a binary variable with value one if the child answers 'father' to the question 'When you need permission for school-related activities, who do you generally ask?', zero otherwise; 'Father final decision-maker' which is a binary variable with value one if the child answers 'father' to the question 'Which parent makes the final decision in the school-related case?', zero otherwise; 'Own' which is a binary variable with value one if the child agrees with the statement 'My parents give me many opportunities to make my own decisions about what I am doing', zero if they do not; 'Reason' which is a binary variable with value one if the child agrees with the statement 'When I ask why I have to do, or not do, something, my parents give me good reasons', zero if they do not; and 'Views' which is a binary variable with value one if the child agrees with the statement 'My point of view is very important to my parents when they make important decisions', zero if they do not. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

we note that this is close to half in both cases, which suggests that children perceive both parents as equally likely to be final decision-makers.

We also report on three binary variables that capture children's perceptions of parenting styles. By asking them whether they agree or not with statements related to parents allowing the child to make their own decisions ('Own'), parents giving them good reasons for parental decisions ('Reason'), and parents taking the child's views into account in decision-making ('Views'). We note that the only statistically significant difference between genders is detected for 'Views', where girls are less likely than boys to agree that their views are taken into account in parental decision-making (50% vs 56%). We also note that children feel empowered in decision-making, with high proportions agreeing that they have the opportunity to make their own decisions (∼60% agree for the 'Own' variable).

Last, we derive summary statistics for categorical variables used as regression controls (presented in Table A.1 of Appendix A). These include children's knowledge of their parents' education levels (∼50% of children to not know their parents' education level) and swimming ability (∼40% know their father can swim, and ∼70% know their mother cannot swim), and how often children see each of their parents at home (∼90% see their parents at home everyday). Pearson's χ^2 for test of statistical difference between categorical variables of boys and girls find that the only variable where children's responses are not statistically different is the variable which records how often they see their mother at home.

Since children were randomly assigned to the experimental treatments in the RCT and the CE, we expect differences in characteristics between treatments to be due to chance, for both boys and girls. Appendix B presents balancing tables for the RCT and CE, on the same variables reported in Table 1.1. Using a pairwise t-test on mean differences, Table B.1 shows that treatment assignment for girls in the RCT is balanced for all variables. Table B.2 shows that treatment assignment for boys along the same list of variables is also balanced, except for boys' responses on whether their parents argue about the child (F=4.870; two-sided $p = 0.028$) and whether they seek help from their mother (F=2.845; two-sided $p = 0.092$). All variables that are not balanced between treatments are included as controls in regressions examining RCT treatment differences.

Using an F-test for balance across all four treatments in the CE, Table B.3 shows that treatment assignment for girls in the CE is balanced on all characteristics except girls' responses to whether the father involves the mother in decisions (F=2.122; two-sided $p = 0.097$). For boys, Table B.4 again shows balanced treatment assignment, and that we fail to reject the null of differences across all CE treatments on most variables except boys' risk scores of their mothers (F=2.804; two-sided $p = 0.039$), boys' risk scores of their fathers (F=3.719; two-sided $p = 0.012$) and their response regarding their father being the final decision-maker (F=3.170; two-sided $p = 0.024$). Again, all variables that are not balanced are included as controls in regressions examining CE treatment effects.

5. Average treatment effects

In the following sections, we present the average treatment effects in the RCT and CE. As these are estimated using OLS, as robustness checks we also present the same regressions estimated using logit (see Tables in Appendix D). A comparison of results using these two alternative approaches is offered as part of our discussion in Section 7.

5.1. RCT

In the RCT, the main outcome variable is a binary variable equal to one if the consent form is returned to the school, zero otherwise. We estimate differences in the likelihood that the form is returned between treatments T1 (mother-form) and T2 (father-form), using equation 1.1. We run the same regression on the pooled sample, and the subsamples of boys and girls. We also separate the analyses for the RCT-only sample of 424 children, and the full sample of 1,272 children. These results are presented in Table 1.2.

Table 1.2: RCT treatment differences in form returns, by child's gender

	RCT-only					Full sample		
	(1) Pooled	(2) Girls	(3) Boys	(4) Girls > 10	(5) Boys > 10	(6) Pooled	(7) Girls > 10	(8) Boys > 10
RCT T2: Father-form	-0.050	-0.125*	-0.009	-0.167**	-0.049	-0.030	-0.055	0.006
	(0.056)	(0.073)	(0.073)	(0.081)	(0.096)	(0.022)	(0.037)	(0.030)
Predicted probability under T1: Mother-form	0.23	0.33	0.15	0.36	0.16	0.23	0.31	0.16
N	424	212	212	194	197	1272	584	583

Notes: OLS regressions with the dependent variable equal to one if the child's consent form is returned, zero otherwise. The RCT treatment dummy takes the value one if the child received a father-form, and zero if they received a mother-form. Models 1-5 are run on the RCT-only sample of 424 children, and 6-8 on the full sample of 1,272 children. All regressions include school and enumerator fixed effects, and a set of control variables : the child's age, their attitude to risk, their preference for competition, their preference for swimming lessons, both parents' education and occupations, how often they see both parents at home, their perceptions of both parents' abilities to swim, their perceptions of both parents' attitudes to risk preferences, their perceptions on whether their parents argue or not, and their perceptions of relative parenting styles, including which parent they think is more likely to scold, punish, help, worry, be more cautious, or be more likely to grant permission for swimming lessons. We include two binary variables that are unbalanced across RCT treatments for boys: boys' responses on whether their parents argue about the child and whether they seek help from their mother. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The predicted probability that a mother-form (T1) is returned is 23% for the pooled RCT-only sample (column 1), and this is not significantly different from treatment T2, where the father-form is used. However, when we look at the gender-disaggregated subsamples, there are statistically significant differences between both treatments. Specifically, in column 2, we observe that girls who do not participate in the CE (the RCT-only sample) are 12.5 percentage points or 38% less likely to return the father-form (T2) than the mother-form (T1), and this increases to 16.7 percentage points or a 46% reduction for girls above 10 years (column 4). We do not detect any treatment differences for girls in the full sample, nor for boys in any of the samples. We also note that the probability of the form being returned is already quite low for boys in T1 compared to girls (∼15% vs ∼33%, i.e., ∼55% less). Taken together, these findings confirm hypothesis 1.1 that uptake of swimming lessons is higher in T1 than T2 for girls, but do not support hypothesis 1.2 that the uptake of swimming lessons is higher in T2 than T1 for boys.

5.2. CE

In the CE conducted with the subsample of 848 children, the main outcome variable of interest is the child's form choice, which takes the value one if the child chose a father-form and zero if they chose a mother-form. Table 1.3 reports estimates of the treatment effects on the likelihood of the child choosing the father-form.

We run separate models on the full CE sample, and the subsamples of girls and boys. We first note that the probability of the father-form being chosen in treatment T1 (neutral framing) is 35%,

Table 1.3: CE treatment effects on form choices, by gender

	(1) Pooled	(2) Girls	(3) Boys	(4) Girls > 10	(5) Boys > 10
CE T2: Risk	0.038	-0.108	0.145**	-0.094	0.171***
	(0.045)	(0.067)	(0.055)	(0.070)	(0.060)
CE T3: Conflict	0.009	-0.155**	0.167**	-0.146**	0.162**
	(0.031)	(0.061)	(0.069)	(0.064)	(0.066)
CE T4: Both	-0.066	-0.161**	0.068	-0.130*	0.119
	(0.040)	(0.069)	(0.066)	(0.075)	(0.070)
Predicted mean under T1: Neutral	0.35	0.42	0.28	0.40	0.26
F-test: T2=T3	0.524	0.492	0.721	0.462	0.875
F-test: T2=T4	0.018	0.394	0.291	0.570	0.478
F-test: T3=T4	0.060	0.926	0.103	0.832	0.417
N	848	424	424	390	386

Notes: OLS regressions with the dependent variable equal to one if the child chooses the father-form, zero if they choose the mother-form. All regressions include school and enumerator fixed effects, and a set of control variables: the child's age, their attitude to risk, their preference for competition, their preference for swimming lessons, both parents' education and occupations, how often they see both parents at home, their perceptions of both parents' abilities to swim, their perceptions of both parents' attitudes to risk preferences, their perceptions on whether their parents argue or not, and their perceptions of relative parenting styles, including which parent they think is more likely to scold, punish, help, worry, be more cautious, or be more likely to grant permission for swimming lessons. We also include the binary variable capturing child views on who the final decision-maker is, as this was unbalanced for boys in the CE. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

42%, and 28%, for the pooled, girls-only, and boys-only samples, respectively. Looking at the pooled sample (column 1), we find no statistically significant treatment effects of T2, T3, or T4.

When we examine treatment effects disaggregated by gender, the results reveal that the pooled sample masks opposing tendencies amongst boys and girls. Specifically, we find that girls are less likely to select the father-form in T3 and T4 compared to T1 (column 2, negative and significant coefficients on T3 and T4, respectively), whilst boys are more likely to select the father-form in T2 and T3 compared to T1 (column 3, positive and significant coefficients on T2 and T3, respectively).

When we reinterpret our hypotheses in light of these differences, we find that boys' choices align with the original predictions: they are more likely to select the father-form in T2 and T3 compared to T1, consistent with hypotheses 1.3 and 1.4. Girls, in contrast, display the opposite pattern, being less likely to choose the father-form in T3 and T4 relative to T1—thus contradicting hypotheses 1.4 and 1.5. In sum, these results suggest that for girls, it is the conflict framing and not the risk framing which is influential in shifting preferences, and for boys, either the risk or conflict framing (but not both combined) have an influence on preferences.

We also note that the additional effect of combining both risk and conflict framings, compared to just one framing, yields some interesting results in the pooled sample as well as the subsamples. In the pooled sample, as the coefficients of T2 and T4 are statistically different from each other (two-sided F-test $p = 0.018$), taking the difference of the coefficients, we find that the probability of a child choosing the father-form is ∼10.4 percentage points lower in treatment T4 than T2. Similarly, as the coefficients of treatments T3 and T4 are statistically different (two-sided F-test $p = 0.060$), we also find the probability of a child choosing the father-form is ∼7.5 percentage points lower in treatment T4 than T3 (taking the difference of the coefficients). In the case of boys, we also find boys are ∼9.9 percentage points *less* likely to choose the father-form in T4 compared to T3 (column

3, taking the difference of the T4 and T3 coefficients yields two-sided F-test $p = 0.103$). These results are contrary to what we expect as part of hypothesis 1.5, which anticipates a reinforcing effect of T2 and T3 in T4.

Before proceeding to mechanisms, it is important to note that the average treatment effects from the CE are in line with results from the same regressions run with logit specifications as robustness checks (reported in Appendix D).

6. Mechanisms

6.1. For CE treatment effects

In this section, we examine a range of mechanisms that could help explain differences in CE treatment effects between girls and boys. These mechanisms include children's perceptions of:

i) their parents' attitudes to risk,
ii) how both parents interact with each other,
iii) parental involvement in school-related decisions,
iv) their agency in their parents' decision-making.

Insights on mechanisms (i) and (ii) would shed light on treatment effects in T2 and T3, respectively, and mechanisms (iii) and (iv) on how child preferences and agency moderate treatment effects across all treatments. To ensure meaningful responses, only children above 10 years are considered.

Parents' risk attitudes

We first examine whether children's perceptions of their parents' relative risk preferences might help explain shifting choices in T2 (where the risk of swimming is highlighted). We use three binary variables to capture children's perceptions of their parents' relative attitudes to risk: 'Mother more risk-averse', 'Father more cautious', and 'Mother more cautious'. We construct the first variable using absolute risk scores, where children are asked to consider how they view their father's/mother's willingness to take risk on a scale of 0-10, where zero means 'not willing to take risk at all', and 10 means 'very willing to take risk'. 'Mother more risk-averse' then takes the value one if the mother's risk score is lower than the father's risk score, zero otherwise. 'Father/Mother more cautious' each take the value one if the child answers with 'Father/Mother' to the question 'Which parent is more cautious in life?', zero otherwise. We interact each of these variables with the treatment indicator variables using equation 1.4. We focus on the interactions with T2 and T4, as these are the treatments which vary the risk framing of swimming lessons. Table 1.4 presents the results.

Looking at 'Father more cautious', we see that girls who believe their father is more cautious than their mother are less likely to choose the father form in the T2 and T4, respectively, than T1 (taking the sum of the treatment coefficient and its interaction term), and these are statistically significant (column 1, coef = -0.662, two-sided $p = 0.014$ for T2; coef = -0.627, two-sided $p = 0.061$ for T4). Similarly, looking at 'Mother more risk-averse', we see that girls who see the mother as the more risk-averse parent are less likely to choose the father-form in T4 than T1 (column 5, coef = -0.190, two-sided $p = 0.038$ for T4).

Table 1.4: Parents' relative risk attitudes

	Father more cautious		Mother more cautious		Mother more risk-averse	
	(1) Girls > 10	(2) Boys > 10	(3) Girls > 10	(4) Boys > 10	(5) Girls > 10	(6) Boys > 10
CE T2: Risk	-0.051	0.160**	-0.132	0.121	-0.178*	0.255***
	(0.072)	(0.072)	(0.100)	(0.108)	(0.106)	(0.098)
CE T3: Conflict	-0.116	0.142**	-0.090	0.260**	-0.100	0.267***
	(0.072)	(0.072)	(0.102)	(0.102)	(0.103)	(0.094)
CE T4: Both	-0.108	0.100	-0.106	0.188*	-0.043	0.128
	(0.068)	(0.072)	(0.104)	(0.099)	(0.102)	(0.097)
Yes	0.548**	0.119	-0.165	-0.117	0.065	0.034
	(0.229)	(0.145)	(0.117)	(0.101)	(0.105)	(0.099)
CE T2 Risk × Yes	-0.612**	0.089	0.072	0.040	0.140	-0.192
	(0.278)	(0.217)	(0.139)	(0.142)	(0.143)	(0.139)
CE T3 Conflict × Yes	-0.333	0.199	-0.107	-0.180	-0.068	-0.197
	(0.286)	(0.184)	(0.141)	(0.136)	(0.140)	(0.137)
CE T4 Both × Yes	-0.520	0.164	-0.042	-0.130	-0.148	-0.019
	(0.341)	(0.189)	(0.142)	(0.133)	(0.139)	(0.135)
Predicted mean under T1	0.41	0.25	0.40	0.26	0.40	0.26
lincom test: T2+T2 × Yes	0.014	0.217	0.533	0.064	0.688	0.509
lincom test: T3+T3 × Yes	0.103	0.046	0.041	0.373	0.076	0.473
lincom test: T4+T4 × Yes	0.061	0.127	0.104	0.513	0.038	0.242
N	390	386	390	386	390	386

Notes: OLS regressions with the dependent variable equal to one if the child chooses the father-form, zero if they choose the mother-form. The sample only includes girls and boys above 10 years old. The CE treatments are interacted with "Yes" variables, which are two dummies related to parents' relative risk preferences as perceived by the child. In models 1 and 2, "Yes" is a dummy variable taking the value one if the child does *not* respond with 'Father' to the question "Which parent do you think is more cautious in life?", and takes the value zero if the answer is 'Father'. In models 3 and 4, "Yes" is a dummy variable which takes the value one if the child rates the father as more risk-seeking than the mother, zero otherwise, where relative risk attitudes are compared by looking at the child's responses to separate quantitative measures of father and mother risk attitudes (these ask the child to consider how they rate each parent's willingness to take risk on a scale of 0-10, where zero means "not willing to take risk at all", and 10 means "very willing to take risk"). The only control variables we use here are the child's age and the child's own risk attitudes (also measured using the same risk scale used for the parents). Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

For boys, the sum of coefficients on T2/T4 and its interaction terms are not statistically significant for 'Father more cautious' (column 2), but the sum of coefficients on T2 and its interaction term are significant for 'Mother more cautious', which tells us that boys who think their mother is more cautious are *more* likely to choose the father form in T2 relative to T1 (column 4, coef = 0.160, two-sided $p = 0.064$ for T2). However, using the 'Mother more risk-averse' variable, we do not find any effect.

Taken together, the positive effects for boys in T2 in Table 1.3 appear to be driven by boys who perceive their mother as the more cautious parent when we look at 'Mother more cautious', but not when we look at 'Mother more risk-averse'. Similarly, the negative effects for girls in T4 in Table 1.3 seem to be driven by girls who perceive the father as the more cautious parent. In summary, under risk framings in T2 and T4, children appear to be discouraged from choosing the parent that they view as more cautious.

Spousal interactions

Next, we examine whether children's perceptions on how their parents interact with each other can shed light on the observed treatment effects in the CE. We focus on the interactions with T3

and T4, as these are the treatments which vary the conflict framing of swimming lessons. Table 1.5 presents regression results in which we interact CE treatments with three binary variables that capture whether the child thinks the father always involves the mother in decisions (columns 1 and 2), their parents argue frequently (columns 3 and 4), and their parents frequently argue specifically about them (columns 5 and 6).

Table 1.5: Parental interactions

	Father involves mother		Parents argue		Parents argue about child	
	(1) Girls > 10	(2) Boys > 10	(3) Girls > 10	(4) Boys > 10	(5) Girls > 10	(6) Boys > 10
CE T2: Risk	0.101	0.254*	-0.058	0.225***	-0.071	0.214***
	(0.160)	(0.153)	(0.088)	(0.080)	(0.073)	(0.075)
CE T3: Conflict	0.067	0.311**	-0.145	0.266***	-0.147**	0.175**
	(0.153)	(0.151)	(0.090)	(0.083)	(0.072)	(0.075)
CE T4: Both	0.133	0.062	-0.153*	0.206**	-0.111	0.157**
	(0.169)	(0.154)	(0.088)	(0.081)	(0.070)	(0.074)
Yes	0.208	-0.017	0.018	0.058	-0.030	0.034
	(0.137)	(0.130)	(0.108)	(0.101)	(0.158)	(0.118)
CE T2 Risk × Yes	-0.227	-0.093	-0.083	-0.160	-0.142	-0.185
	(0.176)	(0.172)	(0.142)	(0.143)	(0.203)	(0.181)
CE T3 Conflict × Yes	-0.260	-0.177	0.002	-0.297**	-0.027	0.006
	(0.173)	(0.170)	(0.146)	(0.138)	(0.236)	(0.168)
CE T4 Both × Yes	-0.316*	0.067	0.071	-0.286**	-0.195	-0.204
	(0.187)	(0.173)	(0.145)	(0.142)	(0.221)	(0.171)
Predicted mean under T1	0.39	0.25	0.40	0.26	0.41	0.25
lincom test: T2+T2 × Yes	0.096	0.033	0.208	0.577	0.264	0.856
lincom test: T3+T3 × Yes	0.015	0.079	0.205	0.781	0.441	0.228
lincom test: T4+T4 × Yes	0.013	0.078	0.461	0.477	0.142	0.754
N	390	386	390	386	390	386

Notes: OLS regressions with the dependent variable equal to one if the child chooses the father-form, zero if they choose the mother-form. The sample only includes girls and boys above 10 years old. The CE treatments are interacted with "Yes" variables capturing children's perceptions of parental interactions. In models 1 and 2, "Yes" is a dummy variable taking the value one if the child answers yes to the question "For any decision your father takes, does he always involve your mother?", and zero otherwise. In models 3 and 4, "Yes" is a dummy variable taking the value one if the child answers 'false' to the statement "My parents never argue", and zero otherwise. In models 5 and 6, "Yes" is a dummy variable taking the value one if the child answers 'true' to the statement "My parents usually argue or disagree because of things that I do", and zero otherwise. The only control variables we use here are the child's age and the child's own risk attitudes. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We first examine the variable capturing the (perceived) mother's involvement (columns 1 and 2). Taking the difference of coefficients on T3/T4 and its interactions in columns 1 (column 2), we find that girls (boys) who think their father involves their mother in all decisions are significantly less (more) likely to choose the father-form in T3/T4 relative to T1 (column 1, coef = -0.1935, two-sided $p = 0.015$ in T3; coef = -0.183, two-sided $p = 0.013$ in T4 for girls; column 2, coef = 0.133, two-sided $p = 0.079$ in T3; coef = 0.129, two-sided $p = 0.078$ in T4 for boys). Taken together, this indicates that the negative (positive) treatment effects in CE treatments for girls (boys) in T3/T4 in Table 1.3 are driven by girls (boys) who believe their father involves their mother.

We next examine the two variables on child perceptions around parents arguing in general (columns 3 and 4) and about the child specifically (columns 5 and 6). For girls, the negative coefficient on T4 in column 3, and the negative coefficient on T3 in column 5 indicate that girls who *do not* perceive their parents as prone to arguments are less likely to choose the father-form in T4 and T3 respectively, relative to T1, and this is not statistically different for girls who perceive

parents as spouses who argue, which indicates that girls' perceptions of parents arguing does not matter for treatment effects.

For boys, the positive and significant coefficients on T2, T3, and T4 in columns 4 and 6 suggest that for boys who do not think their parents argue, either in general or about them specifically, the likelihood of choosing the father form is higher in T2, T3, and T4 respectively, relative to T1. However, since the sum of these coefficients and their interaction terms are not statistically significant, we can conclude that for boys, the shift to father-form choices in T2 and T3 in Table 1.3, relative to T1, are driven by boys who perceive that parents do *not* argue.

Parental involvement in decisions

As another potential mechanism, we examine child perceptions around parental involvement in decision-making. The first variable captures who the child normally seeks help from, as this would indicate an existing preference or tendency to involve one parent over another. The second and third variables specifically relate to parental involvement in school-related decisions, as the swimming lessons are distributed via the school. One variable captures who the child normally contacts first for school-related permissions, the other records which parent the child perceives as the final decision-maker in school-related decisions. We present interactions of treatment indicators with these three binary variables in Table 1.6, which we examine in turn.

The first binary variable 'Seek help from mother', which takes the value one if the child answers with 'mother' to the question 'Which parent do you prefer to go to when you need help with a problem?', and zero otherwise. For girls (column 1), the sum of the treatment coefficients on T3/T4 and its interaction terms are statistically significant, which tell us that girls who prefer to seek help from their mother are less likely to select the father-form in T3/T4, relative to the same group in T1 (column 1, coef = -0.225, two-sided $p = 0.007$ in T3; coef = -0.143, two-sided $p = 0.075$ in T4).

The second binary variable captures which parent the child usually goes to for permission on school-related issues, where 'father' takes the value one and 'mother' takes the value 0. For girls (column 3), the treatment coefficients are negative and significant for T2, T3, and T4, suggesting a reduced preference for the father-form in these treatments relative to T1, for girls who usually approach the mother first. However, the sum of the coefficients tell us that these effects are not statistically significant for girls who generally seek permission from their father. For boys (column 4), we find that the sum of the treatment coefficients on T2/T3/T4 and their interaction terms are statistically significant, which tell us that boys who prefer to seek permission from their father first are more likely to select the father-form in these treatments, relative to the same group in T1 (column 4, coef = 0.301, two-sided $p = 0.011$ in T2; coef = 0.257, two-sided $p = 0.036$ in T3; coef = 0.227, two-sided $p = 0.047$ in T4).

The third binary variable looks at which parent the child sees as the final decision-maker in school-related decisions, which takes the value one if 'father' is chosen, zero otherwise. We only find significant treatment effects for boys in T2 (column 6) when we look at the test on the sum of coefficients (coef = 0.220, two-sided $p = 0.023$ in T2), which suggests that boys who see their father as the final decision-maker are more likely to choose the father form in T2 than T1. We do not find

Table 1.6: Parental involvement in school-related decisions

	Seek help from mother		Seek permission from father		Father final decision-maker	
	(1) Girls > 10	(2) Boys > 10	(3) Girls > 10	(4) Boys > 10	(5) Girls > 10	(6) Boys > 10
CE T2: Risk	0.132	0.346***	-0.160*	0.104	-0.127	0.096
	(0.145)	(0.106)	(0.082)	(0.078)	(0.106)	(0.095)
CE T3: Conflict	0.126	0.219**	-0.219***	0.118	-0.173	0.149*
	(0.144)	(0.105)	(0.079)	(0.077)	(0.106)	(0.088)
CE T4: Both	-0.008	0.177*	-0.162**	0.052	-0.154	0.098
	(0.139)	(0.104)	(0.077)	(0.077)	(0.104)	(0.083)
Yes	0.069	0.056	-0.001	0.083	-0.106	0.008
	(0.122)	(0.103)	(0.114)	(0.103)	(0.109)	(0.098)
CE T2 Risk × Yes	-0.259	-0.295**	0.333**	0.198	0.131	0.124
	(0.173)	(0.141)	(0.153)	(0.142)	(0.148)	(0.138)
CE T3 Conflict × Yes	-0.351**	-0.097	0.372**	0.140	0.058	-0.003
	(0.171)	(0.139)	(0.163)	(0.145)	(0.145)	(0.135)
CE T4 Both × Yes	-0.136	-0.112	0.172	0.175	0.079	0.030
	(0.166)	(0.136)	(0.158)	(0.139)	(0.144)	(0.140)
Predicted mean under T1	0.39	0.26	0.39	0.26	0.39	0.26
lincom test: T2+T2 × Yes	0.127	0.572	0.175	0.011	0.966	0.023
lincom test: T3+T3 × Yes	0.007	0.171	0.279	0.036	0.242	0.155
lincom test: T4+T4 × Yes	0.075	0.444	0.938	0.047	0.432	0.243
N	390	386	390	386	390	386

Notes: OLS regressions with the dependent variable equal to one if the child chooses the father-form, zero if they choose the mother-form. The sample only includes girls and boys above 10 years old. The CE treatments are interacted with "Yes" variables capturing children's perceptions of parental involvement in decisions. In models 1 and 2, "Yes" is a dummy variable taking the value one if the child answers 'mother' to the question "Which parent do you prefer to go to when you need help with a problem?", and zero otherwise. In models 3 and 4, "Yes" is a dummy variable taking the value one if the child answers 'father' to the question "When you need permission for school-related activities, who do you generally ask?", and zero otherwise. In models 5 and 6, "Yes" is a dummy variable taking the value one if the child answers 'father' to the question "Which parent makes the final decision in the school-related case?", and zero otherwise. The only control variables we use here are the child's age and the child's own risk attitudes. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

any significant treatment effect differences for girls who see their father as the final decision-maker (column 5).

Child agency in joint decision-making

Last, we examine whether the extent to which the child feels their parents afford them agency in decision-making can help explain treatment effects. To this end, we use three binary variables capturing different elements of child perceptions on parenting styles: whether the child feels their parents give them opportunities to make their own decisions ('Own'), whether the child feels their parents give them good reasons if they question why they need to do something ('Reason'), and whether the child feels their point of view is important to parents when they make important decisions ('Views'). A heterogeneity analysis of treatment effects along these three variables is reported in Table 1.7. From Table 1.1, we note that around ∼60% of boys and girls feel they are given the opportunity to make decisions on their own, ∼50% of boys and girls feel that their views are taken into account in child-parent decisions, and most (∼88%) feel that parents give sufficient reasons for the decisions made on their behalf.

Looking at 'Own', as there are no significant test results for the sum of any treatment coefficients and their interaction terms for girls (column 1), there are no treatment effects among girls for whom

Table 1.7: Child agency in joint decision-making

	Own		Reason		View	
	(1) Girls > 10	(2) Boys > 10	(3) Girls > 10	(4) Boys > 10	(5) Girls > 10	(6) Boys > 10
CE T2: Risk	-0.230**	0.110	-0.355	0.116	-0.065	0.175
	(0.110)	(0.110)	(0.238)	(0.197)	(0.094)	(0.109)
CE T3: Conflict	-0.223*	0.119	-0.483**	-0.278	-0.213**	0.194*
	(0.113)	(0.106)	(0.225)	(0.211)	(0.096)	(0.108)
CE T4: Both	-0.157	0.053	-0.279	-0.186	-0.258***	0.059
	(0.106)	(0.111)	(0.215)	(0.198)	(0.092)	(0.098)
Yes	-0.190*	-0.085	-0.130	-0.296**	-0.250**	-0.004
	(0.110)	(0.100)	(0.169)	(0.147)	(0.107)	(0.109)
CE T2 Risk × Yes	0.227	0.099	0.283	0.057	-0.006	0.000
	(0.146)	(0.141)	(0.248)	(0.210)	(0.139)	(0.146)
CE T3 Conflict × Yes	0.130	0.070	0.371	0.491**	0.145	-0.036
	(0.145)	(0.138)	(0.239)	(0.222)	(0.147)	(0.144)
CE T4 Both × Yes	0.020	0.107	0.167	0.345	0.292**	0.125
	(0.142)	(0.145)	(0.231)	(0.211)	(0.144)	(0.140)
Predicted mean under T1	0.41	0.26	0.41	0.25	0.39	0.25
lincom test: T2+T2 × Yes	0.978	0.015	0.322	0.015	0.491	0.054
lincom test: T3+T3 × Yes	0.296	0.033	0.133	0.003	0.516	0.085
lincom test: T4+T4 × Yes	0.128	0.063	0.121	0.021	0.738	0.051
N	390	386	390	386	390	386

Notes: OLS regressions with the dependent variable equal to one if the child chooses the father-form, zero if they choose the mother-form. The sample only includes girls and boys above 10 years old. The CE treatments are interacted with "Yes" variables capturing children's perceptions of their agency in joint parental decision-making. In models 1 and 2, "Yes" is a dummy variable taking the value one if the child agrees with the statement "My parents give me many opportunities to make my own decisions about what I am doing", and zero otherwise ('Own'). In models 3 and 4, "Yes" is a dummy variable taking the value one if the child agrees with the statement "When I ask why I have to do, or not do, something, my parents give me good reasons", and zero otherwise ('Reason'). In models 5 and 6, "Yes" is a dummy variable taking the value one if the child agrees with the statement "My point of view is very important to my parents when they make important decisions", and zero otherwise ('Views'). The only control variables we use here are the child's age and the child's own risk attitudes. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

'Own' equals one. Girls who feel they cannot make their own decisions ('Own' = 0), in contrast, are more likely to choose the father-forms in T2 and T3, relative to T1, as shown by the significant coefficients of T2 and T3. For boys, we find that boys who feel they are given the space to make their own decisions are more likely to choose the father-form in T2, T3, and T4, relative to T1, as indicated by the sum of the coefficients of the treatments and their interactions, reported in the lower panel (column 2, coef = 0.208, two-sided $p = 0.015$ for T2; coef = 0.189, two-sided $p = 0.033$ for T3; coef = 0.160, two-sided $p = 0.063$ for T4).

Looking at the 'Reason' variable, we again see a significant reduction in the probability of choosing the father-form amongst girls who feel they are not given enough reason when parents make a decision, in T3 relative to the same group of girls in T1 (column 3). However, none of the tests on the sum of any treatment coefficients and their interactions are significant. In the case of boys (column 4), we find that boys who feel they are given enough reason when parents make decisions on their behalf are more likely to choose the father forms in T2, T3, and T4, relative to T1 (column 4, coef = 0.172, two-sided $p = 0.015$ for T2; coef = 0.212, two-sided $p = 0.003$ for T3; coef = 0.159, two-sided $p = 0.021$ for T4).

Last, we look at the 'Views' variable. For girls (column 5), the only significant finding is a negative effect in T3 and T4 for girls who do not feel their point of view is important in decision-

making. Looking at the tests on the sum of treatment coefficients and their interaction terms, we do not find any significant effects, which suggests that the negative effects of conflict framing and both framings (conflict and risk) found in Table 1.3 for girls are driven by girls who do not feel their views are taken into account by parents. For girls who do feel their view is important, these negative T4 treatment effects disappear. For boys (column 6), we find that the sum of treatment coefficients and their interaction terms are all significant, indicating that boys who feel their views are given due consideration in decision-making are more likely to pick the father-form in T2, T3, and T4, relative to T1 (column 6, coef = 0.175, two-sided $p = 0.054$ for T2; coef = 0.157, two-sided $p = 0.085$ for T3; coef = 0.184, two-sided $p = 0.051$ for T4).

6.2. For RCT treatment differences

In this section, we link the results of the choice experiment and the RCT. In the absence of direct parental responses, combining the two experiments acts as an approximation for a mechanisms analysis of form-return behaviour in the RCT. We note that children's responses on parental behaviour reflect perceptions rather than observed decision-making, which limits the precision of inferences drawn from this approach. Nevertheless, the CE–RCT link provides the most direct way to examine the central question of this paper: how children's preferences and agency shape parental decisions. Since the CE captures children's initial, unconstrained preferences while the RCT reflects the ultimate parental choice (where children were asked to follow instructions, restricting their agency), combining them allows us to test whether a child's expressed preference impacts the final decision. We employ two distinct approaches to combine the experimental results: (1) we test whether receiving the preferred form interacts with the treatment effect in the RCT, and (2) we examine whether the hypothetical treatments presented in the CE influence the likelihood that forms are returned in the RCT. The results of the first approach are presented in Table 1.8, and the second in Table 1.9.

Table 1.8: Influence on form returns from receiving one's form choice

	Child receives form they chose (y/n)	
	(1) Girls > 10	(2) Boys > 10
RCT T2: Father-form	-0.158**	-0.021
	(0.072)	(0.060)
Yes	-0.099	0.003
	(0.075)	(0.075)
RCT T2: Father-form × Yes	0.208*	0.027
	(0.113)	(0.097)
Predicted mean under T1: Mother-form	0.32	0.18
N	390	386

Notes: OLS regressions with the dependent variable equal to one if the child's consent form is returned, zero otherwise. The RCT treatment dummy takes the value zero if the child received a mother-form, and one if they received a father-form. Interactions are with a dummy variable which takes the value one if the child receives the form in the RCT that corresponds to the form they chose in the CE, and zero if not. The only control variables we use here are the child's age and the child's own risk attitudes. We also include enumerator fixed effects. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Recall that from Table 1.1, we find that on average, 34% of girls and boys received the form they preferred. Looking at Table 1.8, the significant negative coefficient on 'RCT T2: Father-form' tells

us that for girls who did not receive the form they wanted, the probability of the form being returned is 16 percentage points lower in T2 than in T1. Moreover, taking the sum of the coefficients on T2 and its interaction, this negative treatment difference disappears for girls who *do* get the form they want (coef = 0.049 and not statistically significant, with two-sided $p = 0.499$). This suggests that girls have a very good sense of what form is best to choose to maximise the chance that they are signed. We do not find any significant coefficients for boys.

We also examine the effects of the CE framings on the probability of the form being returned, by interacting the RCT treatment indicator with the CE treatment indicators. We present the results in Table 1.9. Here, we detect some effects of the CE framings on girls' form return rates (column 1). Specifically, the positive and significant coefficients on 'CE T2: Risk', 'CE T3: Conflict' and 'CE T4: Both' tell us that, relative to girls in CE T1 (neutral framing), girls who receive a mother-form in RCT T1 are 14 percentage points more likely to return the mother-form when risk was made salient in CE (T2), 10 percentage points more likely to return the mother-form when parental conflict was made salient in CE (T3), and 17 percentage points more likely to return the mother-form if both parental conflict and the risks of swimming lessons were made salient (T4). These effects disappear when a father-form is given to the girls in the RCT. Specifically, there are no statistically significant differences in return rates for girls who receive the father-form under the CE T2, T3 or T4 treatments, compared to if they are under CE T1. This is confirmed by looking at the sum of the treatment coefficients and its interactions (coef = -0.115 and two-sided $p = 0.182$ for T2; coef = -0.072 and two-sided $p = 0.457$ for T3; coef = -0.114 and $p = 0.194$ for T4).

7. Discussion

In this section, we revisit our results in light of the predictions derived from our model. We also consider how the results might differ if our model assumptions did not hold, and discuss potential underlying mechanisms that might be driving treatment effects.

Revisiting RCT findings

The RCT allows us to test our prediction that the uptake of swimming lessons would be higher in T1, where maternal consent was required, than in T2, where paternal consent was required by the child. Average treatment effects presented in Table 1.2 support this prediction for girls (confirming hypothesis 1.1) but not boys (refuting hypothesis 1.2). Within the framework of our model, this result can be interpreted as a stronger preference for children's human capital development amongst mothers than fathers, where girls are concerned. This is in line with empirical studies such as Dizon-Ross and Jayachandran (2023), which find that fathers have relatively lower preferences for their daughters' human capital development, and wider evidence of mothers (fathers) favouring daughters (sons) (Godoy et al., 2006; Lundberg and Rose, 2002; Mammen, 2011; Nikiforidis et al., 2018; Raley and Bianchi, 2006; Thomas, 1994). Still, this interpretation regarding parental preferences rests on two key assumptions.

The first assumption is that spouses make decisions independently, so that treatments capture true individual preferences. In practice, however, we cannot rule out intrahousehold consultation.

Table 1.9: Influence of CE on form returns in RCT

	Forms returned	
	(1) CE & RCT Girls	(2) CE & RCT Boys
RCT T2	0.085	0.039
	(0.072)	(0.079)
CE T2 Risk	0.142*	0.069
	(0.081)	(0.094)
CE T3 Conflict	0.097*	0.113
	(0.053)	(0.080)
CE T4 Both	0.166*	0.001
	(0.090)	(0.093)
RCT T2 × CE T2 Risk	-0.202*	-0.004
	(0.101)	(0.122)
RCT T2 × CE T3 Conflict	-0.164	-0.149
	(0.120)	(0.111)
RCT T2 × CE T4 Both	-0.204*	-0.008
	(0.117)	(0.121)
Predicted mean under T1	0.31	0.18
lincom test: RCT T2+RCT T2 × CE T2	0.182	0.631
lincom test: RCT T2+RCT T2 × CE T3	0.457	0.175
lincom test: RCT T2+RCT T2 × CE T4	0.194	0.731
N	390	386

Notes: OLS regressions with a binary dependent variable equal to one if the child's consent form is returned, zero otherwise. The RCT treatment dummy takes the value zero if the child received a mother-form, and one if they received a father-form. Interactions are with the CE categorical variable which has four treatment groups, where T1 'Neutral' is the baseline treatment group. We include enumerator and school fixed effects. We include child-related control variables including the child's age, attitude to risk, competitiveness preference, whether they would like to learn to swim. As controls, we also include variables that were unbalanced in both experiments, that is: the child's scores for mother and father risk attitudes, whether the child perceives that parents argue about them, whether they think the father involves the mother in decisions, whether they think the father is the final decision-maker, and whether they are more likely to seek help from their mother instead. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The parent who receives the form may discuss the decision with their spouse, in which case observed outcomes would reflect joint rather than individual preferences.

This affects interpretation only if the likelihood or influence of consultation differs systematically between mothers and fathers. Evidence from this context suggests that such asymmetries are likely. As shown in Chapter 2, fathers are less likely than mothers to consult or defer decision-making to their spouse in child-related matters. Moreover, when consultation does occur and preferences diverge, fathers' views tend to carry greater weight (Bakhtiar et al., 2024; Carlsson et al., 2013), particularly in patriarchal settings where men typically hold greater bargaining power. Thus, in the father-form treatment (T2), consultation is not only less likely but also less consequential for outcomes, whereas in the mother-form treatment (T1), consultation with, and influence from, husbands is more likely. As a result, any observed difference between treatments is likely to understate the true effect of assigning decision-making authority to mothers. Put differently, the observed treatment difference where mothers show a stronger preference for their daughters learning to swim is likely a *conservative lower-bound estimate.*

The second assumption is that children followed instructions regarding which parent to seek consent from. As we did not monitor children after instructions were given, it is possible that children did not comply with our instructions. We discuss five motivations of non-compliance, their likelihood, and possible implications for the interpretation of our results.

First, where there is a strong patriarchal culture of the father being viewed as the ultimate decision-maker, even when instructed to give the form to their mother, children may still seek approval from the father, which would narrow the gap between treatments. However, responses from children on this subject reveal that on average, boys and girls are as likely to see the father as the final decision-maker as they are to see the mother as final decision-maker (∼50% on 'Father final decision-maker' in Table 1.1). This suggests there is no systematic default towards paternal authority that would bias the mother-form treatment.

Second, where the mother is more commonly at home than the father, non-compliance might be more common in the father-form treatment than in the mother-form treatment, as the child will seek out the parent who is more readily available to sign the form. However, once again, responses from children in Table A.1 reveal that children are 88 and 97% likely to see their fathers and mothers, respectively, at home every day, ruling out this reason as a possible case of non-compliance and a confounding source of treatment differences.

Third, children could have opted not to involve either parent because they did not want swimming lessons. While stated demand for swimming lessons is extremely high—over 95% of both boys and girls report wanting to learn (Table 1.1)—this measure should be interpreted with caution. The question ("Would you like to learn to swim?") is binary and affirmatively framed, and therefore likely captures general or latent interest rather than a fully considered willingness to participate. Moreover, children were not asked to distinguish between different training modalities at this stage, such as controlled environments versus open ocean settings, which limits the precision with which this measure reflects demand for the specific intervention offered. That said, in the study context of rural coastal Zanzibar, swimming lessons are almost universally conducted in open water due to the absence of pool infrastructure. It is therefore likely that children interpreted 'swimming lessons' within this prevailing local understanding, although this cannot be directly verified. Importantly, behavioural responses suggest that interest is not merely hypothetical. Given that girls traditionally have less exposure to swimming, any overstatement of demand would likely be stronger among girls, and we would expect lower follow-through when participation requires real action, such as sharing consent forms with parents. Instead, return rates for girls are approximately twice as high as those for boys, indicating strong revealed demand that persists even when participation entails real costs, including navigating parental approval and engaging in open-water settings. This pattern is inconsistent with children strategically avoiding parental involvement due to lack of interest.

Fourth, non-compliance may have arisen from children signing the consent form themselves. We view this as highly unlikely as the experimental script explicitly accounts for such a possibility, and warns them of the consequences of doing so ('Do NOT sign it yourself. We may call one of your parents, and there will be serious consequences if we find out that you signed the form yourself.').

The last potential source of non-compliance relates to children's agency. In principle, children could deliberately deviate from instructions by involving the parent they preferred (not the parent they were assigned). Table 1.8 provides suggestive evidence against this interpretation. In particular, when girls are assigned the form corresponding to the parent they did not prefer to approach, consent rates fall sharply. If children were routinely disregarding instructions and approaching their preferred parent anyway, outcomes should be similar regardless of whether they received the

form they preferred or not. Instead, the substantial decline in consent when girls receive their non-preferred form indicates meaningful compliance with the assigned instruction, and that this compliance carries a cost in terms of success. In other words, girls have a very good sense of which form is likely to maximise consent, and restricting their agency increases treatment differences, reducing the likelihood of obtaining consent in the father-form treatment.

A final note on the RCT results: we do not find significant treatment differences for boys. We see this as the result of return rates being too small to detect any meaningful difference between treatments for boys. This is supported by the low predicted mean in T1 (Table 1.2). One reason for low overall return rates for boys could be that boys are more likely to already know the basics of swimming than girls. Although the sample went through a vetting process to ensure that all children who participated in the study could not swim, it is plausible that a proportion of boys in the sample already knew how to swim (even though they claimed they did not), and were thus less motivated to follow through with instructions. Suggestive evidence that more boys than girls already have enough swim skills to save themselves is taken from responses on reasons given for wanting to swim: there is a statistically significantly lower proportion of boys than girls who see the main reason for learning to swim as being 'to be able to save myself if we get into trouble at sea' (chi2 = 13.229, two-sided $p = 0.000$), with no other statistically significant differences amongst boys and girls for other reasons given for wanting to learn to swim (e.g., 'because my parents want me to learn' or 'so I can join my friends'). Moreover, follow up discussions with teachers revealed that boys have historically had a lower response rate than girls for most school requests (e.g., homework), requiring more regular follow-ups and reminders. This suggests the possibility that boys are less careful with the form given, or that, compared to girls, they are less likely to put in effort to secure consent. Additionally, discussions with Panje revealed that when communicating the benefits of swimming lessons to parents, there is stronger pushback from the parents of boys, who hold the belief that their sons would use swimming lessons/swimming as an excuse to disengage from academic responsibilities at school.

Revisiting CE findings

The CE allows us to test predictions regarding child preferences. Specifically, hypotheses 1.3, 1.4, and 1.5 posited that father-forms would be chosen more often in treatments that emphasised risk (T2), conflict (T3), and both (T4), respectively, relative to the neutral treatment (T1). The pooled sample of boys and girls provided evidence that runs counter to these predictions, particularly hypothesis 1.5, as children appear less likely to prefer the father-form in T4 relative to T3 and T2 (see statistically significant F-tests in the last two rows of column 1, Table 1.3).

However, the gender-disaggregated analysis reveals that the pooled sample masks opposing tendencies amongst boys and girls. Specifically, boys' choices align with the original predictions, as they are more likely to select the father-form in T2 and T3 compared to T1, consistent with hypotheses 1.3 and 1.4. Girls, in contrast, display choices opposite to our predictions. They are less likely to choose the father-form in T3 and T4, relative to T1, thus contradicting hypotheses 1.4 and 1.5, respectively. Taken together, three patterns emerge.

First, it appears that the conflict framings induce girls and boys to have a stronger preference for involving the same-gender parent, compared to the neutral framing. Where boys and girls switch their preference more towards the same-gender parent in treatments that highlight the possibility of inter-parental conflict (i.e., treatments T3 and T4), one plausible explanation is that children prefer to go to the same-gender parent if they believe this parent will not involve the other parent, as a strategy to mitigate the potential for inter-parental conflict. This is supported by results in Table 1.5, which show that girls are less likely to choose the father-form in T4 if they think the father involves the mother, and boys are more likely to choose the father-form in T3 if they think the father does not involve the mother.

Second, for girls, the conflict framing is more effective than the risk framing. One interpretation here is that girls are more sensitive to spousal conflict than boys. This interpretation is supported with evidence from the summary statistics on conflict perceptions. On average, girls are more likely to report that their parents argue than boys are (see Table 1.1). This difference cannot be plausibly attributed to systematic differences between boys' parents and girls' parents, and is more likely attributed to a difference in sensitivity to or perceptions around parental conflict. Still, it might be that the risk framing matters for certain subgroups of girls. Indeed, the subgroup analysis presented in Table 1.4 reveals that girls who see their father as relatively more risk-averse than their mother are significantly *less* likely to choose the father-form under the risk framing.

Third, for boys, the combined framing appears to be less effective than separate framings. This might be because boys were more sensitive to information overload in the combined treatment, resulting in both effects being diluted. However, the combined treatment might also be more effective for certain subgroups, which we find to be the case. Specifically, we find that treatment effects in T4 become statistically significant for boys who report that their parents do *not* argue, making them more likely to choose the father-form in T4. Amongst boys who report that their parents do argue, this effect disappears.

8. Conclusion

In this chapter, we examine three research questions: (1) whether the gender of the parent involved in a decision about their child's human capital development matters for uptake, (2) what children's preferences are for involving one parent over the other in a decision made on their behalf, and to what extent this is influenced by the nature of the decision and child perceptions around parental characteristics and family dynamics, and (3) whether children's agency in parental decision-making can influence the final development outcome. To provide insights into these questions, we report the results of a field experiment conducted with 1,272 schoolchildren in Tanzania, in which children are offered the opportunity to enrol in swimming lessons by securing consent from one parent.

To answer the first research question, we use results from an RCT which randomised whether the child was given a consent form for swimming lessons that had to be signed by either their mother or father. We find that the gender of the parent confronted with the human capital decision *does* matter, but only for girls. Specifically, girls are 12.5 percentage points or 38% less likely to return

the father-form (RCT T2) than the mother-form (RCT T1), and this increases to 16.7 percentage points or a 46% reduction for girls above 10 years old.

To answer the second research question, we examine the results from a choice experiment run on a subsample of 848 children, where we elicited child preferences between mother-forms and father-forms using a 2 × 2 treatment design to experimentally vary the framing of swimming lessons as neutral (the control group), high-risk (i.e., since swimming lessons take place in open ocean), high-conflict (i.e., with the potential to cause inter-parental disagreement since only one parental signature is required), and a combined high-risk and high-conflict framing. Here, results reveal significant gender differences in how treatments influence children's preferences. Girls prefer to obtain consent from their mother when parental conflict is highlighted, whereas boys tend to choose fathers where risk or parental conflict are made salient, but not both. Additionally, we find treatment effects to be moderated by children's perceptions of their parents, in particular, how parents involve each other in their decisions.

Integrating the findings from both experiments helps shed light on our third research question around child agency in parental decisions. Specifically, we find that emphasising parental conflict reduces girls' success in securing paternal consent but increases their success with maternal consent. We also find that the likelihood of securing parental consent is higher where girls are given the form they want. These findings indicate that children are not passive recipients of parental decisions: they hold meaningful preferences over which parent to approach, act strategically in response to family dynamics, and their agency has real consequences for whether a human capital investment is realised.

That said, these results are likely conservative estimates of the true effects. To the extent that mothers in the mother-form treatment are more likely than fathers to have consulted their spouse—and evidence from Chapter 2 suggests such asymmetric consultation does occur—any observed treatment difference understates the true effect of assigning decision authority to mothers. Similarly, although the CE elicits stated preferences over a hypothetical choice, the consistency between girls' stated preferences and their revealed behaviour in the RCT suggests these preferences are meaningful. If anything, the true magnitude of both the parental gender effect and children's strategic agency is larger than our estimates indicate.

Taken together, our findings underscore the importance of recognising children as active participants in household decision-making. We demonstrate that parental decision authority—i.e., *who* gets to decide—is not fixed but can be endogenously shaped by the child. This dynamic matters, not least because the degree of parental involvement in child-parent decisions has direct consequences for children's human capital development.

To our knowledge, this is the first study to experimentally vary which parent is confronted with a child's human capital decision while simultaneously eliciting children's own preferences over parental involvement, providing direct evidence of child agency in intrahousehold decision-making. Such evidence suggests that interventions aimed at improving children's human capital may be more efficient when targeted directly at children rather than exclusively at parents, at least in contexts where children are the proximate actors in securing parental consent.

References

Armand, Alex, Orazio Attanasio, Pedro Carneiro, and Valérie Lechene. 2020. "The effect of gender-targeted conditional cash transfers on household expenditures: Evidence from a randomized experiment." *The Economic Journal*, 130(631): 1875–1897.

Ashraf, Nava. 2009. "Spousal control and intra-household decision making: An experimental study in the Philippines." *American Economic Review*, 99(4): 1245–1277.

Attanasio, Orazio P, Costas Meghir, and Ana Santiago. 2012. "Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa." *The Review of Economic Studies*, 79(1): 37–66.

Bakhtiar, M Mehrab, Marcel Fafchamps, Markus Goldstein, Kenneth L Leonard, and Sreelakshmi Papineni. 2024. "To Defer or to Differ: Experimental Evidence on the Role of Cash Transfers in Nigerian Couples' Decision–Making." *The Economic Journal*.

Becker, Gary S. 1981. *A treatise on the family.* Harvard university press.

Becker, Gary S, and Nigel Tomes. 1976. "Child endowments and the quantity and quality of children." *Journal of political Economy*, 84(4, Part 2): S143–S162.

Besta, Naima Abdallah. 2013. "Seaweed Farming and Intra-Household Gender Relations on Songo Songo Island, Tanzania." PhD diss. University of East Anglia, UK.

Borgonovi, Francesca, Helke Seitz, and Irina Vogel. 2022. "Swimming skills around the world: evidence on inequalities in life skills across and within countries."

Browning, Martin, Pierre-André Chiappori, and Yoram Weiss. 2014. *Economics of the Family.* Cambridge University Press.

Brugere, Cecile, Flower E. Msuya, Narriman Jiddawi, Betty Nyonje, and Ritha Maly. 2020. "Can innovation empower? Reflections on introducing tubular nets to women seaweed farmers in Zanzibar." *Gender, Technology and Development*, 24(1): 89–109.

Carlsson, Fredrik, Peter Martinsson, Ping Qin, and Matthias Sutter. 2013. "The influence of spouses on household decision making under risk: an experiment in rural China." *Experimental Economics*, 16(3): 383–401.

Chiappori, Pierre-André. 1992. "Collective labor supply and welfare." *Journal of Political Economy*, 100(3): 437–467.

Croson, Rachel, and Uri Gneezy. 2009. "Gender differences in preferences." *Journal of Economic literature*, 47(2): 448–474.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97(2): 31–47.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78(3): 883–931.

de la Torre-Castro, Maricela, Sara Fröcklin, Sanna Börjesson, Janine Okupnik, and Narriman S. Jiddawi. 2017. "Gender analysis for better coastal management – Increasing our understanding of social-ecological seascapes." *Marine Policy*, 83: 62–74.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household Choices and Child Development." *The Review of Economic Studies*, 81(1 (286)): 137–185.

Dizon-Ross, Rebecca, and Seema Jayachandran. 2023. "Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation." *American Economic Review: Insights*, 5(4): 445–459.

Doepke, Matthias, and Fabrizio Zilibotti. 2017. "Parenting with style: Altruism and paternalism in intergenerational preference transmission." *Econometrica*, 85(5): 1331–1371.

Duflo, Esther. 2003. "Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa." *The World Bank Economic Review*, 17(1): 1–25.

Duflo, Esther, and Christopher Udry. 2004. "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices." NBER Working Paper No. w10498, Available at SSRN: https://ssrn.com/abstract=552103.

D'Exelle, Ben, and Charlotte Ringdal. 2022. "Women's use of family planning services: An experiment on the husband's involvement." *Journal of Development Economics*, 158: 102915.

D'Exelle, Ben, and Charlotte Ringdal. 2025. "Who decides in the household when risk preferences conflict?" *Review of Economics of the Household*.

D'Exelle, Ben, Aurélia Lépine, Richard Bakyono, and Ludovic D. G. Tapsoba. 2023. "Fertility and polygyny: Experimental evidence from Burkina Faso." *Journal of Development Economics*, 164: 103134.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics*, 133(4): 1645–1692.

Felter, Claire Elizabeth. 2015. "In Zanzibar, Swim School to Fight Drownings."

Godoy, Ricardo, Victoria Reyes-García, Thomas McDade, Susan Tanner, William R. Leonard, Tomás Huanca, Vincent Vadez, and Karishma Patel. 2006. "Why do mothers favor girls and fathers, boys?" *Human Nature*, 17(2): 169–189.

Graham, Anne P, Mary Ann Powell, Donnah Anderson, Robyn Fitzgerald, and Nicola Taylor. 2013. "Ethical research involving children." UNICEF Office of Research Innocenti.

Haddad, Lawrence, John Hoddinott, and Harold Alderman. 1997. "Intrahousehold resource allocation in developing countries: models, methods and policies." *International Food Policy Research Institute*.

Hall, Anthony. 2008. "Brazil's Bolsa Família: A double-edged sword?" *Development and change*, 39(5): 799–822.

Heckman, James J, and Stefano Mosso. 2014. "The economics of human development and social mobility." *Annual Review of Economics*, 6(1): 689–733.

King, Nancy M. P., and Larry R. Churchill. 2000. "Ethical Principles Guiding Research on Child and Adolescent Subjects." *Journal of Interpersonal Violence*, 15(7): 710–724.

Lundberg, Shelly, and Elaina Rose. 2002. "The effects of sons and daughters on men's labor supply and wages." *Review of Economics and Statistics*, 84(2): 251–268.

Lundberg, Shelly, and Robert A Pollak. 1993. "Separate spheres bargaining and the marriage market." *Journal of political Economy*, 101(6): 988–1010.

Lundberg, Shelly, Jennifer L. Romich, and Kwok Ping Tsang. 2009. "Decision-making by children." *Review of Economics of the Household*, 7(1): 1–30.

Lundberg, Shelly J, Robert A Pollak, and Terence J Wales. 1997. "Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit." *Journal of Human Resources*, 32(3): 463–480.

Makame, Makame Omar, Ali Rashid Hamad, Muhammad Suleiman Said, Alice Mushi, and Khadija Sharif. 2021. "Moving Seaweed Farms from Shallow to Deep Seawater to Cope with Warming and Diseases in Zanzibar. Current Socio-Economic and Cultural Barriers." *Journal of Sustainable Development*, 14(5): p29. Number: 5.

Mammen, Kristin. 2011. "Fathers' time investments in children: do sons get more?" *Journal of Population Economics*, 24(3): 839–871.

McCarthy, Aine Seitz. 2019. "Intimate partner violence and family planning decisions: Experimental evidence from rural Tanzania." *World Development*, 114: 156–174.

Nikiforidis, Lambrianos, Kristina M Durante, Joseph P Redden, and Vladas Griskevicius. 2018. "Do mothers spend more on daughters while fathers spend more on sons?" *Journal of Consumer Psychology*, 28(1): 149–156.

Raley, Sara, and Suzanne Bianchi. 2006. "Sons, Daughters, and Family Processes: Does Gender of Children Matter?" *Annual Review of Sociology*, 32(Volume 32, 2006): 401–421.

Robinson, Carol. 2025. "Ethical considerations of children's involvement in school-based research: balancing children's provision, protection, and participation rights." *Research Ethics*, 21(3): 480–502.

Royal National Lifeboat Institution. 2019. "Evidence, Perceptions and Knowledge of drowning risk in Zanzibar."

Stacke, Sarah. 2017. "On 'Burkini Island' Muslim Girls Can Finally Learn to Swim." Section: Photography.

Thomas, Duncan. 1990. "Intra-Household Resource Allocation: An Inferential Approach." *The Journal of Human Resources*, 25(4): 635–664.

Thomas, Duncan. 1994. "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height." *The Journal of Human Resources*, 29(4): 950–988.

Tungodden, Jonas, and Alexander Willén. 2023. "When Parents Decide: Gender Differences in Competitiveness." *Journal of Political Economy*, 131(3): 751–801.

Appendix

A. Descriptive statistics for categorical variables

Table A.1: Descriptive statistics for categorical variables

	Boys (%)	Girls (%)	Total (%)	Pearson's χ^2
Can your father swim?				
No	31	27	29	0.052
Yes	41	38	40	
Don't know	28	35	31	
Can your mother swim?				
No	73	70	72	0.020
Yes	8	6	7	
Don't know	19	25	22	
How often do you see your father at home?				
Never	0	0	0	<0.001
A few times a year	2	1	2	
A few times a week	14	6	10	
Once a day	16	18	17	
Several times a day	68	73	71	
How often do you see your mother at home?				
Never	0	0	0	0.544
A few times a year	1	1	1	
A few times a week	2	1	2	
Once a day	4	3	3	
Several times a day	93	95	94	
What is your mother's highest level of education completed?				
No education	2	2	2	0.027
Primary	10	8	9	
Secondary	29	27	28	
Higher education	8	4	6	
Don't know	51	58	54	
Vocational training	0	0	0	
What is your father's highest level of education completed?				
No education	1	1	1	<0.001
Primary	6	6	6	
Secondary	25	15	20	
Higher education	6	6	6	
Don't know	61	71	66	

Notes: Summary statistics for categorical variables used as control variables in main regressions. Pearson's χ^2 test p-values for statistical difference between boys and girls for categorical variables are reported in the last column.

B. Balancing tests

Balancing tables for all variables in Table 1.1 are presented below. Note that this does not include the categorical variables presented in Appendix A, given that t-tests would treat these as continuous variables, which is most likely statistically invalid, and binary variables of the same categorical variables would mask important variation. However, as a precaution, we include all categorical variables presented in Table A.1 as controls in all regressions.

Table B.1: Balancing tests for RCT treatment groups, for girls

Variable	(1) RCT T1 Mean/(SE)	(2) RCT T2 Mean/(SE)	(1)-(2) Pairwise t-test Mean difference
Education	0.453	0.453	0.000
	(0.028)	(0.028)	1.000
Age	13.418	13.330	0.088
	(0.109)	(0.110)	0.571
Siblings	5.604	6.031	-0.428
	(0.183)	(0.200)	0.116
Preference for swimming lessons	0.947	0.959	-0.013
	(0.013)	(0.011)	0.455
Child's self-score of competition	7.368	7.226	0.142
	(0.134)	(0.139)	0.463
Child's self-score of risk	5.094	5.031	0.063
	(0.151)	(0.142)	0.762
Child's risk score of father	7.409	7.336	0.072
	(0.142)	(0.139)	0.716
Child's risk score of mother	5.648	5.667	-0.019
	(0.145)	(0.145)	0.927
Mother more risk-averse	0.579	0.575	0.003
	(0.028)	(0.028)	0.936
Father more cautious	0.085	0.091	-0.006
	(0.016)	(0.016)	0.780
Mother more cautious	0.509	0.516	-0.006
	(0.028)	(0.028)	0.874
Father involves mother	0.767	0.774	-0.006
	(0.024)	(0.024)	0.851
Parents argue	0.330	0.349	-0.019
	(0.026)	(0.027)	0.616
Parents argue about child	0.110	0.113	-0.003
	(0.018)	(0.018)	0.900
Seek help from mother	0.726	0.704	0.022
	(0.025)	(0.026)	0.539
Seek permission from father	0.274	0.245	0.028
	(0.025)	(0.024)	0.416
Father final decision-maker	0.535	0.503	0.031
	(0.028)	(0.028)	0.428
Own	0.610	0.563	0.047
	(0.027)	(0.028)	0.228
Reason	0.887	0.865	0.022
	(0.018)	(0.019)	0.401
Views	0.525	0.484	0.041
	(0.028)	(0.028)	0.303
N	318	318	636

Notes: Balancing table for the following variables: 'Education' which is a dummy taking the value one if they are in primary school, zero if they are not; 'Age' which captures the child's age; 'Siblings' which is the total number of siblings the child has; 'Preference for swimming lessons (y/n)' which is a dummy taking the value one if the child answers yes to the question 'Would you like to learn to swim?', zero if the answer is no; Child's self-score of competition' which is a continuous variable that asks the child to consider how willing they are to compete on a scale of 0-10, where zero means 'not willing to compete', and 10 means 'very willing to compete'; 'Child's self-score of risk' which is a continuous variable that asks the child to consider how they view their willingness to take risk on a scale of 0-10, where zero means 'not willing to take risk at all', and 10 means 'very willing to take risk'; 'Child's risk score of father/mother' which offers the same scale, but to capture how the child views their father's/mother's willingness to take risk; 'Mother more risk-averse' which is a binary variable with value one if the mother's risk score is lower than the father's risk score, zero otherwise; 'Father/Mother more cautious' which are two binary variables with value one if the child answers with 'Father/Mother' to the question 'Which parent is more cautious in life?', zero otherwise; 'Father involves mother' which is a binary variable taking the value one if the child answers yes to the question 'For any decision your father takes, does he always involve your mother?', and zero otherwise; 'Parents argue' which is a binary variable with value one if the child answers 'false' to the statement 'My parents never argue', and zero otherwise; 'Parents argue about child' which is a binary variable taking the value one if the child answers 'true' to the statement 'My parents usually argue or disagree because of things that I do', zero otherwise; 'Seek help from mother' which is a binary variable with value one if the child answers 'mother' to the question 'Which parent do you prefer to go to when you need help with a problem?', zero otherwise; 'Seek permission from father' which is a binary variable with value one if the child answers 'father' to the question 'When you need permission for school-related activities, who do you generally ask?', zero otherwise; 'Father final decision-maker' which is a binary variable with value one if the child answers 'father' to the question 'Which parent makes the final decision in the school-related case?', zero otherwise; 'Own' which is a binary variable with value one if the child agrees with the statement 'My parents give me many opportunities to make my own decisions about what I am doing', zero if they do not; 'Reason' which is a binary variable with value one if the child agrees with the statement 'When I ask why I have to do, or not do, something, my parents give me good reasons', zero if they do not; and 'Views' which is a binary variable with value one if the child agrees with the statement 'My point of view is very important to my parents when they make important decisions', zero if they do not. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Balancing tests for RCT treatment groups, for boys

Variable	(1) RCT T1 Mean/(SE)	(2) RCT T2 Mean/(SE)	(1)-(2) Pairwise t-test Mean difference
Education	0.450	0.450	0.000
	(0.028)	(0.028)	1.000
Age	13.629	13.538	0.091
	(0.121)	(0.116)	0.586
Siblings	5.513	5.352	0.160
	(0.175)	(0.182)	0.524
Preference for swimming lessons	0.984	0.965	0.019
	(0.007)	(0.010)	0.129
Child's self-score of competition	7.371	7.208	0.164
	(0.132)	(0.144)	0.404
Child's self-score of risk	6.506	6.204	0.302
	(0.144)	(0.158)	0.159
Child's risk score of father	6.075	6.123	-0.047
	(0.137)	(0.141)	0.810
Child's risk score of mother	5.126	5.035	0.091
	(0.150)	(0.151)	0.669
Mother more risk-averse	0.522	0.503	0.019
	(0.028)	(0.028)	0.635
Father more cautious	0.148	0.167	-0.019
	(0.020)	(0.021)	0.514
Mother more cautious	0.575	0.538	0.038
	(0.028)	(0.028)	0.339
Father involves mother	0.777	0.739	0.038
	(0.023)	(0.025)	0.267
Parents argue	0.547	0.547	0.000
	(0.028)	(0.028)	1.000
Parents argue about child	0.160	0.230	-0.069**
	(0.021)	(0.024)	0.028
Seek help from mother	0.610	0.544	0.066*
	(0.027)	(0.028)	0.092
Seek permission from father	0.343	0.286	0.057
	(0.027)	(0.025)	0.125
Father final decision-maker	0.428	0.443	-0.016
	(0.028)	(0.028)	0.690
Own	0.604	0.607	-0.003
	(0.027)	(0.027)	0.935
Reason	0.881	0.887	-0.006
	(0.018)	(0.018)	0.805
Views	0.575	0.544	0.031
	(0.028)	(0.028)	0.425
N	318	318	636

Notes: Balancing table for the following variables: 'Education' which is a dummy taking the value one if they are in primary school, zero if they are not; 'Age' which captures the child's age; 'Siblings' which is the total number of siblings the child has; 'Preference for swimming lessons (y/n)' which is a dummy taking the value one if the child answers yes to the question 'Would you like to learn to swim?', zero if the answer is no; Child's self-score of competition' which is a continuous variable that asks the child to consider how willing they are to compete on a scale of 0-10, where zero means 'not willing to compete', and 10 means 'very willing to compete'; 'Child's self-score of risk' which is a continuous variable that asks the child to consider how they view their willingness to take risk on a scale of 0-10, where zero means 'not willing to take risk at all', and 10 means 'very willing to take risk'; 'Child's risk score of father/mother' which offers the same scale, but to capture how the child views their father's/mother's willingness to take risk; 'Mother more risk-averse' which is a binary variable with value one if the mother's risk score is lower than the father's risk score, zero otherwise; 'Father/Mother more cautious' which are two binary variables with value one if the child answers with 'Father/Mother' to the question 'Which parent is more cautious in life?', zero otherwise; 'Father involves mother' which is a binary variable taking the value one if the child answers yes to the question 'For any decision your father takes, does he always involve your mother?', and zero otherwise; 'Parents argue' which is a binary variable with value one if the child answers 'false' to the statement 'My parents never argue', and zero otherwise; 'Parents argue about child' which is a binary variable taking the value one if the child answers 'true' to the statement 'My parents usually argue or disagree because of things that I do', zero otherwise; 'Seek help from mother' which is a binary variable with value one if the child answers 'mother' to the question 'Which parent do you prefer to go to when you need help with a problem?', zero otherwise; 'Seek permission from father' which is a binary variable with value one if the child answers 'father' to the question 'When you need permission for school-related activities, who do you generally ask?', zero otherwise; 'Father final decision-maker' which is a binary variable with value one if the child answers 'father' to the question 'Which parent makes the final decision in the school-related case?', zero otherwise; 'Own' which is a binary variable with value one if the child agrees with the statement 'My parents give me many opportunities to make my own decisions about what I am doing', zero if they do not; 'Reason' which is a binary variable with value one if the child agrees with the statement 'When I ask why I have to do, or not do, something, my parents give me good reasons', zero if they do not; and 'Views' which is a binary variable with value one if the child agrees with the statement 'My point of view is very important to my parents when they make important decisions', zero if they do not. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Balancing tests for CE treatment groups, for girls

Variable	(1) CE T1: Neutral Mean/(SE)	(2) CE T2: Risk Mean/(SE)	(3) CE T3: Conflict Mean/(SE)	(4) CE T4: Both Mean/(SE)	F-test for balance across all groups F-stat/P-value	(1)-(2) Mean difference	(1)-(3) Mean difference	(1)-(4) Mean difference	(2)-(3) Mean difference	(2)-(4) Mean difference	(3)-(4) Mean difference
						Pairwise t-test					
Education	0.453	0.453	0.453	0.453	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.049)	(0.049)	(0.049)	(0.049)	1.000						
Age	13.160	13.368	13.302	13.443	0.391	-0.208	-0.142	-0.283	0.066	-0.075	-0.142
	(0.188)	(0.201)	(0.193)	(0.187)	0.760						
Siblings	6.085	5.962	5.849	5.104	1.759	0.123	0.236	0.981**	0.113	0.858*	0.745
	(0.332)	(0.338)	(0.377)	(0.276)	0.154						
Preference for swimming lessons	0.943	0.962	0.943	0.953	0.182	-0.019	0.000	-0.009	0.019	0.009	-0.009
	(0.023)	(0.019)	(0.023)	(0.021)	0.908						
Child's self-score of competition	7.094	7.349	7.292	7.226	0.209	-0.255	-0.198	-0.132	0.057	0.123	0.066
	(0.242)	(0.233)	(0.253)	(0.230)	0.890						
Child's self-score of risk	5.170	5.028	5.217	5.047	0.130	0.142	-0.047	0.123	-0.189	-0.019	0.170
	(0.259)	(0.268)	(0.233)	(0.261)	0.942						
Child's risk score of father	7.660	7.443	7.491	6.925	1.707	0.217	0.170	0.736**	-0.047	0.519	0.566*
	(0.240)	(0.250)	(0.212)	(0.266)	0.165						
Child's risk score of mother	5.764	5.915	5.887	5.434	0.755	-0.151	-0.123	0.330	0.028	0.481	0.453
	(0.268)	(0.264)	(0.232)	(0.250)	0.520						
Mother more risk-averse	0.557	0.575	0.566	0.575	0.035	-0.019	-0.009	-0.019	0.009	0.000	-0.009
	(0.048)	(0.048)	(0.048)	(0.048)	0.991						
Father more cautious	0.047	0.123	0.075	0.047	1.993	-0.075**	-0.028	0.000	0.047	0.075**	0.028
	(0.021)	(0.032)	(0.026)	(0.021)	0.114						
Mother more cautious	0.509	0.500	0.472	0.557	0.525	0.009	0.038	-0.047	0.028	-0.057	-0.085
	(0.049)	(0.049)	(0.049)	(0.048)	0.665						
Father involves mother	0.821	0.792	0.698	0.821	2.122*	0.028	0.123**	0.000	0.094	-0.028	-0.123**
	(0.037)	(0.040)	(0.045)	(0.037)	0.097						
Parents argue	0.321	0.396	0.302	0.443	1.999	-0.075	0.019	-0.123*	0.094	-0.047	-0.142**
	(0.046)	(0.048)	(0.045)	(0.048)	0.113						
Parents argue about child	0.094	0.170	0.085	0.113	1.502	-0.075	0.009	-0.019	0.085*	0.057	-0.028
	(0.029)	(0.037)	(0.027)	(0.031)	0.214						
Seek help from mother	0.726	0.726	0.774	0.698	0.525	0.000	-0.047	0.028	-0.047	0.028	0.075
	(0.044)	(0.044)	(0.041)	(0.045)	0.665						
Seek permission from father	0.264	0.311	0.217	0.255	0.823	-0.047	0.047	0.009	0.094	0.057	-0.038
	(0.043)	(0.045)	(0.040)	(0.043)	0.482						
Father final decision-maker	0.594	0.509	0.491	0.519	0.881	0.085	0.104	0.075	0.019	-0.009	-0.028
	(0.048)	(0.049)	(0.049)	(0.049)	0.451						
Own	0.604	0.585	0.632	0.538	0.686	0.019	-0.028	0.066	-0.047	0.047	0.094
	(0.048)	(0.048)	(0.047)	(0.049)	0.561						
Reason	0.858	0.896	0.868	0.896	0.376	-0.038	-0.009	-0.038	0.028	0.000	-0.028
	(0.034)	(0.030)	(0.033)	(0.030)	0.770						
Views	0.481	0.538	0.481	0.472	0.384	-0.057	0.000	0.009	0.057	0.066	0.009
	(0.049)	(0.049)	(0.049)	(0.049)	0.764						
Number of observations	106	106	106	106	424	212	212	212	212	212	212

Notes: 'Education' which is a dummy taking the value one if they are in primary school, zero if they are not; 'Age' which captures the child's age; 'Siblings' which is the total number of siblings the child has; 'Preference for swimming lessons (y/n)' which is a dummy taking the value one if the child answers yes to the question 'Would you like to learn to swim?', zero if the answer is no; Child's self-score of competition' which is a continuous variable that asks the child to consider how willing they are to compete on a scale of 0-10, where zero means 'not willing to compete', and 10 means 'very willing to compete'; 'Child's self-score of risk' which is a continuous variable that asks the child to consider how they view their willingness to take risk on a scale of 0-10, where zero means 'not willing to take risk at all', and 10 means 'very willing to take risk'; 'Child's risk score of father/mother' which offers the same scale, but to capture how the child views their father's/mother's willingness to take risk; 'Mother more risk-averse' which is a binary variable with value one if the mother's risk score is lower than the father's risk score, zero otherwise; 'Father/Mother more cautious' which are two binary variables with value one if the child answers with 'Father/Mother' to the question 'Which parent is more cautious in life?', zero otherwise; 'Father involves mother' which is a binary variable taking the value one if the child answers yes to the question 'For any decision your father takes, does he always involve your mother?', and zero otherwise; 'Parents argue' which is a binary variable with value one if the child answers 'false' to the statement 'My parents never argue', and zero otherwise; 'Parents argue about child' which is a binary variable taking the value one if the child answers 'true' to the statement 'My parents usually argue or disagree because of things that I do', zero otherwise; 'Seek help from mother' which is a binary variable with value one if the child answers 'mother' to the question 'Which parent do you prefer to go to when you need help with a problem?', zero otherwise; 'Seek permission from father' which is a binary variable with value one if the child answers 'father' to the question 'When you need permission for school-related activities, who do you generally ask?', zero otherwise; 'Father final decision-maker' which is a binary variable with value one if the child answers 'father' to the question 'Which parent makes the final decision in the school-related case?', zero otherwise; 'Own' which is a binary variable with value one if the child agrees with the statement 'My parents give me many opportunities to make my own decisions about what I am doing', zero if they do not; 'Reason' which is a binary variable with value one if the child agrees with the statement 'When I ask why I have to do, or not do, something, my parents give me good reasons', zero if they do not; and 'Views' which is a binary variable with value one if the child agrees with the statement 'My point of view is very important to my parents when they make important decisions', zero if they do not. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Balancing tests for CE treatment groups, for boys

Variable	(1) CE T1: Neutral Mean/(SE)	(2) CE T2: Risk Mean/(SE)	(3) CE T3: Conflict Mean/(SE)	(4) CE T4: Both Mean/(SE)	F-test for balance across all groups F-stat/P-value	(1)-(2) Mean difference	(1)-(3) Mean difference	(1)-(4) Mean difference	(2)-(3) Mean difference	(2)-(4) Mean difference	(3)-(4) Mean difference
						Pairwise t-test					
Education	0.443	0.453	0.453	0.453	0.009	-0.009	-0.009	-0.009	0.000	0.000	0.000
	(0.048)	(0.049)	(0.049)	(0.049)	0.999						
Age	13.642	13.453	13.434	13.528	0.211	0.189	0.208	0.113	0.019	-0.075	-0.094
	(0.204)	(0.201)	(0.209)	(0.206)	0.889						
Siblings	5.528	5.264	5.264	5.151	0.287	0.264	0.264	0.377	0.000	0.113	0.113
	(0.299)	(0.311)	(0.313)	(0.271)	0.835						
Preference for swimming lessons	0.981	0.972	0.972	0.962	0.227	0.009	0.009	0.019	0.000	0.009	0.009
	(0.013)	(0.016)	(0.016)	(0.019)	0.878						
Child's self-score of competition	7.396	7.358	6.943	6.915	1.151	0.038	0.453	0.481	0.415	0.443	0.028
	(0.233)	(0.230)	(0.254)	(0.250)	0.328						
Child's self-score of risk	6.585	6.179	6.189	6.075	0.729	0.406	0.396	0.509	-0.009	0.104	0.113
	(0.252)	(0.255)	(0.271)	(0.273)	0.535						
Child's risk score of father	6.283	6.368	5.377	5.849	3.719**	-0.085	0.906**	0.434	0.991***	0.519	-0.472
	(0.238)	(0.237)	(0.259)	(0.208)	0.012						
Child's risk score of mother	5.434	5.160	4.594	4.585	2.804**	0.274	0.840**	0.849**	0.566	0.575	0.009
	(0.256)	(0.259)	(0.246)	(0.250)	0.039						
Mother more risk-averse	0.472	0.557	0.481	0.585	1.324	-0.085	-0.009	-0.113*	0.075	-0.028	-0.104
	(0.049)	(0.048)	(0.049)	(0.048)	0.266						
Father more cautious	0.151	0.104	0.189	0.160	1.028	0.047	-0.038	-0.009	-0.085*	-0.057	0.028
	(0.035)	(0.030)	(0.038)	(0.036)	0.380						
Mother more cautious	0.613	0.642	0.519	0.557	1.329	-0.028	0.094	0.057	0.123*	0.085	-0.038
	(0.048)	(0.047)	(0.049)	(0.048)	0.265						
Father involves mother	0.821	0.717	0.774	0.774	1.081	0.104*	0.047	0.047	-0.057	-0.057	0.000
	(0.037)	(0.044)	(0.041)	(0.041)	0.357						
Parents argue	0.547	0.585	0.538	0.519	0.328	-0.038	0.009	0.028	0.047	0.066	0.019
	(0.049)	(0.048)	(0.049)	(0.049)	0.805						
Parents argue about child	0.226	0.132	0.198	0.198	1.108	0.094*	0.028	0.028	-0.066	-0.066	0.000
	(0.041)	(0.033)	(0.039)	(0.039)	0.346						
Seek help from mother	0.585	0.585	0.594	0.594	0.013	0.000	-0.009	-0.009	-0.009	-0.009	0.000
	(0.048)	(0.048)	(0.048)	(0.048)	0.998						
Seek permission from father	0.292	0.321	0.292	0.302	0.089	-0.028	0.000	-0.009	0.028	0.019	-0.009
	(0.044)	(0.046)	(0.044)	(0.045)	0.966						
Father final decision-maker	0.425	0.491	0.472	0.302	3.170**	-0.066	-0.047	0.123*	0.019	0.189***	0.170**
	(0.048)	(0.049)	(0.049)	(0.045)	0.024						
Own	0.613	0.575	0.575	0.585	0.138	0.038	0.038	0.028	0.000	-0.009	-0.009
	(0.048)	(0.048)	(0.048)	(0.048)	0.937						
Reason	0.887	0.887	0.906	0.877	0.150	0.000	-0.019	0.009	-0.019	0.009	0.028
	(0.031)	(0.031)	(0.029)	(0.032)	0.929						
Views	0.594	0.623	0.604	0.481	1.782	-0.028	-0.009	0.113*	0.019	0.142**	0.123*
	(0.048)	(0.047)	(0.048)	(0.049)	0.150						
Number of observations	106	106	106	106	424	212	212	212	212	212	212

Notes: Balancing table for the following variables: 'Education' which is a dummy taking the value one if they are in primary school, zero if they are not; 'Age' which captures the child's age; 'Siblings' which is the total number of siblings the child has; 'Preference for swimming lessons (y/n)' which is a dummy taking the value one if the child answers yes to the question 'Would you like to learn to swim?', zero if the answer is no; Child's self-score of competition' which is a continuous variable that asks the child to consider how willing they are to compete on a scale of 0-10, where zero means 'not willing to compete', and 10 means 'very willing to compete'; 'Child's self-score of risk' which is a continuous variable that asks the child to consider how they view their willingness to take risk on a scale of 0-10, where zero means 'not willing to take risk at all', and 10 means 'very willing to take risk'; 'Child's risk score of father/mother' which offers the same scale, but to capture how the child views their father's/mother's willingness to take risk; 'Mother more risk-averse' which is a binary variable with value one if the mother's risk score is lower than the father's risk score, zero otherwise; 'Father/Mother more cautious' which are two binary variables with value one if the child answers with 'Father/Mother' to the question 'Which parent is more cautious in life?', zero otherwise; 'Father involves mother' which is a binary variable taking the value one if the child answers yes to the question 'For any decision your father takes, does he always involve your mother?', and zero otherwise; 'Parents argue' which is a binary variable with value one if the child answers 'false' to the statement 'My parents never argue', and zero otherwise; 'Parents argue about child' which is a binary variable taking the value one if the child answers 'true' to the statement 'My parents usually argue or disagree because of things that I do', zero otherwise; 'Seek help from mother' which is a binary variable with value one if the child answers 'mother' to the question 'Which parent do you prefer to go to when you need help with a problem?', zero otherwise; 'Seek permission from father' which is a binary variable with value one if the child answers 'father' to the question 'When you need permission for school-related activities, who do you generally ask?', zero otherwise; 'Father final decision-maker' which is a binary variable with value one if the child answers 'father' to the question 'Which parent makes the final decision in the school-related case?', zero otherwise; 'Own' which is a binary variable with value one if the child agrees with the statement 'My parents give me many opportunities to make my own decisions about what I am doing', zero if they do not; 'Reason' which is a binary variable with value one if the child agrees with the statement 'When I ask why I have to do, or not do, something, my parents give me good reasons', zero if they do not; and 'Views' which is a binary variable with value one if the child agrees with the statement 'My point of view is very important to my parents when they make important decisions', zero if they do not. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C. Experimental scripts

The following text is the experimental script used for this study. Instructions for the enumerator are between brackets. Text in bold specifies the procedures followed.

[Introduce yourself using the following script.]

Hello! My name is [insert name]. I am part of a research team that is working with a University. We are interested in learning more about Zanzibari students. We would like to ask you some questions, if you are happy to participate in this survey, we may begin.

Once the child provides assent, and passes the eligibility check (i.e., that they are between ages 8-18 years, live with both parents, and do not know how to swim), the enumerator conducts the interview. After the interview, the enumerator reads out the following.

We are offering you the opportunity to participate in free swimming lessons. These lessons are offered by the Panje Project, a local organisation that provides swimming lessons to children in Zanzibar. You may or may not be selected to participate. If you are selected, you will be expected to attend lessons and spend 3-6 hours each week for 2 weeks. To participate in the swimming lessons, you will need permission from one of your parents.

After this statement is read out, the sample splits into two: those who go through the CE script first and then the RCT script (848 children in the 'CE&RCT' subsample.), and those who jump straight to the RCT script, skipping the CE script (424 children in the 'RCT-only' subsample).

Choice experiment script

[You must read the following text aloud verbatim to the student. Be very careful when you read the text because there are FOUR VERSIONS. The version that is displayed will be chosen at RANDOM. It is essential that you read out the version that is displayed.]

Neutral treatment (T1):

Before we proceed, here's some information about these swimming lessons. Your swimming trainer will be someone of your same gender, and the lessons will take place in the open ocean. While the ocean is different from playing on school grounds because it involves water, it's just another environment where you can have fun and learn as long as you pay attention to simple safety rules.

As you are still a minor, we require the consent of one of your parents. We understand that it is sometimes difficult for children to get signatures from both parents, and so we are happy to accept just one signature.

Risk treatment (T2):

Before we proceed, here's some information about these swimming lessons. Your swimming trainer will be someone of your same gender, and the lessons will take place in the open ocean. The ocean is not like a swimming pool, and can be a dangerous place with strong currents, sudden waves, and deep water. It's crucial to follow all safety instructions from your swimming trainer. Not doing so could lead to serious consequences.

As you are still a minor, we require the consent of one of your parents. We understand that it is sometimes difficult for children to get signatures from both parents, and so we are happy to accept just one signature.

Conflict treatment (T3):

Before we proceed, here's some information about these swimming lessons. Your swimming trainer will be someone of your same gender, and the lessons will take place in the open ocean. While the ocean is different from playing on school grounds because it involves water, it's just another environment where you can have fun and learn as long as you pay attention to simple safety rules.

As you are still a minor, we require the consent of one of your parents. We understand that this might create the possibility for disagreements between your parents since we do not ask for both signatures, but we are happy to accept just one signature.

Combined treatment (T4):

Before we proceed, here's some information about these swimming lessons. Your swimming trainer will be someone of your same gender, and the lessons will take place in the open ocean. The ocean is not like a swimming pool, and can be a dangerous place with strong currents, sudden waves, and deep water. It's crucial to follow all safety instructions from your swimming trainer. Not doing so could lead to serious consequences.

As you are still a minor, we require the consent of one of your parents. We understand that this might create the possibility for disagreements between your parents since we do not ask for both signatures, but we are happy to accept just one signature.

[You must now show the child the two different consent forms and read the following script.]

If you were given the choice, what would you choose from these two different forms in front of you? One is a consent form that needs to be signed by your mother. One is a consent form that needs to be signed by your father. You may only choose one form. Please remember, there is no correct answer, you are free to choose whichever form you wish.

For the 'CE&RCT' subsample, once the child's choice is elicited, a set of buffer questions are asked prior to moving to the RCT script.

RCT script

[You must read the following text aloud verbatim to the student. Be very careful when you read the text because there are FOUR VERSIONS. The version that is displayed will be chosen at RANDOM. It is essential that you read out the version that is displayed.]

Mother-form treatment (T1):

Thank you. We will now randomly give you one of the two consent forms. This might be different from the one you prefer.

[You must hand over the Mother-Only consent form to the student and read the following instructions aloud verbatim]

This form is for your MOTHER. If you want swimming lessons, you must take it home and give it directly to HER. SHE needs to sign it. You do not need the signature of your father. You must return the signed form to your teacher as soon as possible, ideally within one week. Do NOT sign it yourself. We may call one of your parents, and there will be serious consequences if we find out that you signed the form yourself.

Father-form treatment (T2):

Thank you. We will now randomly give you one of the two consent forms. This might be different from the one you prefer.

[You must hand over the Father-Only consent form to the student and read the following instructions aloud verbatim]

This form is for your FATHER. If you want swimming lessons, you must take it home and give it directly to HIM. HE needs to sign it. You do not need the signature of your mother. You must return the signed form to your teacher as soon as possible, ideally within one week. Do NOT sign it yourself. We may call one of your parents, and there will be serious consequences if we find out that you signed the form yourself.

[You must now hand over the form that you told the child you were going to give. You must ensure that you write down the STUDENT ID, NAME and GENDER on the form.]

Thank you for your time today. As a thank you, please accept this small gift. You may now return to your class. Please do NOT talk to other students about the survey.

END

D. Robustness checks

Here, we present the average treatment effects of the CE with an alternative specification that uses logit models.

Table D.1: CE treatment effects on form choices, by gender

	Form choice				
	(1) Pooled	(2) Girls	(3) Boys	(4) Girls > 10	(5) Boys > 10
Which form will you choose?					
CE T2: Risk	0.162	-0.803	1.588***	-0.684	2.242***
	(0.195)	(0.543)	(0.457)	(0.683)	(0.706)
CE T3: Conflict	0.041	-1.683***	1.526***	-1.442**	1.787**
	(0.135)	(0.617)	(0.555)	(0.690)	(0.715)
CE T4: Both	-0.306*	-1.442**	0.596	-1.178*	1.560*
	(0.180)	(0.655)	(0.584)	(0.690)	(0.852)
Controls		✓	✓	✓	✓
Predicted mean under T1	0.35	0.33	0.13	0.28	0.09
F-test: T2=T3	0.518	0.148	0.890	0.188	0.397
F-test: T2=T4	0.010	0.278	0.043	0.368	0.187
F-test: T3=T4	0.044	0.678	0.050	0.663	0.686
N	848	396	403	361	345

Notes: Logit regressions with the dependent variable equal to one if the child chooses the father-form, zero if they choose the mother-form. All regressions include school and enumerator fixed effects, and a set of control variables: the child's age, their attitude to risk, their preference for competition, their preference for swimming lessons, both parents' education and occupations, how often they see both parents at home, their perceptions of both parents' abilities to swim, their perceptions of both parents' attitudes to risk preferences, their perceptions on whether their parents argue or not, and their perceptions of relative parenting styles, including which parent they think is more likely to scold, punish, help, worry, be more cautious, or be more likely to grant permission for swimming lessons. We also include the binary variable capturing child views on who the final decision-maker is, as this was unbalanced for boys in the CE. Standard errors clustered at the school level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 2

Does Autonomy Matter? A Deep Dive into Spousal Involvement in Household Decisions

Abstract

Ensuring uptake of development services remains a central challenge for policymakers and practitioners working to reduce poverty and gender inequality. Most interventions address this by targeting one household member for the uptake decision—either the husband or the wife. Yet little is known about how spouses involve each other in the uptake decision *beyond* this initial contact. Three questions vis-à-vis unobserved spousal involvement are worth exploring: (1) how much autonomy do spouses retain and relinquish in household decisions? (2) what influences whether individuals retain or relinquish autonomy? and (3) does retaining or relinquishing autonomy matter for the final uptake of the services offered? To shed light on these questions, we use an experiment with 664 couples in rural Zanzibar, where husbands and wives are separately asked to consider three decision blocks. In the first block, they are offered a choice between swimming lessons and a consumption good. In the second and third blocks, they are offered the chance to defer their decision to their spouse, and to consult their spouse, respectively. We use a 2×2 treatment design that varies (i) whether swimming lessons are for the child or the wife and hence whether the decision-making domain is joint or the wife's, and (ii) whether the paired consumption good benefits an individual or the household. By experimentally varying which spouse is approached first, who benefits from the decision, and whose domain the decision falls into, we identify their separate influences on the preference for autonomy and spousal involvement. Additionally, by comparing the uptake of swimming lessons across the decision blocks, we examine whether retaining autonomy matters for final uptake. We find four main results: First, uptake is highest for treatments that offer swimming lessons for the child instead of the wife. Second, when given the chance to defer decisions, husbands (wives) are more likely to retain (relinquish) agency when the paired goods fall in the joint decision domain. Third, when consultation takes place, wives (husbands) are more (less) likely to resign to their spouse's preference. Finally, uptake comparisons across decision blocks reveal that once spousal involvement is endogenised, spousal differences in decision outcomes become smaller, reducing the importance of which spouse is initially targeted.

1. Introduction

Policymakers and development practitioners often aim to maximise the uptake of development services offered to households. To do so, they commonly target one specific household member, which, in households with married or co-habiting couples, is one of two spouses. With this approach, there is growing evidence across a range of contexts that the success of a development intervention often hinges on (a) who is approached first (either the wife or the husband); and (b) who is involved in the decision (either the individual or both spouses jointly). For example, commitment savings products yield higher uptake when offered to women rather than men (Ashraf, Karlan and Yin, 2010), transfers directed to women improve child schooling more than transfers directed to men (Schultz, 2004), and the uptake of family planning services depends on whether women are contacted alone or with their husband present (Ashraf, Field and Lee, 2014; D'Exelle and Ringdal, 2022; D'Exelle et al., 2023).

However, these studies focus primarily on a static snapshot of spousal involvement. In reality, the decision to involve one's spouse is endogenous—a dynamic, deliberative process shaped by preferences, beliefs, and anticipated (dis)agreement. Consequently, the observed decision outcome depends not only on which spouse is targeted as 'first mover', but also on how spouses subsequently retain or relinquish decision-making autonomy after this initial offer is presented. Taking this process-driven perspective, very little evidence exists to answer the following questions: (1) how much autonomy do spouses retain and relinquish in household decisions? (2) what influences whether individuals retain or relinquish autonomy? and (3) does retaining or relinquishing autonomy matter for the final uptake of the services offered?

This study is designed to provide insight into these questions. Using an experiment with 664 married or co-habiting couples (1,328 participants) in rural Zanzibar, Tanzania, we track spousal decision outcomes from initial reflection to final resolution using what we call 'Decision Blocks' to *endogenise* the spousal involvement decision process. In Decision Block 1 ('Individual block'), we separately offer husbands and wives a choice between a pair of goods: free swimming lessons or a consumption good. In Decision Block 2 ('Defer block'), individuals are offered the choice to defer decision-making to their spouse and relinquish decision-making autonomy, or retain autonomy and not involve the spouse. In Decision Block 3 ('Consult block'), individuals are offered the choice to engage in a consultation with their spouse or not. If consultation occurs, spouses are offered the choice to resign their choice to the spouse's choice, or retain their original preference.

For the dual purpose of creating exogenous variation in spousal preference differences, and variation in the key determinants of decision-making agency, we use a 2×2 treatment design which varies (i) whether swimming lessons are for the wife or the child and hence whether the decision falls in the decision-making domain of the wife or a joint domain that explicitly requires the involvement of the husband, and (ii) whether the paired consumption good benefits an individual or the household.

Endogenising spousal involvement whilst exogenously varying the determinants of decision-making in this way allows us to examine each of our three research questions. First, to examine how much autonomy spouses have in the decision-making process, we can observe the propensities of husbands and wives to involve their spouse in Decision Blocks 2 and 3, and their propensity to retain or resign their original choices once consultation takes place in Decision Block 3. Second, by experimentally varying who is approached first, who benefits, and whose domain into which the decision falls, we identify the factors that influence preferences for autonomy and spousal involvement. Last, to examine whether autonomy matters, we map the evolution of autonomy onto household decision outcomes, i.e., the uptake of swimming lessons. To do so, we first compare individual preferences in Decision Block 1, testing whether it matters which spouse is initially asked to make the decision on behalf of the household. We then compare whether preference gaps widen or diminish after spousal involvement is endogenised in Decision Blocks 2 and 3, and whether this differs across treatment.

Our focus on the uptake of swimming lessons in coastal communities in Zanzibar—an archipelago off the eastern coast of Tanzania (see panel **(a)** in Figure 1.3 of Chapter 1)—is motivated by several factors. First, for Zanzibari coastal communities, swimming is a critical capability which expands individual agency, broadens livelihood alternatives, and reduces vulnerability to drowning (Royal National Lifeboat Institution, 2019). Second, the economic relevance of swimming is increasing for Zanzibari women. Traditional coastal livelihoods such as seaweed farming, practised predominantly by women, are under threat from climatic changes (de Jong Cleyndert et al., 2021; Makame et al., 2021). This is pushing women toward alternative marine-based activities (e.g., tourism, fishing, deep-water seaweed farming) that require swimming ability (Brugere et al., 2020), which suggests rising returns to aquatic skills in this setting. Third, swimming as a human capital skill displays a striking gender gap: globally, women are less likely than men to be able to swim without assistance across nearly all countries, age cohorts, and education levels, especially in low-income settings (Borgonovi, Seitz and Vogel, 2022). This gap is particularly striking in Zanzibar's predominantly Islamic society. Historical norms around modesty, gender segregation, and female mobility have long shaped views on the appropriateness of swimming for females (Brugere et al., 2020; de la Torre-Castro et al., 2017; Makame et al., 2021; Stacke, 2017). In practice, these norms have played a role in restricting women's and girls' access to learning, which typically relied on unsupervised exposure to the sea (Brugere et al., 2020; Felter, 2015). Such constraints have relaxed over time, particularly following the establishment of the Panje Project—a local non-profit delivering swimming instruction in gender-sensitive ways such as through female instructors and the provision of modest swimwear known as 'burkinis' (Brugere et al., 2020; Stacke, 2017). Yet attitudes remain uneven. Whilst these interventions have increased acceptance of swimming as a survival skill in more accessible areas, in rural settings, residual concerns around propriety and safety continue to limit female participation (Brugere et al., 2020; Felter, 2015; Stacke, 2017). Beyond attitudes toward swimming itself, decision-making in Zanzibar reflects a broader patriarchal structure in which wives are commonly expected to seek their husband's approval before pursuing activities outside the household's routine—whether taking up paid work, visiting relatives, or enrolling in training programmes (Brugere et al., 2020). This makes the *process* of deciding on swimming lessons a particularly salient object of study: who holds decision authority over human capital investments, and whether wives can act on their own preferences without spousal involvement, are open empirical questions in this context.

Our study provides several insights. The first set of insights relates to Decision Block 1. We find that the uptake of swimming lessons is higher in treatments where lessons are offered to children instead of wives, regardless of the consumption good with which it is paired. When the child's swimming lessons are held constant, uptake is higher in the treatment where the consumption good benefits the household instead of the child. These findings hold irrespective of whether the husband or the wife makes the decision. Moreover, although cross-domain treatments successfully vary spousal preference differences, both spouses *underestimate* these preference differences, especially when swimming lessons concern the child instead of the wife.

The second set of insights is derived from Decision Blocks 2 and 3. In Decision Block 2, husbands are less likely to defer decision-making to their wives when children's swimming lessons are offered instead of wives' swimming lessons. Moreover, they reduce deferral when the child's swimming lessons are paired with a good that benefits the household rather than the child. On the contrary, women are more likely to defer when the decision involves children's swimming lessons, and this is higher when paired with a consumption good that benefits the household. In Decision Block 3, treatments influence husbands' consultation behaviour when it concerns the wife's swimming lessons. Specifically, in the wife-only domain, husbands are more likely to consult their wife when the paired good benefits the household instead of the wife directly. Treatments also influence wives' propensity to consult, but only when the child's swimming lessons are offered instead of wives'.

The third set of insights examines whether retaining or relinquishing autonomy matters for the final uptake of swimming lessons. When spouses retain decision-making control, gendered preferences manifest strongly in uptake: husbands select swimming lessons at higher rates than wives across all treatments, with the largest gap where child lessons are paired with a household good. Once the spousal involvement process is endogenised, however, these differences largely dissipate—deferral, and consultation with switching, both produce convergence in final uptake across treatments. This convergence arises partly because wives more frequently defer and switch to their husbands' preferences for swimming. Autonomy therefore matters: gender gaps persist when it is retained, but the spousal involvement process reconciles these differences when it is relinquished, ultimately mitigating the influence of which spouse moved first.

These insights contribute to a few strands of intrahousehold literature. First, we contribute to the literature that compares outcomes from individual versus joint decision-making. A growing body of work shows that decision outcomes can be significantly altered in domains such as health, education, and fertility, depending on whether husbands or wives are involved in decision-making (Ashraf, 2009; D'Exelle and Ringdal, 2022; D'Exelle et al., 2023; McCarthy, 2019). For example, Ashraf, Field and Lee (2014) demonstrate that women's uptake of contraceptives is lower if the husband is involved in the decision process, whereas D'Exelle and Ringdal (2022) find that uptake of family planning services increases with the husband's involvement. D'Exelle et al. (2023) extend this by showing that involvement effects vary between monogamous and polygynous households. Rather than randomising individual versus joint decision-making, we adopt the approach of Bakhtiar et al. (2024) to endogenise spousal involvement. This allows us to examine how decision-making arrangements arise in practice. Then, by comparing descriptive patterns across decision blocks, we can assess whether differences in outcomes between individual and joint decisions align with findings from the existing literature.

Second, we relate to the literature that examines the determinants of spousal involvement (Abbink, Islam and Nguyen, 2020; Afzal et al., 2022; Dykstra, Exley and Niederle, 2022; D'Exelle and Ringdal, 2025; Hillesland et al., 2025). This differs from the earlier literature, as it focuses specifically on the factors that determine whether people prefer to move from individual to joint decision-making. For instance, some studies find that spousal involvement is higher when the cognitive load of deciding alone is substantial (Arugay et al., 2024; Hillesland et al., 2025) or when perceived abilities differ (Abbink, Islam and Nguyen, 2020), whilst other studies find that spousal involvement is lower when, for example, conflict from preference heterogeneity is anticipated (D'Exelle and Ringdal, 2025). We contribute to this literature, by being the first to exogenously vary determinants such as preference heterogeneity, decision beneficiaries, and the decision-making domain, which allows us to comment on the *conditions* under which wives and husbands are more likely to involve their spouse.

Third, we advance the emerging literature documenting that individuals attach an intrinsic value to holding decision-making authority, independent of the final allocation itself (Ertac, Gumren and Gurdal, 2020; Falk and Kosfeld, 2006; Ferreira, Hanaki and Tarroux, 2020; Ferreira et al., 2025; Freundt and Herz, 2025). While this literature typically abstracts away from marital contexts by studying autonomy where control is artificially imposed or delegated, our design allows us to examine how preferences for autonomy operate when spousal involvement is explicitly domain-structured. By endogenising the involvement decision across varied domains, we are able to disentangle autonomy demand (the preference for holding the right to decide) from spousal involvement (the behavioural choice of how to exercise that right). This approach reveals that rather than being a uniform trait, individuals demand autonomy within their own traditional domains but voluntarily cede it in others—a nuance obscured when control is exogenously assigned rather than chosen. This conceptual distinction also aligns with Arugay et al. (2024), who find that decision participation is only weakly correlated with agency and instead, the strongest predictor of agency is the extent to which respondents can make their own personal decisions if they want to.

Fourth, and finally, we connect these procedural decision-making insights to the literature examining the eventual outcomes of joint decision-making and gendered resource allocation within the household. Existing work shows that when spouses choose together, bargaining power often dictates the outcome, typically tilting the final decision toward the husband's preferences (Anuwong and Munro, 2025; Ashraf, Field and Lee, 2014; Carlsson et al., 2012, 2013; D'Exelle and Ringdal, 2025). Concurrently, a vast literature documents that mothers generally exhibit a stronger preference than fathers for allocating resources to children's development over private consumption (Armand et al., 2020; Dizon-Ross and Jayachandran, 2023; Duflo, 2003; Duflo and Udry, 2004; Haddad, Hoddinott and Alderman, 1997; Lundberg, Pollak and Wales, 1997; Thomas, 1990, 1994). For example, Armand et al. (2020) show that in Burkina Faso, cash transfers targeted to mothers rather than fathers led to larger improvements in children's nutrition and clothing expenditures, and Ringdal and Sjursen (2021) find that increasing women's bargaining power improves gender equality in allocation to children's education. We bridge these insights by evaluating how the endogenised decision process maps onto the final uptake of swimming lessons. Our setting provides novel evidence on how gendered preferences over human capital investments interact with dynamic bargaining structures, showing that while initial preferences diverge sharply along gender lines, the endogenous process of deferral and consultation ultimately narrows these gaps in the final equilibrium allocation.

The rest of this paper is organised as follows. In Section 2, we explain the experimental procedures of the field experiment, followed by a description of the treatments, a simple theoretical model and a series of exploratory hypotheses to test, our empirical strategy, and finally, a brief description of how the experiment was implemented in the field. In Section 3, we present the results, first offering some descriptive statistics of the sample and demographic correlates of decision-making preferences for swimming lessons, before moving to differences across experimental treatments and investigations of how beliefs mediate, and individual characteristics moderate, spouses' choices in each of the decision blocks. A discussion of our results ensues in Section 4, after which we conclude in Section 5 with some reflections.

2. Design

In this section, we first present a summary of the experimental procedures. We then explain the treatment design, followed by a simple theoretical model and a series of exploratory hypotheses to test, our empirical strategy, and finally, a brief description of how the experiment was implemented in the field.[9]

2.1. Experimental procedures

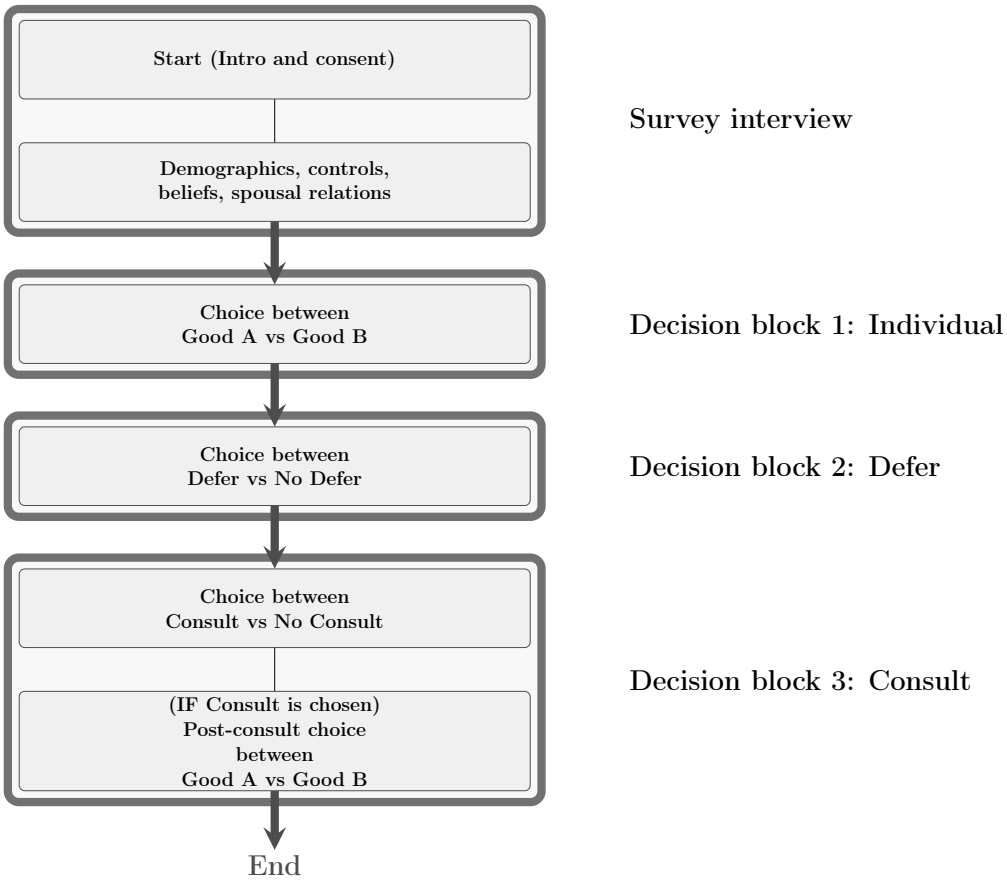
The experiment was conducted with a sample of 664 married couples.[10] Upon providing consent to participate, the experimental sequence for each couple proceeded in the following way (Figure 2.1).

First, both spouses in the couple were ushered into separate rooms, where they were each interviewed by same-gender enumerators using a standard set of household survey questions. After the household questionnaire, each spouse was asked to consider three Decision Blocks, where it was explained that *one* of the

[9]The pre-analysis plan for this study was registered in the OSF Registries on the 14th March 2025 and can be accessed here: https://osf.io/zt9hj.

[10]The full sample was subject to two lab-in-the-field experiments, one of which is discussed in this chapter (experiment 1), the other of which is the focus of Chapter 3 of this thesis (experiment 2). In order to ensure that the experiments did not influence each other, we randomised the order of the experiments. Specifically, half the sample followed the order of survey, experiment 1, then experiment 2, whereas the other half followed the order of survey, experiment 2, then experiment 1. Please take note of this when reading the full experimental script presented in Appendix B.

Figure 2.1: Experimental procedures for spousal choices



couple's decisions between two goods in each block would be randomly implemented for payment (i.e., one of six choices, three made by each spouse). Specifically, the following script was read out:

> *Now I will ask you to make THREE SETS of decisions. Each set of decisions involves a choice between two goods. Your spouse will also be asked the same THREE SETS of decisions. ONE of the SIX decisions made by both of you will be selected at random, and may be given to your household at the end of the interview. Therefore, they are ALL equally important, and you must think carefully about what you are going to choose.*

In Decision Block 1, which we call the **'Individual block'**, individuals were offered a choice between two goods, Good A (always some form of swimming lessons in each treatment) and Good B (an alternative household consumption good). They were asked to make an individual decision about which good they wanted for their household, and their individual choices were recorded.

In Decision Block 2, which we call the **'Defer block'**, individuals were asked to decide whether or not they would like to defer this household decision to their spouse in the next room. It was explained that if they chose to do so, their own choice in the 'Individual block' would be disregarded and instead their spouse's choice would be used for random implementation.

In Decision Block 3, which we call the **'Consult block'**, individuals were asked whether they would like to share their initial choice with the other spouse, and whether they would like to hear their spouse's initial choice. Conditional on the respondent agreeing to hear—and their spouse agreeing to share—individual choices made in the 'Individual block', we showed respondents what their spouse actually chose in the 'Individual block'. For this subset of individuals who met the necessary conditions to induce a 'consultation',

we asked them if they would like to keep their original choice or switch to the other good. After their final choice was made, the interview ended.

2.2. Treatments

We implemented four treatments, in a 2 × 2 design, randomly assigned at the couple level (Figure 2.2). In all treatments, each spouse needed to consider a pair of goods, Good A and Good B, and go through all three Decision Blocks outlined above. In this way, although treatments were assigned at the couple-level, we recorded both spouses' first responses, making it possible for us to track gender differences in outcomes based on the first mover in the decision-making process. We also varied the *types of goods* in the pairing under consideration by treatment. Specifically, we experimentally varied pairings along two dimensions. Along the **'decision-making domain'** dimension, we offered pairings such that Goods A and B fell in either the wife's decision-making domain (wife swimming lessons, wife clothes, or washing powder, i.e., row 1 of Figure 2.2) or the joint decision-making domain requiring the involvement of the husband (child swimming lessons, school books, or rice, i.e., row 2 of Figure 2.2). Along the '**who benefits from Good B'** dimension, we kept Good A fixed (either wife swimming lessons or child swimming lessons) and varied Good B to benefit either an individual (new clothes for wife in the wife domain and exercise books for the child in the joint domain, i.e., column 1 of Figure 2.2) or the entire household (washing powder in the wife domain and rice in the joint domain, i.e., column 2 of Figure 2.2). We note that comparisons across decision-making domains—T1 against T3, and T2 against T4—simultaneously vary both the domain and the swimming beneficiary (wife to child), and thus effects in these comparisons cannot be attributed to either dimension alone. We therefore treat them as bundled comparisons (see Section 5 for an extended reflection on this).

Figure 2.2: Experimental treatments for spousal choices

		Who benefits	
		Good B benefits the individual (wife/child)	Good B benefits the household
Whose decision domain	Wife's decision-making domain	T1 {Wife swimming lessons, wife clothes}	T2 {Wife swimming lessons, washing powder}
	Joint decision-making domain	T3 {Child swimming lessons, books}	T4 {Child swimming lessons, rice}

2.3. Theoretical framework and hypotheses

In this section, we formulate hypotheses on how decisions vary between treatments in each of the three Decision Blocks. To structure our arguments, we first put forth a simple theoretical framework to model spousal behaviour in the experiment.

Consider a household consisting of a wife w, a husband h, and a child c. In Decision Block 1, each spouse $s \in \{w, h\}$ independently chooses between two goods $x \in \{A, B\}$. Good A is always swimming lessons—wife swimming (WS) or child swimming (CS)—while good B varies across treatments.

Each household member $j \in \{w, h, c\}$ derives individual utility $u_j(x)$ from the selected good. Goods differ in the identity of the primary beneficiary. Some goods generate primarily private benefits for one household

member, while others generate benefits shared by the entire household. Spouse s evaluates alternatives according to the utility function:

$$U_s(x) = \beta_s^w u_w(x) + \beta_s^h u_h(x) + \beta_s^c u_c(x), \tag{2.1}$$

where $\beta_s^j \geq 0$ for all j and $\sum_j \beta_s^j = 1$. Spouse s chooses good A over B whenever:

$$U_s(A) - U_s(B) > 0. \tag{2.2}$$

The experimental treatments vary the bundle of goods A and B along two dimensions. First, the *decision-making domain* determines whether the bundle falls within the wife's domain or within the joint household domain. Second, the *beneficiary dimension* varies whether good B of the bundle benefits a single household member (private good) or the entire household (household good).

To allow for potential disagreement, we assume spouses may differ in the relative weight they place on each household member's utility. In particular, individuals may place greater weight on their own individual utility than their spouse does (on this see, e.g., Browning, Chiappori and Weiss (2014)), such that $\beta_w^w > \beta_w^h$ and $\beta_h^h > \beta_h^w$. Thus, variation across treatments operates solely through differences in the distribution of benefits across household members. Using this model, we are able to make the following predictions under our three Decision Blocks.

Decision Block 1

From the model above, when presented with a choice in Decision Block 1, each spouse compares the weighted sum of individual utilities generated by goods A and B. We expect spouses to evaluate the choices differently between treatments, given that a) they have different individual utilities, b) they place different weights on household members' utilities. Put differently, we expect spousal choices in Decision Block 1 to reflect diverging preferences where treatments vary good B, keeping good A fixed. We formalise this with the following hypothesis at the *couple-level*:

Hypothesis 2.1: *Spousal preference differences (SPDs) in Decision Block 1 differ between treatments T1 and T2 and between treatments T3 and T4, i.e.,* $|SPD(T2) - SPD(T1)| > 0$ *and* $|SPD(T4) - SPD(T3)| > 0$.

Treatment variation between T1 (WS/clothes) and T2 (WS/soap) arises from changes in the identity of the primary beneficiary of good B (wife vs household). Similarly, treatment variation between T3 (CS/books) and T4 (CS/rice) arises from changes in the identity of the primary beneficiary of good B (child vs household). We expect husbands and wives to evaluate the choices differently between treatments, given that they make trade-offs based on individual preferences and weights. Thus, we expect the proportion of couples with spousal preference differences (SPDs) in choices (i.e., where $x_w \neq x_h$) to differ between T1 and T2 and between T3 and T4.

Decision Blocks 2 and 3

Decision Blocks 2 and 3 introduce opportunities for spousal involvement in the decision-making process. After making an initial choice in Block 1, individuals are given the opportunity to modify how the final decision is determined. In Decision Block 2, respondents must decide whether to keep their own initial choice or defer the decision to their spouse. Deferral implies that the spouse's initial choice becomes the final decision, as the individual relinquishes decision authority to the spouse. In Decision Block 3, respondents decide whether to engage in consultation i.e., they must decide whether or not to share their own choices and listen to

one's spouse's choices made in Decision Block 1. Note, the individual's choice to engage in consultation is insufficient to induce 'spousal consultation' i.e., the opportunity to *revise* one's choice after hearing one's spouse's choices. This transpires only if the individual agrees to listen to their spouse's initial choice, *and their spouse* agrees to share this initial choice. The subset of respondents that reach this consultation will learn of their spouse's initial choice, and must decide whether to switch to their spouse's choice or retain their original choice.

The key behavioural consideration in these blocks is the possibility of disagreement between spouses. In Block 1, individuals make decisions independently and without knowledge of their spouse's choice. In Blocks 2 and 3, however, respondents know that their spouse has also made a choice, and that diverging preferences may emerge when final decisions are revealed at the end of the experiment. Anticipating such disagreement may motivate individuals either to defer the decision to their spouse or to seek consultation before the final outcome is determined. In particular, the likelihood of deferral or consultation may depend on two factors: 1) decision domains, and 2) spousal preference differences. We elaborate on these below with some hypotheses from the pre-analysis plan.

Comparisons between decision domains: In joint domains where decision-making authority is less clearly defined within the household (Liaqat et al., 2026), spouses may perceive decision authority to be shared rather than individual. In this case, individuals would be more willing to defer or consult the decision with their spouse. We formalise this in the following way:

Hypothesis 2.2: *Deferral in Decision Block 2 is more common in the joint decision-making domain than in the wife's decision-making domain, i.e.,* $DEF(T3) > DEF(T1)$ *and* $DEF(T4) > DEF(T2)$.

Hypothesis 2.3: *Consultation in Decision Block 3 is more common in the joint decision-making domain than in the wife's decision-making domain, i.e.,* $CON(T3) > CON(T1)$ *and* $CON(T4) > CON(T2)$.

Comparisons between spousal preference differences (SPDs): If individuals anticipate that their spouse may prefer a different option, they would be more likely to defer or consult the decision in order to avoid disagreement and allow the spouse's preference to be prioritised. Given hypothesis 2.1 that SPDs are positive between T1–T2 and T3–T4, we would also expect differences in deferral and consultation rates between these treatment pairs, i.e., between treatments where spousal preferences are amplified. This is formalised in the following way:

Hypothesis 2.4: *Deferral in Decision Block 2 is more common when spousal preference differences are larger, i.e.,* $|DEF(T2) - DEF(T1)| > 0$ *and* $|DEF(T4) - DEF(T3)| > 0$.

Hypothesis 2.5: *Consultation in Decision Block 3 is more common when spousal preference differences are larger, i.e.,* $|CON(T2) - CON(T1)| > 0$ *and* $|CON(T4) - CON(T3)| > 0$.

We note that if hypothesis 2.1 does not hold, then hypotheses 2.4 and 2.5 cannot be interpreted as tests of whether SPD variation affects deferral and consultation. In this case, any T1–T2 or T3–T4 differences in deferral and consultation rates would reflect the direct effect of the paired-good variation, not its effect via SPDs. Finally, given local patriarchal gender norms around permission-seeking from the head of the household (i.e., the husband) (Brugere et al., 2020), we expect the differences under hypotheses 2.2-2.5 to be larger among wives than husbands. Formally, we have:

Hypothesis 2.6: *We expect more pronounced treatment effects for wives than husbands, under hypotheses 2.2-2.5.*

Put differently, on average, we expect wives to be more likely to defer and consult than husbands when we consider our relevant treatment pairs.

2.4. Empirical strategy

Prior to evaluating our primary causal treatment effects, we examine the baseline demographic and behavioural determinants of individual swimming preferences within each decision-making domain. To isolate which observable characteristics predict an individual's propensity to choose swimming instruction, we estimate pooled baseline ordinary least squares (OLS) specifications of the following form:

$$\mathrm{Y}_{ic} = \alpha_0 + \mathbf{X}'_{ic}\boldsymbol{\gamma} + \omega_{ic}, \tag{2.3}$$

where for each individual i in couple c, Y_{ic} is our indicator variable First Choice$_{ic}$ from Decision Block 1, which takes a value of 1 if the participant selects swimming lessons and 0 otherwise. The vector $\mathbf{X}_{ic}$ represents a set of individual demographic characteristics, including the respondent's age, an indicator for the spousal age difference, the count of sons and daughters in the household, a binary measure tracking the respondent's own swimming capability, and measures of time preferences, risk aversion, educational attainment, and total years of marriage. The parameter ω_{ic} captures the idiosyncratic error term. We estimate equation (2.3) separately by gender and pooled across decision-making domains. Specifically, columns (1) and (3) of our baseline analysis pool treatments 1 and 2 to capture correlates within the separate female domain, while columns (2) and (4) pool treatments 3 and 4 to isolate correlates within the joint child-investment domain. As a robustness check and to ensure transparency, we also estimate this specification disaggregated by individual treatment arms (Appendix G).

To test for treatment effects at the individual-level, we regress:

$$Y_{ic} = \beta_1 + \beta_2 \mathrm{T1}_{ic} + \beta_3 \mathrm{T3}_{ic} + \beta_4 \mathrm{T4}_{ic} + \beta_5 \mathrm{X}_{ic} + \mu_e + \varepsilon_{ic} \tag{2.4}$$

where for each spouse i in couple c, Y_{ic} is one of the following dependent variables: First choice$_{ic}$, a binary variable with value 1 if the individual chooses swimming lessons, zero otherwise; Defer$_{ic}$, a binary with value 1 if the individual chooses to defer their decision to their spouse, zero otherwise; and Consult$_{ic}$, a binary variable with value 1 if the individual is open to consultation with their spouse (i.e., they are willing to share *or* listen to their spouse's preferences), zero otherwise. In this model, T2 is the reference group, and β_1 is the mean outcome under T2. T1, T3, and T4 are dummy variables for each treatment group, where coefficients β_2, β_3 and β_4 estimate the treatment effects of T1, T3, and T4 respectively, relative to T2. We also include fixed effects for the specific enumerator who conducts the interview (μ_e for enumerator e). X_{ic} is a vector of control variables used to increase the efficiency of the estimates. They include the respondent's age, education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with their spouse. Last, ε_{ic} is the error term.

We estimate the equation by fitting seemingly unrelated regression (SUREG) models. We do this to simultaneously estimate the choices of wives and husbands, taking into account possible correlations in unobserved factors that might influence the choices of both partners. We also account for within-village correlation by clustering standard errors at the village level. However, with only 19 village clusters, conventional cluster-robust standard errors may underestimate true sampling variability and yield overly optimistic p-values. To address this, following Cameron, Gelbach and Miller (2008), we bootstrap standard errors clustered at the village level and report the same regression results as robustness checks in Appendix F (section F.1).

To test hypotheses 2.1-2.6, we examine the coefficient on the T_{ic} whenever the hypothesis compares T_{ic} with T2 (our base treatment in the SUREGs). Where the hypothesis compares T_{ic} and T_{jc} where $i \neq j \neq 2$,

we perform a Wald test with the null hypothesis of $H_0 : \beta_i = \beta_j$ to test for statistically significant differences in choices between treatments. We are also interested in examining whether certain subgroups might drive treatment differences, to be able to say something about possible mechanisms. For this, we run the following regression model:

$$Y_{ic} = \theta_1 + \sum_{j=2}^{4} \theta_j \mathrm{T}_j + \mathrm{Var}_{ic} + \sum_{j=2}^{4} \gamma_j (\mathrm{T}_j \times \mathrm{Var}_{ic}) + \mu_e + \varepsilon_{ic} \tag{2.5}$$

where for each spouse i in couple c each treatment indicator T_j for $j = 1, 3, 4$ is interacted with a variable Var_{ic}, which represents any variable for which we are interested in examining heterogeneity. For example, if we would like to test that conflict aversion increases the proportion of individuals who defer, Y_{ic} would be Defer_{ic} and Var_{ic} would be some variable capturing the respondent's aversion to conflict. Examining each interaction terms' γ_j coefficients for $j = 1, 3, 4$ tells us if there are differences in treatment effects on deferral behaviour in T1, T3, and T4, respectively, relative to T2, for spouses who are averse to conflict.

Beyond individual heterogeneity, we formally assess whether respondent's expectations regarding their partner's initial preference might mediate spousal involvement choices. This mediation analysis is specified as follows:

$$Y_{ic} = \alpha_1 + \sum_{j=2}^{4} \alpha_j \mathrm{T}_j + \lambda \mathrm{ESPD}_{ic} + X_i' \delta + \epsilon_{ic} \tag{2.6}$$

where Y_{ic} is an indicator for spousal involvement behaviour (e.g., deferral or consultation choice), T_i denotes treatment assignment, ESPD_{ic} is the indicator for expected spousal preference differences, and X_i is a vector of baseline controls. Standard errors are clustered at the village level. Because subjective beliefs are themselves endogenously shaped by the initial treatment assignment, we interpret the parameter λ as a descriptive behavioural channel rather than a causal mechanism per se. By comparing the baseline treatment effects β_j in equation (2.4) with the attenuated direct effects α_j in equation (2.6), we trace the extent to which expected disagreement explains the observed demand for spousal involvement.

2.5. Sampling and implementation

Our research site was chosen on the basis that the majority of women and children in Unguja—the Zanzibar archipelago's biggest and most populous island—do not know how to swim (Royal National Lifeboat Institution, 2019). This is of policy interest as Zanzibari women and children are at high risk of drowning, given frequent exposure to the ocean through aquatic leisure activities, waterborne transport, and coastal labour in the local fisheries, tourism, and seaweed-farming sectors. The sample is also of empirical interest as conservative Islamic values have historically had some influence on swimming being viewed as a male-specific activity.

We recruited the sample of 664 married couples through the following sampling approach. We partnered with a local non-profit organisation, the Panje Project (Panje henceforth), which offers free aquatic survival training to Zanzibari women and children—a rare opportunity for them to build their human capital. Panje provided us with a list of 19 coastal villages covering 5 of the 7 districts in Unguja (see Figure A.1 in Appendix A for village locations), where coastal labour requiring swimming skills is in demand. We worked with each village's respective leader—a local government official known as the *Sheha*—to select a sample of 40 married couples per village (higher than required to account for possible no-shows on the day). The only eligibility criterion we enforced was that the married couple had at least one child of age six years or older but younger than 18 years—the age group to which Panje offers swimming lessons. To recruit the sample,

the village leader worked together with local mobilisers, calling or visiting each household and inviting them to participate in the study on a particular date and time, at a specified location. The only information participants received about the study was that it was for research aiming to improve their livelihoods, that they would be compensated TZS 4,000 (≈USD 1.5) for their time, and that the household would have the chance to receive additional goods.

3. Results

The following sections present the main results from our analysis. We start with some simple descriptive statistics of our sample. We then move to an analysis of choices made in Decision Block 1, looking at both average treatment effects (ATEs) and whether spousal preference differences arise from our treatments, as per hypothesis 2.1. We also examine some correlates with choices in Decision Block 1, such as spousal beliefs around each other's choices. Next, we move to Decision Blocks 2 and 3 where once again we examine ATEs on deferral and consultation choices, in order to examine hypotheses 2.2-2.5. Presenting these ATEs by gender also allows us to examine whether deferral and consultation differ between husbands and wives. ATEs are followed by regressions where treatments are interacted with specific variables to examine possible behavioural mechanisms (including beliefs and conflict aversion). We also present results of a heterogeneity analysis which examines whether treatment effects differ by certain subgroups. Last, we examine the effects of deferral and consultation on the final uptake of swimming lessons, including a closer look at switching behaviour.

3.1. Sample

Table 2.1 presents some gender-disaggregated summary statistics of the 1,328 spouses, indicating p-values of t-tests for statistical difference in mean outcomes for husbands and wives.

The average ages of husbands and wives are 49 and 41 years, respectively. Over 86% of spouses have received at least a primary education, and this does not differ by gender. Husbands on average earn significantly more than wives, with TZS 400,000 (∼ USD 163) compared to TZS 80,000 (∼ USD 32) in a good month, and TZS 158,000 (∼ USD 64) compared to TZS 37,000 (∼ USD 15) in a bad month, respectively. Husbands also contribute the majority of costs (over 80%) for child, personal, and household expenses. Although we would expect to see the same mean years of marriage/cohabitation given the sample is of married/cohabiting couples, this is slightly different for men (20 years) and women (19 years). In line with the overall finding in the literature that men are less risk-averse than women, husbands in our sample score lower than wives on our risk aversion scale. Similarly in line with the competition literature, the sample indicates that husbands have (slightly) stronger competition preferences than women as they score higher on our competitiveness index. Husbands also score higher than wives on our self-efficacy index and marital trust scale. However, contrary to the literature, both spouses appear to have the same average scores on our conflict aversion index.

Since couples were randomly assigned to our experimental treatments, we should expect differences in characteristics between treatments to be due to chance, for both husbands and wives. Appendix A presents balancing tables on a range of indicators of relevance to treatment outcomes, including age, education risk aversion, years of marriage, etc. For wives, we fail to reject the null that the treatments are balanced on all characteristics. For husbands, treatments are balanced on all characteristics except on how much income is earned in a good month, with those in T4 showing a lower mean monthly income than other treatments. This is due to a few outliers with very high monthly incomes in other treatments. To control for these differences, we include monthly incomes as controls in our main regressions, and run separate robustness checks (these are reported in Appendix F, under section F.2).

Table 2.1: Summary statistics of socio-demographic characteristics

Variable	(1) Husbands Mean/(SD)	(2) Wives Mean/(SD)	(1)-(2) Pairwise t-test P-value
Age	48.831	41.161	0.000***
	(13.291)	(11.099)	
Primary education	0.886	0.860	0.162
	(0.319)	(0.347)	
Monthly income in good month	400.339	81.306	0.000***
	(791.455)	(118.762)	
Monthly income in bad month	158.142	36.968	0.000***
	(359.415)	(77.877)	
Years of marriage/cohabitation	20.253	19.017	0.074*
	(12.697)	(12.497)	
% of personal costs covered	83.554	56.536	0.000***
	(24.801)	(27.092)	
% of household costs covered	84.834	38.614	0.000***
	(16.171)	(22.247)	
% of child costs covered	80.648	42.937	0.000***
	(17.950)	(23.256)	
Attitude to risk	1.125	2.273	0.000***
	(1.091)	(0.933)	
Marital trust	8.827	7.697	0.000***
	(1.542)	(2.378)	
Competitiveness index	21.593	19.931	0.000***
	(4.777)	(5.031)	
Self-efficacy index	17.051	15.074	0.000***
	(3.160)	(2.699)	
Conflict aversion index	23.407	23.345	0.765
	(4.051)	(3.456)	
Number of observations	664	664	1328

Notes: Summary statistics of the following variables: 'Age' which captures the respondent's age; 'Primary education' which is a dummy taking the value 1 if they have at least a basic primary education; 'Income in a good/bad month' which is the respondents' estimate of monthly income in a good/bad month given in Tanzanian shillings (in 000's); 'Years of marriage' gives the number of years the respondent has been married; '% of child/personal/household costs covered' is the respondents' estimate between 0-100 of how much they contribute to each expense item; 'Attitude to risk' which is a categorical variable capturing how willing the respondent is to take risks, coded as 0 'Takes risk a lot', 1 'Takes risk but not a lot', 2 'Avoids risk but not a lot', and 3 'Avoids risk a lot'; 'Marital trust' which is a continuous variable of responses to the question 'On a scale of 1 to 10, where 1 is 'not at al' and 10 is 'completely', how much do you trust your spouse's judgment when making important decisions that affect both of you?'; 'Competitiveness index' which is an index that captures the sum of the respondent's answers to a set of six Likert-scale questions on their preference for competition, with 6=lowest score for competitive preferences and 30=highest score for competitive preferences; 'Self-efficacy index' which is an index that captures the sum of the respondent's answers to a set of six Likert-scale questions on their self-efficacy beliefs, with 6=lowest self-efficacy score and 30=highest self-efficacy score; and 'Conflict aversion index' which is an index that captures the sum of the respondent's answers to a set of six Likert-scale questions on their attitudes to conflict, with 6=lowest conflict aversion score and 30=highest conflict aversion score. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Before turning to treatment effects, we examine baseline correlates of swimming preferences. Table 2.2 reports OLS regressions of first-choice swimming lesson preference on key demographic characteristics, pooled within each decision domain (full results by separate treatments are reported in Table G.1.). Specifically, columns (1) and (3) pool treatments 1 and 2 to capture correlates within the wives' domain, while columns (2) and (4) pool treatments 3 and 4 to isolate correlates within the joint domain. A few patterns stand out. In the wife's domain (cols 1 and 3), education is positively associated with swimming lesson uptake for both wives and husbands, and age is negatively associated for both. Risk aversion exhibits a striking gender asymmetry: more risk-averse wives are less likely to prefer swimming lessons, whereas more risk-averse

husbands are more likely to do so. In the joint domain (cols 2 and 4), having sons predicts higher preferences for child swimming lessons among wives, and wives who already swim are substantially more likely to want their child to take lessons (this is in line with evidence presented by Borgonovi, Seitz and Vogel (2022)). Beyond these patterns, most coefficients are small in magnitude and individually insignificant, and no single characteristic predicts preferences consistently across both domains or both genders. Because observable characteristics we measure appear to account for only a small fraction of the overall variation in preferences, it is more likely that preference heterogeneity is derived from the experimental design, which we examine next.

Table 2.2: Demographic correlates of decision-making preferences

	Wives		Husbands	
	(1) T1/T2	(2) T3/T4	(3) T1/T2	(4) T3/T4
Age	-0.004*	0.006	-0.004*	0.003
	(0.002)	(0.004)	(0.002)	(0.005)
Spousal age difference	-0.004**	-0.003	-0.003	-0.011
	(0.002)	(0.004)	(0.005)	(0.007)
Sons	0.001	0.035**	-0.000	0.027
	(0.016)	(0.014)	(0.012)	(0.016)
Daughters	-0.021	-0.016	-0.008	-0.001
	(0.014)	(0.018)	(0.014)	(0.015)
Swimming skill	0.051	0.225*	0.068	0.082
	(0.100)	(0.109)	(0.055)	(0.075)
Time preference	-0.039	-0.012	-0.045	-0.090*
	(0.041)	(0.057)	(0.039)	(0.046)
Risk aversion	-0.107**	-0.122*	0.124*	0.050
	(0.049)	(0.064)	(0.063)	(0.067)
Education	0.075**	0.021	0.096**	0.025
	(0.030)	(0.104)	(0.041)	(0.118)
Years married	0.002	-0.008**	0.002	-0.001
	(0.002)	(0.003)	(0.002)	(0.005)
N	332	332	332	332

Notes: OLS regressions with 'First choice' (= 1 if swimming lessons chosen in Decision Block 1, zero otherwise) as the dependent variable. Columns (1) and (3) pool T1 and T2 (wife swimming, wife's domain); columns (2) and (4) pool T3 and T4 (child swimming, joint domain). Variables are defined as follows: 'Age' captures the respondent's age in years; 'Spousal age difference' is the absolute difference in years between the wife's and husband's age; 'Sons' and 'Daughters' give the number of sons and daughters in the household, respectively; 'Swimming skill' is a binary variable equal to 1 if the respondent reports being able to swim in deep water for 10 minutes without stopping, zero otherwise; 'Time preference' is a binary equal to 1 if the respondent chose the earlier, smaller payment option (Tsh 10,000 tomorrow over Tsh 15,000 in one month) in all three rounds of a sequential binary time preference elicitation task, indicating the most impatient type, zero otherwise; 'Risk aversion' is a categorical variable coded as 0 ('Takes risk a lot'), 1 ('Takes risk but not a lot'), 2 ('Avoids risk but not a lot'), and 3 ('Avoids risk a lot'); 'Education' is a binary equal to 1 if the respondent has at least a primary education, zero otherwise; and 'Years married' gives the number of years the respondent has been married. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.2. Decision Block 1

In this section, we examine how the initial choices made by spouses in Decision Block 1 vary across treatments. To do so, we examine two variables: 'SPDs', a binary variable that captures spousal preference differences by assigning a value of one to couples where the initial choices are different between spouses, and zero if their choices are the same, and 'First choice', a binary variable which takes the value one if the respondent's choice in Decision Block 1 was swimming lessons, zero otherwise. In both cases, we look at average treatment effects (ATEs). Within-domain comparisons, i.e., T1–T2 and T3–T4, hold both the decision-making domain and the swimming beneficiary fixed while varying only the type of Good B, and therefore permit clean identification of any Good B type effect. As an exploratory analysis, we also compare treatment pairs where we vary

decision domains i.e., T1–T3 and T2–T4. These cross-domain comparisons simultaneously change both the decision-making domain and the swimming beneficiary (from wife to child), which means that effects cannot be attributed cleanly to either dimension alone (we discuss this in Section 5).

Table 2.3 presents treatment comparisons around spousal preference differences (SPDs) in Decision Block 1 (column 1), which allows us to test hypothesis 2.1, and treatment comparisons around spouses' initial choices in Decision Block 1 (column 2 for wives and column 3 for husbands), to examine the uptake of swimming lessons. In all regressions, we use T2 as the base treatment.

Table 2.3: Treatment effects for Decision Block 1

	SPDs	First choice	
	Couples	Wives	Husbands
T1 - WS/Kanga	-0.036	0.006	-0.018
	(0.053)	(0.041)	(0.033)
T3 - CS/Books	0.127*	0.133**	0.090**
	(0.067)	(0.053)	(0.045)
T4 - CS/Rice	0.133**	0.283***	0.343***
	(0.061)	(0.043)	(0.051)
Predicted mean outcome under T2	0.313	0.157	0.229
F-test p-value for T1 vs T3	0.011	0.001	0.019
F-test p-value for T3 vs T4	0.938	0.002	0.000

Notes: Seemingly unrelated regressions are run on the 'First choice' dependent variable, which is equal to one if swimming lessons are chosen in Decision Block 1, zero otherwise. Wife and husband regressions have 664 observations each. OLS regression run on 'SPDs', a dependent variable equal to one if the first choices of spouses in a couple differ from each other, zero if choices are the same. 664 couples in this latter regression. In all regressions, 'T2 - WS/Soap' is the base treatment. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Comparing treatment pairs where good A remains fixed and good B is varied i.e., T1–T2 and T3–T4, we note the following results. Looking at column 1, we find no statistically significant spousal preference differences between treatment pairs for both T1–T2 (col 1, T1 coef = -0.036) and T3–T4 (col 1, two-sided F-test T3 vs T4, $p = 0.938$). Thus, we cannot confirm hypothesis 2.1. Looking at columns 2 and 3, we find the uptake of (wife) swimming lessons (WS) is not statistically significantly different between T1 and T2, for both wives and husbands. However, we find wives and husbands 15 percentage points and 25.3 percentage points, respectively, more likely to choose (child) swimming lessons (CS) in T4 than T3 (by taking the difference in T4 and T3 coefficients), and these results are statistically significant for wives (col 2, two-sided F-test T3 vs T4, $p = 0.002$) and husbands (col 3, two-sided F-test T3 vs T4, $p = 0.000$).

As an exploratory analysis, we also compare treatment pairs where we vary decision domains i.e., T1–T3 and T2–T4. We find significant SPDs between treatments for both pairs T1–T3 (col 1, two-sided F-test of T1 vs T3, $p = 0.011$) and T2–T4 (col 1, T4 coef = 0.133^{**}). Similarly, examining initial choices, wives and husbands are 12.7 percentage points and 10.8 percentage points, respectively, more likely to choose CS in T3 than WS in T1 (col 1, two-sided F-test $p = 0.001$; col 3, two-sided F-test $p = 0.019$), and wives and husbands are 28.3 percentage points and 34.3 percentage points, respectively, statistically significantly more likely to choose CS in T4 than WS in T2 (col 2, coef=0.283^{**}; col 3, coef = 0.343^{***}). We next examine whether spousal beliefs independently shape first choices in Decision Block 1.

Table E.1 (Appendix E) presents results of three interaction models, where treatment indicators are interacted with 'Beliefs' (whether the individual believes their spouse chose swimming), 'Spouse's choice' (the spouse's actual first choice), and 'Spouse's belief' (whether the spouse believes the individual chose swimming). We find significant coefficients on own beliefs (Model 1, columns 1 and 2) and on spouse's beliefs (Model 3, columns 5 and 6). Specifically, in the base treatment, individuals who believe their spouse chose

swimming, and those whose spouses believe they will choose swimming,[11] are more likely to choose swimming themselves. By contrast, the coefficient on 'Spouse's actual choice' is not significant (Model 2, columns 3 and 4), indicating no evidence of preference coordination in the individual block.

The interaction models further reveal that some treatment effects differ across belief groups. Wives' higher CS uptake in T3 relative to T2 is driven by those who believe their husbands also chose swimming (column 1, T3 × Beliefs = 0.218), and wives whose husbands believe the wife will choose swimming are 27 percentage points less likely to choose swimming in T1 relative to T2. However, we note that correlations between choices and beliefs/spouses' beliefs are not necessarily causal. For instance, since individual preferences are elicited before beliefs are elicited, it might be that spouses' first choices influence their beliefs about spousal preferences.

Finally, it is worth noting the differences in swimming lessons as first choices between wives and husbands for all treatments. The predicted means under T2 are 16% for wives and 23% for husbands, and these are statistically different (two-sided F-test, chi2=5.27, $p = 0.0168$). Similarly, first choices are statistically different between genders for T4 (two-sided F-test, chi2=3.86 $p = 0.0495$). However, we find that they are not significantly different between genders for either T1 (two-sided F-test, chi2= 1.98, $p = 0.1591$) or T3 (two-sided F-test, chi2= 0.59, $p = 0.4410$). We summarise results from this section in the following way:

Result 1: *Within-domain treatment variation (T1–T2 and T3–T4) generates no statistically significant spousal preference differences, and the uptake of wife swimming lessons is not significantly affected. Exploratory cross-domain comparisons (T1–T3 and T2–T4) reveal both greater spousal preference divergence and higher uptake of child than wife swimming lessons. Beliefs about spousal preferences—but not spouses' actual choices—are correlated with individual first choices, with certain treatment effects driven by belief-holding subgroups; the direction of this relationship is unclear given the timing of belief elicitation.*

3.3. Decision Blocks 2 & 3

In this section, we examine how treatments affect choices in Decision Blocks 2 and 3. Specifically, Table 2.4 presents average treatment effects for two variables: 'Defer' in Decision Block 2, which takes the value one if the respondent decides to defer the choice to the spouse, zero otherwise; and 'Consult' in Decision Block 3, which is equal to one if they are willing to consult with the spouse (i.e., if they are open to either hearing their spouse's first choice or sharing their own first choice with their spouse), zero otherwise. Once again, all regressions use T2 as the base treatment.

Comparing treatment pairs which move from individual to joint decision domains—i.e., T1–T3 and T2–T4—allow us to test hypotheses 2.2 and 2.3. Comparing treatment pairs that change the beneficiary of good B whilst holding good A fixed—i.e., T1–T2 and T3–T4—allow us to test hypotheses 2.4 and 2.5.

Two caveats bear on the interpretation of these comparisons. First, the cross-domain pairs (T1–T3 and T2–T4) simultaneously vary both the swimming beneficiary and the decision-making domain, so any treatment effects on spousal involvement cannot be attributed to either alone; we return to this design limitation in Section 5. Second, the within-domain comparisons (T1–T2 and T3–T4) were constructed on the basis that hypothesis 2.1 held. However, the results from Table 2.3 indicate that SPDs do not vary for these treatment pairs. Since hypothesis 2.1 does not hold—i.e., varying Good B within a fixed Good A fails to shift SPDs—hypotheses 2.4 and 2.5 cannot be interpreted as tests of whether SPD variation affects

[11] Although it is not possible for an individual to know what their spouse thinks, 'Spouse's beliefs' is of interest as correlations suggest that choices co-vary with what individuals *anticipate* their spouse to think—a second-order belief.

deferral/consultation. Instead, any T1–T2 or T3–T4 differences reflect the direct effect of the paired-good variation. We therefore treat these comparisons as exploratory—a point we return to in Section 4.

Table 2.4: Treatment effects for Decision Blocks 2 and 3

	Defer		Consult	
	(1)		(2)	
	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	0.108*	0.012	0.012	-0.072*
	(0.061)	(0.053)	(0.043)	(0.041)
T3 - CS/Books	0.169***	-0.066	0.102***	0.000
	(0.052)	(0.052)	(0.038)	(0.037)
T4 - CS/Rice	0.084**	-0.127**	0.036	0.006
	(0.043)	(0.055)	(0.042)	(0.037)
Predicted outcome under T2	0.392	0.548	0.807	0.867
F-test p-value for T1 vs T3	0.300	0.137	0.015	0.077
F-test p-value for T3 vs T4	0.046	0.286	0.065	0.870

Notes: Seemingly unrelated regression with the following dependent variables. Model (1): defer in Decision Block 2 where the choice to defer decision-making to the spouse takes the value 1, and the choice to make the decision alone takes the value 0; Model (2): consult in Decision Block 3, where the willingness to consult with the spouse takes the value 1 (i.e., if they are open to either hearing their spouse's first choice or sharing their own first choice with their spouse), 0 if they are closed to either hearing or sharing. The base treatment is 'T2 - WS/Soap'. Wife and husband regressions have 664 observations each. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Model (1) in Table 2.4 presents the treatment differences in deferral for wives and husbands, respectively. Comparing treatment pairs which move from individual to joint decision domains, we find the following results. Wives are 8.4 percentage points more likely to defer the decision to husbands in T4 relative to T2 (col 1, positive and significant treatment coefficients on T4), whilst husbands are 12.7 percentage points *less* likely to defer in T4 than T2 (col 2, T4 coef). However, we do not find statistically significant differences in deferral in T3 relative to T1, for wives nor husbands (cols 1 and 2, two-sided F-test of T1 vs T3, $p = 0.300$ and $p = 0.137$, respectively). Thus we have partial support for hypothesis 2.2 that deferral is more common in the joint decision domain than the wife's decision domain. Comparing treatment pairs that change the beneficiary of good B whilst holding good A fixed, we find wives 10.8 percentage points more likely to defer the decision to husbands in T1 relative to T2 (col 1, T1 coef), and 8.5 percentage points more likely to defer to husbands in T3 relative to T4 (col 1, difference in T3 and T4 coefs, two-sided F-test $p = 0.046$). For husbands, we do not find statistically significant deferral rates for both T1–T2 and T3–T4 treatment pairs. Thus, we can confirm hypothesis 2.4, but only for wives (exploratory; see above).

Model (2) in Table 2.4 compares spouses' willingness to consult across treatments. Comparing treatment pairs which move from individual to joint decision domains, we find that wives are 9 percentage points more likely to consult in T3 than T1 (col 3, difference in T3 and T1 coefs, two-sided F-test $p = 0.015$), and husbands are 7 percentage points more likely to consult in T3 than T1 (col 4, difference in T3 and T1 coefs, two sided F-test $p = 0.077$). However, we do not find statistically significant differences in consultation in T4 relative to T2, for wives nor husbands (coefs on T4, cols 3 and 4, respectively). Thus we have partial support for hypothesis 2.3 that consultation is more common in the joint decision domain than the wife's decision domain. Comparing treatment pairs that change the beneficiary of good B whilst holding good A fixed, we find that wives are 7 percentage points more willing to consult in T3 than T4 (col 3, difference in T3 and T4 coefs, two-sided F-test $p = 0.065$), but no treatment effects on consultation in T1 relative to T2 (col 3, coef T1). We also find husbands 7 percentage points less likely to consult in T1 than T2 (col 4, coef T1), but find no treatment effects on consultation in T4 relative to T3 (col 4, two-sided F-test $p = 0.870$). Thus we have partial support for hypothesis 2.5 (exploratory; see above).

Table D.3 (Appendix D) complements the regression results with descriptive rates of deferral, consultation, and switching by treatment and gender. Both spouses display a high willingness to consult across

all treatments (above 80%), with no statistically significant gender differences in openness to consult. Once consultation occurs, however, switching behaviour diverges. In particular, wives are significantly more likely to resign to their spouse's preference than husbands in T1 and T4, with the largest gap in T4 where wives are 8 percentage points more likely to switch. We summarise these results as follows.

Result 2a: *When varying decision domains, we find partial evidence that joint decision domains increase spousal involvement. For deferral, this occurs in gender-asymmetric ways: wives (husbands) become more (less) likely to defer when moving from individual to joint domains, but only in the* ***household****-benefitting treatment pair (T2–T4). For consultation, both wives and husbands become more willing to consult when moving from individual to joint domains, but only in the* ***individual****-benefitting treatment pair (T1–T3). Both spouses display a high baseline willingness to consult (above 80% across all treatments), with no statistically significant gender differences in openness to consult.*

Result 2b: *When varying the beneficiary of good B whilst holding good A fixed, wives are more likely to defer in both the wife's domain (T1–T2) and the joint domain (T3–T4), but treatment effects on husbands' deferral behaviour are generally absent. For consultation, treatment effects are only found in the wife's domain (T1–T2), with wives (husbands) consulting more (less) in T1 than T2. Once consultation occurs, switching behaviour is gender-asymmetric: wives are significantly more likely to resign to their husband's preference in T1 and T4, with the largest gap in T4 (8 percentage points).*

Since within-domain treatment variation shifts deferral and consultation independently of actual SPDs, we examine in Section 3.4 whether *expected* spousal preference differences—i.e., beliefs about whether one chose differently to one's partner—serve as the operative mechanism.

3.4. Mediating effects of beliefs on spousal involvement behaviour

Having established that preferences for spousal involvement vary systematically across decision domains, we now explore whether expected preference differences serve as the operative mechanism. This analysis was not pre-specified in our pre-analysis plan; we treat it as exploratory.

A key assumption underlying the interpretation of treatment effects on spousal involvement is that individuals have full information about their partner's preferences. If this assumption holds, then observed behaviour around spousal involvement reflects the causal effect of treatment. However, a common finding in the intrahousehold literature is that spouses often do *not* have full information regarding their spouse's preferences (Afzal et al., 2022; D'Exelle and Ringdal, 2022; Tagat, Kapoor and Kulkarni, 2024). Instead, spouses form beliefs—which are often inaccurate—around what they expect their spouse to prefer. In such contexts, behaviour may be driven not only by actual preference differences, but also by *expected* spousal preference differences.

In our sample, beliefs were elicited prior to Decision Blocks 2 and 3 (Appendix B). From these, we construct four measures. First, *Beliefs* is a binary indicator equal to one if the respondent believes their spouse selected swimming lessons, and zero otherwise. Second, we construct an indicator for expected spousal preference differences (*ESPD*), equal to one when respondents expect their spouse to have chosen a different option, zero otherwise. Third, *ESPD no swim* captures the specific case in which the respondent chooses swimming lessons but expects their spouse not to do so. Fourth, we construct an indicator for incorrect beliefs (*ICBs*), equal to one when a respondent's belief about their spouse's first choice differs from their spouse's actual first choice, zero otherwise.

Descriptive statistics reveal substantial systematic variation in beliefs across treatments (Appendix D). Both wives and husbands become increasingly likely to believe that their spouse selected swimming lessons

as the decision domain shifts from WS (T1) to CS (T4). For example, the proportion of wives holding this belief rises from 9% in T1 to 43% in T4, while the corresponding figures for husbands increase from 8% to 37% (Table D.2). Additionally, beliefs are statistically significantly different for all treatment comparisons, except T1–T2 for wives, and T2–T3 for husbands.

Despite this treatment responsiveness, beliefs remain markedly inaccurate. Expected spousal preference differences are relatively rare, with mean ESPD values below 20% across all treatments, against actual preference differences occurring in approximately 40% of couples on average—rising from around 30% in T1–T2 to roughly 44% in T3–T4 (Table D.2). Consistent with this, *ICBs* show that misperceptions are relatively limited in T1–T2 but increase substantially in T3–T4, where roughly one-third to nearly one-half of respondents hold incorrect expectations about their spouse's choice (Table D.2). Both husbands and wives substantially underestimate the likelihood of disagreement, with no statistically significant gender differences in the prevalence of ESPDs.

We therefore test whether treatment-induced variation in ESPDs predicts spousal involvement behaviour according to the empirical strategy for a mediation analysis as specified by equation 2.6. We restrict this analysis to deferral, as anticipated disagreement has a clear directional prediction for deferral (conflict avoidance motivates relinquishing authority) but not for consultation, where the direction is theoretically ambiguous. Figure 2.3 depicts the causal structure under investigation.

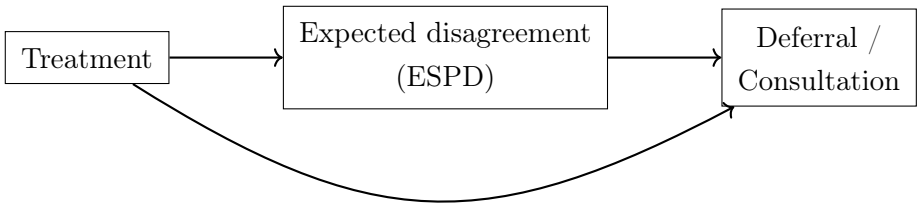


Figure 2.3: Beliefs as a mediating mechanism

Table 2.5 examines whether treatment effects on spousal involvement behaviour operate through changes in expected spousal preference differences. For ease of reference, column (1) reports the same results from Table 2.4. Column (2) reports treatment effects on ESPDs: while treatment variation does not significantly affect perceived disagreement among wives, husbands become more likely to expect disagreement between certain treatment pairs, particularly T2–T4. Column (3) introduces ESPDs into the deferral regression. For husbands, perceived disagreement is positively associated with deferral behaviour, and the inclusion of ESPDs slightly attenuates treatment coefficients, consistent with a belief-mediated mechanism. In contrast, ESPDs are not significantly associated with wives' deferral behaviour, and treatment coefficients remain largely unchanged.

Taken together, these results indicate that beliefs about spousal preferences do not substantially mediate the relationship between decision domain and spousal involvement. For wives, neither the pathway from treatment to expected disagreement nor the pathway from expected disagreement to deferral is statistically supported; treatment effects on wives' deferral remain unchanged when ESPDs are introduced. Yet despite largely perceiving preference alignment, wives' deferral varies significantly across domains. For husbands, perceived disagreement is positively associated with deferral and modestly attenuates treatment coefficients, consistent with partial mediation; however, a substantial direct effect of domain on deferral persists. Across both spouses, therefore, the primary driver of spousal involvement is the decision-making domain itself—specifically, normative expectations about who holds decision authority in wife's versus joint domains—rather than strategic responses to anticipated conflict. This finding strengthens the interpretation that domain-

Table 2.5: Mediating effects of expected spousal preference differences on deferral

	Defer (total)		ESPDs		Defer (direct)	
	(1)		(2)		(3)	
	Wives	Husbands	Wives	Husbands	Wives	Husbands
T1: Kanga	0.108*	0.012	0.024	0.054	0.106*	0.006
	(0.061)	(0.053)	(0.033)	(0.043)	(0.061)	(0.053)
T3: Books	0.169***	-0.066	-0.006	0.102**	0.169***	-0.077
	(0.052)	(0.052)	(0.033)	(0.041)	(0.050)	(0.053)
T4: Rice	0.084**	-0.127**	0.042	0.090**	0.080*	-0.136**
	(0.043)	(0.055)	(0.038)	(0.042)	(0.042)	(0.056)
ESPDs					0.091	0.103**
					(0.058)	(0.051)

Notes: Seemingly unrelated regression with the following dependent variables. Models (1) and (3): defer in Decision Block 2 where the choice to defer decision-making to the spouse takes the value 1, and the choice to make the decision alone takes the value 0; Model (2): ESPDs which takes the value 1 if the respondent believes their spouse's first choice differs from their own first choice, zero otherwise. The base treatment is 'T2 - WS/Soap'. Wife and husband regressions have 664 observations each. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

based norms, not belief-updating, are the mechanism through which treatment variation shapes spousal involvement.

3.5. Heterogeneity analysis

We examine heterogeneity in treatment effects on spousal involvement through two moderating variables: conflict aversion and bargaining power, the latter proxied by the share of child-related costs borne by the respondent[12]. Both become salient in Decision Blocks 2 and 3, where the possibility of disagreement is foregrounded: spousal involvement becomes common knowledge, and the beliefs elicitation process prompts individuals to consider whether their initial choice differs from their spouse's. The extent to which individuals are conflict averse, or hold greater bargaining power within the household, may therefore shape whether they choose to defer or consult.[13]

Table 2.6 presents treatment effects interacted with conflict aversion and share of child costs borne by the respondent, for deferral (panel 1) and consultation (panel 2). From panel 1, deferral is lower in T4 than T2 for wives who are not conflict averse, and this difference attenuates as conflict aversion increases (column 1). The positive treatment effects of deferral in T3 and T4 relative to T1 are driven by wives who bear no share of child costs; as the wife's contribution to child costs increases, deferral in the joint domain decreases (column 3), consistent with greater bargaining power reducing deference. From panel 2, the negative treatment effect of T1 on husbands' consultation relative to T2 is concentrated among those with low conflict aversion and attenuates as conflict aversion increases (column 2, positive and significant T1 × Var coefficient).

Result 3: *Conflict aversion and bargaining power moderate treatment effects on spousal involvement. For wives, reduced deferral in T4 relative to T2 is concentrated among those with low conflict aversion; greater contributions to child costs attenuate deferral in the joint domain, consistent with higher bargaining power reducing deference. For husbands, reduced consultation in T1 relative to T2 is driven by those with low conflict aversion, and attenuates as conflict aversion increases.*

[12]These heterogeneity dimensions were pre-specified in the pre-analysis plan (Section 5.5), though no directional hypotheses were stated. The analyses are therefore pre-specified rather than purely exploratory, but should not be interpreted as confirmatory tests.

[13]This mechanism is weaker in Decision Block 1, where individuals retain plausible deniability regarding spousal involvement and choices are made without the possibility of observing or influencing their spouse's decision within the task.

Table 2.6: Heterogeneity in treatment effects on spousal involvement: conflict aversion and bargaining power

	Defer			
	Conflict aversion		Child cost contributions	
	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	-0.405	-0.136	0.148	-0.173
	(0.327)	(0.323)	(0.118)	(0.210)
T3 - CS/Books	-0.364	-0.112	0.361***	-0.244
	(0.457)	(0.398)	(0.129)	(0.170)
T4 - CS/Rice	-0.900**	-0.160	0.330***	-0.301
	(0.399)	(0.327)	(0.111)	(0.263)
Var	-0.011	0.012	0.003	-0.002
	(0.012)	(0.010)	(0.002)	(0.002)
T1 - WS/Kanga × Var	0.021	0.007	-0.001	0.002
	(0.014)	(0.013)	(0.002)	(0.003)
T3 - CS/Books × Var	0.022	0.002	-0.005*	0.002
	(0.019)	(0.016)	(0.003)	(0.002)
T4 - CS/Rice × Var	0.041**	0.002	-0.006**	0.002
	(0.016)	(0.014)	(0.002)	(0.003)
	Consult			
	Conflict aversion		Child cost contributions	
	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	-0.383	-0.729***	0.060	-0.159
	(0.322)	(0.234)	(0.087)	(0.120)
T3 - CS/Books	-0.091	-0.278	0.090	-0.407**
	(0.351)	(0.175)	(0.072)	(0.185)
T4 - CS/Rice	-0.216	-0.227	0.060	-0.194
	(0.259)	(0.277)	(0.082)	(0.202)
Var	0.000	0.002	-0.000	-0.001
	(0.010)	(0.007)	(0.001)	(0.002)
T1 - WS/Kanga × Var	0.016	0.028***	-0.001	0.001
	(0.013)	(0.009)	(0.002)	(0.001)
T3 - CS/Books × Var	0.008	0.012*	0.000	0.005**
	(0.014)	(0.007)	(0.002)	(0.002)
T4 - CS/Rice × Var	0.011	0.010	-0.001	0.002
	(0.011)	(0.011)	(0.002)	(0.002)

Notes: Seemingly unrelated regression with the following dependent variables: Panel 1 uses 'Defer' where the choice in Decision Block 2 to defer decision-making to the spouse takes the value one, and the choice to make the decision alone takes the value zero; Panel 2 uses 'Consult' where the willingness to consult with the spouse in Decision Block 3 takes the value one (i.e., if they are open to either hearing their spouse's first choice or sharing their own first choice with their spouse), zero if they are closed to either hearing or sharing. The base treatment is 'T2 - WS/Soap'. Wife and husband regressions have 664 observations each. All models report treatment effects where treatment dummies are interacted with 'Var', where 'Var' is one of two variables. The first is 'Conflict aversion' which is the sum of the respondent's answers to a set of six Likert-scale questions on their attitudes to conflict, with 6=lowest conflict aversion score and 30=highest conflict aversion score. The second is the '% of child costs covered' which captures respondents' estimate between 0-100 of how much they contribute to child-related costs. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

To shed further light on the psychological motives underlying these channels, Appendix E (Table E.2) explores the interactive effects between treatment assignment and baseline beliefs. Because ESPDs are themselves shifted by the treatments (as shown in Section 3.4), these interaction models are strictly exploratory and supplementary to our main mediation analysis. Nevertheless, the pattern of heterogeneity is highly consistent with conflict-avoidance motives. Notably, wives' deferral in T3 is substantially higher among those who anticipate disagreement, while husbands' reduced deferral in T4 is concentrated among those who expect preference alignment. These results do not contradict the mediation analysis in Section 3.4: the mediation question is whether the treatment induces ESPDs, which then drive deferral; the interaction question is

whether wives or husbands with high ESPDs respond differently to treatment. For wives, the treatment does not reliably induce ESPDs, so there is no mediated pathway overall—but when disagreement is anticipated, it still predicts deferral, suggesting conflict-avoidance as a latent motive that the treatment does not systematically activate. Crucially, accounting for these belief dynamics does not displace the direct effect of the decision domains established in Section 3.3, reinforcing that domain-specific norms and subjective beliefs operate in tandem.

3.6. Uptake of swimming lessons

Our final research question asks whether it matters that one spouse has autonomy over decision-making for the final uptake of swimming lessons. To answer this, we compare final uptake across the three decision blocks by the gender of the first mover. Given that our treatments were not designed to identify causal effects on uptake directly, comparisons across decision blocks are exploratory.

Table 2.7 records the percentage of swimming lessons taken up in each treatment under each decision block. In Decision Block 1, this is simply the proportion of husbands and wives who choose swimming lessons as their first choice. In Decision Block 2, the recorded uptake represents cases where swimming is the first choice of the individual if they chose not to defer, and cases where swimming is the first choice of the *spouse* if the individual chose to defer. In Decision Block 3, uptake represents cases where swimming is the first choice of the individual if they chose not to consult, and cases where swimming is the choice of the *spouse* if they chose to consult *and* switch.

Table 2.7: Uptake of swimming lessons by decision block

	Decision block 1			Decision block 2			Decision block 3		
Variable	(1) Husbands	(2) Wives	(1)-(2) Pairwise t-test	(3) Husbands	(4) Wives	(3)-(4) Pairwise t-test	(5) Husbands	(6) Wives	(5)-(6) Pairwise t-test
T1	0.211	0.163	0.048	0.175	0.163	0.012	0.205	0.145	0.060
	(0.409)	(0.370)		(0.381)	(0.370)		(0.405)	(0.353)	
T2	0.229	0.157	0.072*	0.181	0.163	0.018	0.187	0.151	0.036
	(0.421)	(0.365)		(0.386)	(0.370)		(0.391)	(0.359)	
T3	0.319	0.289	0.030	0.295	0.241	0.054	0.295	0.259	0.036
	(0.468)	(0.455)		(0.458)	(0.429)		(0.458)	(0.439)	
T4	0.572	0.440	0.133**	0.512	0.452	0.060	0.536	0.494	0.042
	(0.496)	(0.498)		(0.501)	(0.499)		(0.500)	(0.501)	
N	166	166	332	166	166	332	166	166	332

Notes: In Decision Block 1, uptake of swimming corresponds to the swimming preferences of 166 husbands and 166 wives in each treatment. In Decision Block 2, column 3 combines cases where swimming is the first choice of the husband if he chose not to defer, and where swimming is the first choice of the *wife* if he chose to defer decision-making. In other words, column 3 is the end result of deferral on swimming uptake of the husband as first-mover. Similarly, column 4 combines cases where swimming is the first choice of the wife if she chose not to defer, and where swimming is the first choice of the *husband* if she chose to defer decision-making. In other words, column 4 is the end result of deferral on swimming uptake of the wife as first-mover. In Decision Block 3, column 5 combines cases where swimming is the first choice of the husband if he chose not to switch after consultation, and where swimming is the first choice of the *wife* if he chose to switch after consultation. In other words, column 5 is the end result of consultation on swimming uptake of the husband as first-mover. Similarly, column 6 combines cases where swimming is the first choice of the wife if she chose not to switch after consultation, and where swimming is the first choice of the *husband* if she chose to switch after consultation. In other words, column 6 is the end result of consultation on swimming uptake of the wife as first-mover. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A few interesting results stand out. In all treatments under Decision Block 1, swimming uptake is higher amongst husbands (column 1) than amongst wives (column 2). Moreover, T4 records the highest proportion of swimming uptake in Decision Block 1 for both spouses, *and* the largest spousal preference difference in swimming of any treatment (husbands are 13 percentage points more likely to choose (child) swimming lessons). However, when we account for deferral choices, these differences disappear, with none

of the treatments indicating statistically significant differences in uptake between husbands as first movers (column 3) and wives as first movers (column 4) for any treatments. It is also worth noting that the *absolute* levels of uptake are lower than in Decision Block 1 in both columns, and across all treatments.

When we allow for switching to take place in Decision Block 3, once again, uptake differences get smaller to the point that there are no statistical differences between spouses in any treatment. It is interesting to note that similar to Decision Block 1, uptake is higher where the husband is first mover (column 5) than where the wife is (column 6). Given the lower switching rates of husbands observed in Table D.3, this uptake is likely driven by husbands' first choices. Still, since some switching takes place, the absolute levels of uptake are lower for husbands in the final allocation than the initial reflection of choices.

To examine whether *retaining autonomy* matters for uptake, we present Table H.1 (Appendix H), which uses the same approach as Table 2.7, but restricts the samples in Decision Blocks 2 and 3 to those who retained autonomy, i.e., those that chose not to defer and not to consult. In doing so, we note that the differences in uptake between spouses are even more pronounced than in Decision Block 1: from those who do not defer, husbands are 10, 16, and 22 percentage points more likely than wives to choose swimming lessons in T2, T3, and T4, respectively. Given the very low samples of husbands and wives who do not consult in each treatment, we do not comment on uptake differences between spouses in Decision Block 3.

To examine uptake among those who did not retain autonomy, we restrict the samples in Decision Blocks 2 and 3 to those who did not retain autonomy, i.e., those that chose to defer and to consult. Table H.2 (Appendix H) presents the results. Focusing on this subsample, we note that the gap in uptake now reduces between spouses, with the highest uptake recorded in column 4 for wives (52%). This is driven by the high swimming preferences of husbands in Decision Block 1, combined with the high deferral rates of wives as reported in Table D.3. In fact, given the preferences of husbands, swimming uptake is now higher in column 4 (wives) for all treatments except T1. Last, as with Table 2.7, we again see that differences in uptake converge to non-significant differences in Decision Block 3, for all treatments.

4. Discussion

This experiment was designed to answer three questions: how much autonomy spouses retain in household decisions, what shapes the decision to involve a spouse, and whether retaining or relinquishing autonomy affects the final uptake of swimming lessons. Across the three decision blocks, a consistent picture emerges: domain-based norms, rather than anticipated conflict about preferences, are the primary driver of spousal involvement; and while gender gaps in initial swimming preferences are substantial, they narrow considerably once the spousal involvement process is endogenised. We discuss each in turn. We begin with spousal preferences for swimming lessons as revealed in Decision Block 1, then turn to the determinants of spousal involvement in Decision Blocks 2 and 3, and finally examine what these processes imply for the uptake of swimming lessons.

4.1. Spousal preferences for swimming lessons

In Decision Block 1, both spouses consistently prefer child swimming lessons (CS) over wife swimming lessons (WS) across all treatments. Within decision domains, treatment variation generates no statistically significant spousal preference differences, which means we are unable to confirm hypothesis 2.1. Still, when CS is held fixed, uptake is higher when the paired good benefits the household (rice) rather than the child (books), suggesting that goods which compete directly for child welfare crowd out demand for CS. Exploratory cross-domain comparisons reveal both higher CS uptake and larger spousal preference differences. One interpretation that unifies these patterns is that the two types of swimming hold different implications for

the household: CS may be viewed as an investment opportunity—more valuable than books or rice—whilst WS is possibly viewed more as an opportunity cost of women's time, rendering it less appealing than clothes or soap. This asymmetry resonates with evidence that women in low-income settings systematically undervalue investments in their own human capital relative to those for their children (Borgonovi, Seitz and Vogel, 2022; Duflo and Udry, 2004).

Beyond treatment variation, individuals appear to anticipate their spouse's preferences even in the individual block (Table E.1, Appendix E). Both wives and husbands who believe their spouse chose swimming are themselves more likely to do so, and some treatment effects—notably, wives' higher CS uptake in T3 relative to T2—are driven by wives who believe their husbands also chose swimming. This co-movement between choices and spousal beliefs may reflect a desire to minimise anticipated conflict, or simply that aligned preferences and beliefs tend to co-occur within couples. What the pattern suggests, however, is that the spousal involvement process does not begin cleanly at Decision Block 2: anticipated preferences already shape individual choices, foreshadowing the involvement dynamics we examine next.

4.2. Spousal involvement in decision-making

Revisiting the choices made by individuals on whether to defer decisions to, or consult with, their spouse (presented in Table 2.4), the most striking pattern across Decision Blocks 2 and 3 is the gender asymmetry in deferral. In the T2–T4 comparison alone, husbands become 12.7 percentage points *less* likely to defer while wives become 8 percentage points *more* likely. These opposing directions are consistent with an interpretation that both spouses share a normative expectation that husbands hold primary decision authority in joint domains. The consultation results reinforce this reading: while both spouses show a high baseline willingness to consult (over 80% across all treatments, Table D.3), wives are more likely to consult in T3 than T2, T1 or T4. For husbands, an interesting result emerges in consultation: within the wife's domain, husbands are less willing to consult when Good B benefits the wife individually (T1) than when it benefits the household (T2). This is consistent with husbands asserting decision authority even over goods that primarily affect their spouse, a pattern also noted in Abbink, Islam and Nguyen (2020). This also echoes the concern that the absence of consultation can reflect a husband deciding for, rather than with, his wife, a form of control that is observationally distinct from deferral but no less consequential for women's effective agency (Arugay et al., 2024).

The mediation analysis in Section 3.4 sheds further light on what drives these patterns. Although treatment assignment shifts spouses' beliefs about likely disagreement (ESPDs), these beliefs do not substantially mediate deferral behaviour for wives—neither the treatment-to-ESPD link nor the ESPD-to-deferral link is statistically strong enough to constitute a mediated pathway. For husbands, partial mediation is present: perceived disagreement is positively associated with deferral and modestly attenuates treatment coefficients, consistent with husbands partly deferring to avoid anticipated conflict. Yet, a substantial direct effect of decision domain on deferral persists for both spouses. This points to domain-specific norms—rather than strategic belief-updating—as the primary driver of spousal involvement.

Supplementary interaction analyses in Appendix E (Table E.2) are broadly consistent with this reading. Among wives who do anticipate disagreement, deferral in T3 is substantially elevated; among husbands, reduced deferral in T4 is concentrated among those expecting preference alignment. Both patterns are consistent with conflict-avoidance as a motivation. However, since the treatment does not reliably induce ESPDs for wives, this motive is not systematically activated. These belief dynamics operate against a backdrop of systematic underestimation: mean ESPDs remain below 20% across all treatments, substantially below the actual rate of preference disagreement, which rises from around 30% of couples in T1-T2 to roughly 44% in T3–T4 (Table D.2, Appendix D). The deferral and consultation rates we observe therefore represent a lower bound on what would obtain if spouses held correct beliefs. These patterns also resonate with the

literature documenting an intrinsic value in holding decision rights (Ertac, Gumren and Gurdal, 2020; Falk and Kosfeld, 2006; Ferreira, Hanaki and Tarroux, 2020). The fact that deferral and consultation are lowest for wives in their own personal domain suggests that wives' reluctance to cede control in that domain reflects a genuine valuation of decision authority, not merely a response to anticipated outcomes.

4.3. The uptake of swimming lessons

Turning to final uptake of swimming lessons across the decision blocks, several patterns surfaced. In individual decisions (Decision Block 1), husbands selected swimming lessons at higher rates than wives in every treatment, with the largest gender gap in T4 where child swimming lessons were paired with rice. Yet, these differences largely dissipated once we allowed spouses to defer decision-making (Decision Block 2): uptake converged and there were no statistically significant differences in outcomes between husband-first and wife-first cases across all treatments (Table 2.7, columns 3 and 4). Allowing consultation and switching (Decision Block 3) produced a similar convergence (Table 2.7, columns 5 and 6). This reflects the fact that autonomy by way of first mover seems to matter less once spousal involvement becomes an option.

Restricting the analysis to those who retain autonomy (i.e., do not defer or consult), Table H.1 provides an important counterpoint. Among this subsample, husbands are substantially more likely to choose swimming lessons than wives. This suggests that where spouses insist on making decisions independently, their gendered preferences manifest strongly in final uptake. This is consistent with (Arugay et al., 2024), who find that the ability to make personal decisions if desired is more strongly associated with agency than participation in decision-making per se.

In contrast, among those who relinquish autonomy (i.e., defer or consult), Table H.2 shows that spousal preference differences narrow, and uptake patterns shift in favour of husbands because wives more frequently defer and switch (after consulting) to their husbands' preferences for swimming. This pattern of wives deferring and accommodating their husbands' preferences is consistent with Bakhtiar et al. (2024), who find that wives defer in 67–68% of split-choice cases in Nigeria, compared to only 20–26% of husbands, and accommodate their husbands' communicated preferences far more than the reverse—a pattern they attribute to social norms rather than instrumental concerns.

Taken together, gender differences in uptake matter only insofar as spouses retain control; once they relinquish decision-making power or engage in consultative processes that involve switching, the identity of the first mover ceases to matter for final uptake.

5. Conclusion

This paper examines whether, and how, spousal involvement in household decision-making varies with the decision domain, using a field experiment with 664 married couples in Zanzibar. The results show that domain shapes both the direction and the mechanism of spousal involvement: wives defer more, and husbands less, in the joint domain, with wives' deferral driven by role-based norms rather than anticipated disagreement, and the endogenous process ultimately narrowing initial preference gaps in final swimming lesson uptake.

It is worth discussing these findings in relation to the growing body of work that interrogates the meaning of autonomous decision-making in households, and how it relates to women's empowerment (Afzal et al., 2016, 2022; Arugay et al., 2024; Bakhtiar et al., 2024; Donald et al., 2020; Duflo, 2012; Dykstra, Exley and Niederle, 2022; Edmonds, Feigenberg and Leight, 2023; Hillesland et al., 2025; Jayachandran and Voena, 2025; Lecoutere and Wuyts, 2021; Mishra and Tripathi, 2011). Consistent with studies that use similar experimental approaches (e.g., Bakhtiar et al., 2024), we find that wives are more likely than husbands to defer or, after consulting, to switch, to their spouse's preferred choice. This pattern aligns with broader

evidence that husbands' preferences carry greater weight in joint decision-making (Alem, Hassen and Köhlin, 2023; Carlsson et al., 2012, 2013). However, whilst scholars link women's lower autonomy, bargaining power or influence in decision-making to lower levels of women's empowerment (Jayachandran and Voena, 2025; Munro, 2018), and in many instances this is indeed the case, there are two important nuances.

For one, autonomy does not only manifest through decision-making (Arugay et al., 2024; Donald et al., 2020). Although 'being involved' in making decisions is often used as a proxy for agency or autonomy and hence women's empowerment, Arugay et al. (2024) provide evidence that being a decision-maker is only weakly related with agency, and instead the ability to make *personal* decisions if desired is more strongly associated with agency. This is consistent with our finding that deferral and consultation is *lowest* for women in their own personal domain, which indicates a meaningful reflection of autonomy in our sample. It is also in line with the literature documenting an intrinsic value in holding the right to decide (Ertac, Gumren and Gurdal, 2020; Falk and Kosfeld, 2006; Ferreira, Hanaki and Tarroux, 2020).

For another, making decisions is not always a reflection of agency. As Hillesland et al. (2025) argue, agency may also lie in *choosing not to decide*. In their framework, deferral can reflect a deliberate and meaningful exercise of agency, particularly in settings where spouses specialise across decision domains. Our findings resonate with this interpretation. If children fall in the decision domain of the husbands, and child swimming lessons more so, given husbands' greater competence in swimming relative to wives, deferral by wives to husbands may simply be motivated by a judgement that their husband is better placed to choose the welfare-maximising option (instead of lower bargaining power). Indeed, the fact that uptake of swimming is *higher* where wives defer to their husbands indicates that deciding not to decide leads to a better outcome for the children. It would therefore be misleading to conclude that empowering women through autonomous decision-making leads to worse household outcomes in our sample.

One design consideration that future work might address is that the decision domain and identity of the swimming beneficiary are collinear across treatment arms. That is, wife's domain treatments always pair wife swimming with wife-benefiting goods, and joint-domain treatments always pair child swimming with household or child-benefiting goods. As a result, any effect attributed to 'domain' could equally be attributed to 'beneficiary', and vice versa. Future work might employ a $2 \times 2 \times 2$ design which independently varies (i) investment beneficiary (wife vs. child), (ii) the decision-making domain (wife's vs. joint), and (iii) the paired consumption good's beneficiary, which would allow separate identification of domain and beneficiary effects on spousal involvement.

References

Abbink, Klaus, Asad Islam, and Chau Nguyen. 2020. "Whose voice matters? An experimental examination of gender bias in intra-household decision-making." *Journal of Economic Behavior & Organization*, 176: 337–352.

Afzal, Uzma, Giovanna d'Adda, Marcel Fafchamps, and Farah Said. 2016. "Gender and agency within the household: Experimental evidence from Pakistan." *CEPR Discussion Paper No. DP11464*.

Afzal, Uzma, Giovanna d'Adda, Marcel Fafchamps, and Farah Said. 2022. "Intrahousehold Consumption Allocation and Demand for Agency: A Triple Experimental Investigation." *American Economic Journal: Applied Economics*, 14(3): 400–444.

Alem, Yonas, Sied Hassen, and Gunnar Köhlin. 2023. "Decision-making within the household: The role of division of labor and differences in preferences." *Journal of Economic Behavior & Organization*, 207: 511–528.

Anuwong, Anusara, and Alistair Munro. 2025. "Better Together? Evidence On Information Flows and the Impact of Debt Management Training from A Lab-In-The-Field Experiment with Married Couples." Available at SSRN: https://ssrn.com/abstract=5462114 or http://dx.doi.org/10.2139/ssrn.5462114.

Armand, Alex, Orazio Attanasio, Pedro Carneiro, and Valérie Lechene. 2020. "The effect of gender-targeted conditional cash transfers on household expenditures: Evidence from a randomized experiment." *The Economic Journal*, 130(631): 1875–1897.

Arugay, Aries, Aletheia Amalia Donald, Forest Brach Jarvis, Hillary C Johnson, and Aletheia Valenciano. 2024. "When does decision-making reflect agency? Evidence from the rural Philippines." The World Bank.

Ashraf, Nava. 2009. "Spousal control and intra-household decision making: An experimental study in the Philippines." *American Economic Review*, 99(4): 1245–1277.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2010. "Female empowerment: Impact of a commitment savings product in the Philippines." *World development*, 38(3): 333–344.

Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household bargaining and excess fertility: An experimental study in Zambia." *American Economic Review*, 104(7): 2210–37.

Bakhtiar, M Mehrab, Marcel Fafchamps, Markus Goldstein, Kenneth L Leonard, and Sreelakshmi Papineni. 2024. "To Defer or to Differ: Experimental Evidence on the Role of Cash Transfers in Nigerian Couples' Decision–Making." *The Economic Journal*.

Borgonovi, Francesca, Helke Seitz, and Irina Vogel. 2022. "Swimming skills around the world: evidence on inequalities in life skills across and within countries."

Browning, Martin, Pierre-André Chiappori, and Yoram Weiss. 2014. *Economics of the Family*. Cambridge University Press.

Brugere, Cecile, Flower E. Msuya, Narriman Jiddawi, Betty Nyonje, and Ritha Maly. 2020. "Can innovation empower? Reflections on introducing tubular nets to women seaweed farmers in Zanzibar." *Gender, Technology and Development*, 24(1): 89–109.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics*, 90(3): 414–427.

Carlsson, Fredrik, Haoran He, Peter Martinsson, Ping Qin, and Matthias Sutter. 2012. "Household decision making in rural China: Using experiments to estimate the influences of spouses." *Journal of Economic Behavior & Organization*, 84(2): 525–536.

Carlsson, Fredrik, Peter Martinsson, Ping Qin, and Matthias Sutter. 2013. "The influence of spouses on household decision making under risk: an experiment in rural China." *Experimental Economics*, 16(3): 383–401.

de Jong Cleyndert, Georgia, Rebecca Newman, Cecile Brugere, Aida Cuni-Sanchez, and Robert Marchant. 2021. "Adaptation of Seaweed Farmers in Zanzibar to the Impacts of Climate Change." In *African Handbook of Climate Change Adaptation.* , ed. Nicholas Oguge, Desalegn Ayal, Lydia Adeleke and Izael da Silva, 3–28. Cham:Springer International Publishing.

de la Torre-Castro, Maricela, Sara Fröcklin, Sanna Börjesson, Janine Okupnik, and Narriman S. Jiddawi. 2017. "Gender analysis for better coastal management – Increasing our understanding of social-ecological seascapes." *Marine Policy*, 83: 62–74.

Dizon-Ross, Rebecca, and Seema Jayachandran. 2023. "Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation." *American Economic Review: Insights*, 5(4): 445–459.

Donald, Aletheia, Gayatri Koolwal, Jeannie Annan, Kathryn Falb, and Markus Goldstein. 2020. "Measuring Women's Agency." *Feminist Economics*, 26(3): 200–226.

Duflo, Esther. 2003. "Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa." *The World Bank Economic Review*, 17(1): 1–25.

Duflo, Esther. 2012. "Women empowerment and economic development." *Journal of Economic literature*, 50(4): 1051–79.

Duflo, Esther, and Christopher Udry. 2004. "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices." NBER Working Paper No. w10498, Available at SSRN: https://ssrn.com/abstract=552103.

Dykstra, Holly, Christine L Exley, and Muriel Niederle. 2022. "When Do Individuals Give Up Agency? The Role of Decision Avoidance."

D'Exelle, Ben, and Charlotte Ringdal. 2022. "Women's use of family planning services: An experiment on the husband's involvement." *Journal of Development Economics*, 158: 102915.

D'Exelle, Ben, and Charlotte Ringdal. 2025. "Who decides in the household when risk preferences conflict?" *Review of Economics of the Household.*

D'Exelle, Ben, Aurélia Lépine, Richard Bakyono, and Ludovic D. G. Tapsoba. 2023. "Fertility and polygyny: Experimental evidence from Burkina Faso." *Journal of Development Economics*, 164: 103134.

Edmonds, Eric, Ben Feigenberg, and Jessica Leight. 2023. "Advancing the Agency of Adolescent Girls." *The Review of Economics and Statistics*, 105(4): 852–866.

Ertac, Seda, Mert Gumren, and Mehmet Y Gurdal. 2020. "Demand for decision autonomy and the desire to avoid responsibility in risky environments: Experimental evidence." *Journal of Economic Psychology*, 77: 102200.

Falk, Armin, and Michael Kosfeld. 2006. "The hidden costs of control." *American Economic Review*, 96(5): 1611–1630.

Felter, Claire Elizabeth. 2015. "In Zanzibar, Swim School to Fight Drownings."

Ferreira, João V, Nobuyuki Hanaki, and Benoît Tarroux. 2020. "On the roots of the intrinsic value of decision rights: Experimental evidence." *Games and Economic Behavior*, 119: 110–122.

Ferreira, João V, Nobuyuki Hanaki, Fabrice Le Lec, Erik Schokkaert, and Benoît Tarroux. 2025. "Freedom counts: Cross-country empirical evidence." *European Economic Review*, 176: 105022.

Freundt, Jana, and Holger Herz. 2025. "A Cross-Country Study of Preferences for Choice Autonomy." CESifo Working Paper No. 12238.

Haddad, Lawrence, John Hoddinott, and Harold Alderman. 1997. "Intrahousehold resource allocation in developing countries: models, methods and policies." *International Food Policy Research Institute*.

Hillesland, Marya, Cheryl R Doss, Serena Masino, Martina Querejeta, Aletheia Amalia Donald, Greg Seymour, and Clare Clingain. 2025. "Deciding Not to Decide: When Is There Power in Not Deciding?" The World Bank. Policy Research Working Paper. Report Number WPS11065. Washington, D.C. : World Bank Group.

Jayachandran, Seema, and Alessandra Voena. 2025. "Women's Power in the Household."

Lecoutere, Els, and Eva Wuyts. 2021. "Confronting the Wall of Patriarchy: Does Participatory Intrahousehold Decision Making Empower Women in Agricultural Households?" *The Journal of Development Studies*, 57(6): 882–905. _eprint: https://doi.org/10.1080/00220388.2020.1849620.

Liaqat, Sundas, Aletheia Donald, Forest B. Jarvis, Elizaveta Perova, and Hillary C. Johnson. 2026. "Lost in interpretation: Why spouses disagree on who makes decisions." *Review of Economics of the Household*.

Lundberg, Shelly J, Robert A Pollak, and Terence J Wales. 1997. "Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit." *Journal of Human Resources*, 32(3): 463–480.

Makame, Makame Omar, Ali Rashid Hamad, Muhammad Suleiman Said, Alice Mushi, and Khadija Sharif. 2021. "Moving Seaweed Farms from Shallow to Deep Seawater to Cope with Warming and Diseases in Zanzibar. Current Socio-Economic and Cultural Barriers." *Journal of Sustainable Development*, 14(5): p29. Number: 5.

McCarthy, Aine Seitz. 2019. "Intimate partner violence and family planning decisions: Experimental evidence from rural Tanzania." *World Development*, 114: 156–174.

Mishra, Nripendra Kishore, and Tulika Tripathi. 2011. "Conceptualising Women's Agency, Autonomy and Empowerment." *Economic and Political Weekly*, 46(11): 58–65.

Munro, Alistair. 2018. "Intra-household experiments: A survey." *Journal of Economic Surveys*, 32(1): 134–175.

Ringdal, Charlotte, and Ingrid Hoem Sjursen. 2021. "Household bargaining and spending on children: Experimental evidence from Tanzania." *Economica*, 88(350): 430–455.

Royal National Lifeboat Institution. 2019. "Evidence, Perceptions and Knowledge of drowning risk in Zanzibar."

Schultz, T Paul. 2004. "School subsidies for the poor: evaluating the Mexican Progresa poverty program." *Journal of development Economics*, 74(1): 199–250.

Stacke, Sarah. 2017. "On 'Burkini Island' Muslim Girls Can Finally Learn to Swim." Section: Photography.

Tagat, Anirudh, Hansika Kapoor, and Savita Kulkarni. 2024. "Private lives: experimental evidence on information completeness in spousal preferences." *Journal of the Economic Science Association*, 10(1): 62–75.

Thomas, Duncan. 1990. "Intra-Household Resource Allocation: An Inferential Approach." *The Journal of Human Resources*, 25(4): 635–664.

Thomas, Duncan. 1994. "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height." *The Journal of Human Resources*, 29(4): 950–988.

Appendix

A. Balancing tests and sampling map

Table A.1: Balancing tests for treatment groups (women)

Variable	(1) T1: Kanga Mean/(SE)	(2) T2: Soap Mean/(SE)	(3) T3: Books Mean/(SE)	(4) T4: Rice Mean/(SE)	F-test for balance across all groups F-stat/P-value
Age	42.072	40.657	40.072	41.843	1.239
	(0.895)	(0.901)	(0.855)	(0.789)	0.295
Education (y/n)	0.837	0.855	0.843	0.904	1.486
	(0.029)	(0.027)	(0.028)	(0.023)	0.217
Swim skills (y/n)	0.072	0.030	0.042	0.042	1.017
	(0.020)	(0.013)	(0.016)	(0.016)	0.385
Risk averse (y/n)	0.747	0.759	0.765	0.723	0.298
	(0.034)	(0.033)	(0.033)	(0.035)	0.827
Income in a good month (TZS '000s')	78.849	78.886	87.066	80.422	0.162
	(9.472)	(8.426)	(10.003)	(8.965)	0.922
Income in a bad month (TZS '000s')	33.654	38.479	36.398	39.342	0.173
	(5.706)	(6.083)	(5.394)	(6.930)	0.915
Years lived in the village	32.313	32.277	31.319	32.813	0.236
	(1.400)	(1.312)	(1.277)	(1.293)	0.871
Years of marriage	19.524	18.837	18.283	19.422	0.369
	(1.085)	(1.011)	(0.925)	(0.849)	0.776
No. of sons	1.916	1.922	2.145	1.970	0.638
	(0.126)	(0.145)	(0.135)	(0.117)	0.591
No. of daughters	1.819	1.735	1.849	1.861	0.275
	(0.119)	(0.100)	(0.123)	(0.128)	0.843
% of child costs contributed	44.880	42.892	41.566	42.410	0.651
	(1.773)	(1.926)	(1.672)	(1.845)	0.583
% of personal costs contributed	56.867	55.904	55.000	58.373	0.468
	(2.034)	(2.172)	(2.112)	(2.101)	0.704
% of household costs contributed	39.398	38.072	37.289	39.699	0.429
	(1.693)	(1.739)	(1.706)	(1.777)	0.732
Number of observations	166	166	166	166	664

Table A.2: Balancing tests for treatment groups (men)

Variable	(1) T1: Kanga Mean/(SE)	(2) T2: Soap Mean/(SE)	(3) T3: Books Mean/(SE)	(4) T4: Rice Mean/(SE)	F-test for balance across all groups F-stat/P-value
Age	50.422	47.223	49.012	48.669	1.595
	(1.002)	(1.080)	(1.065)	(0.970)	0.189
Education (y/n)	0.867	0.873	0.910	0.892	0.627
	(0.026)	(0.026)	(0.022)	(0.024)	0.598
Swim skills (y/n)	0.795	0.765	0.783	0.753	0.332
	(0.031)	(0.033)	(0.032)	(0.034)	0.802
Risk averse (y/n)	0.319	0.259	0.253	0.271	0.713
	(0.036)	(0.034)	(0.034)	(0.035)	0.544
Income in a good month (TZS '000s')	443.542	441.609	427.228	288.976	4.293***
	(92.429)	(56.699)	(54.132)	(19.323)	0.005
Income in a bad month (TZS '000s')	177.488	156.398	171.241	127.440	1.519
	(42.879)	(26.044)	(22.091)	(10.626)	0.208
Years lived in the village	45.946	43.102	45.434	44.627	0.892
	(1.260)	(1.344)	(1.307)	(1.256)	0.445
Years of marriage	21.133	19.705	19.614	20.560	0.508
	(1.064)	(1.002)	(0.979)	(0.893)	0.677
No. of sons	1.916	1.922	2.145	1.970	0.638
	(0.126)	(0.145)	(0.135)	(0.117)	0.591
No. of daughters	1.819	1.735	1.849	1.861	0.275
	(0.119)	(0.100)	(0.123)	(0.128)	0.843
% of child costs contributed	80.000	79.578	82.530	80.482	0.841
	(1.475)	(1.384)	(1.437)	(1.270)	0.472
% of personal costs contributed	84.819	81.807	84.277	83.313	0.457
	(1.887)	(1.988)	(1.972)	(1.859)	0.712
% of household costs contributed	85.241	83.313	85.482	85.301	0.632
	(1.261)	(1.291)	(1.334)	(1.129)	0.595
Number of observations	166	166	166	166	664

Figure A.1: Map of Unguja showing sampled villages for Chapters 2 & 3

B. Experimental script

The following text is the experimental script used for this study. Instructions for the enumerator are between brackets. Text in bold was emphasised by enumerators.

[Introduce yourself using the following script.]

Hello! My name is [insert name]. Thank you very much for coming. On behalf of our research team, I would like to welcome you to today's session. This session is being conducted as part of our research into coastal communities in Zanzibar. I am part of a research team that is working with a University. We have invited you here, today, because we want to learn about how people in this area take decisions. You are going to be asked to take decisions that allow you to earn some money. The money that results from your decisions will be yours to keep. What you need to do will be explained fully. But first we want to make a few of things clear.

First of all, this is not our money. We belong to a university, and this money has been given to us for research. Second, participation is voluntary. You may still choose not to participate in the exercise. If you decide to not to participate, you will still be compensated for your time. If you participate, make sure that you listen carefully to us. It is important that you follow our instructions. Your interview will last around 45 minutes. For participating in this interview, you will receive TSH 4,000 for sure, in addition to any money you might win during the interview. I will now explain the procedures in more detail.

Today's interview will be in THREE PARTS

In PART 1 we would like to ask you some questions about you and your household. Please remember that there is no right or wrong answer, and your answers will be anonymised for any research purposes. This means that your answers will not be able to be personally linked to you or your household, and you will not be personally identifiable by anyone who uses this information for research. Therefore, we ask you to answer as truthfully as possible. After PART 1, we will move onto PART 2 and PART 3.

In PART 2 and 3, you will be asked to do two different activities. In one activity, I will ask you to make a series of THREE choices, which will be explained to you very clearly. Your spouse will be asked to make the same choices. The choices made here will determine the type of gift your household may receive. So your choices have **REAL** consequences. In another activity, I will demonstrate to you a game that you will be asked to complete **TWO** times. Your performance here will determine how much money you might win to take home.

This next part is very important to understand.

What you will win to take home will be RANDOMLY selected from PART 2 and PART 3. This means that for your payment **ONLY ONE** of the two activities will be implemented. At the end of the interview, our computer will randomly select ONE from the three choices you made, the three choices your spouse made, and the two tasks which you perform, to determine what you will take home. Therefore, **BOTH** activities are equally important because you do not know which one will be chosen for your final payment. Once you receive the final compensation, you can leave. Any money you earn will be paid out to you privately and confidentially **IMMEDIATELY** after the exercise is complete, with the use of Tigo Peza to a mobile number that you specify.

[*Explain PART II using the following script.*]

Now I will ask you to make **THREE SETS** of decisions. Each set of decisions involves a choice between two goods. Your spouse will also be asked the same **THREE SETS** of decisions. **ONE**

of the SIX decisions made by both of you will be selected at random, and may be given to your household at the end of the interview. Therefore, they are **ALL** equally important, and you must think carefully about what you are going to choose.

DECISION BLOCK 1: INDIVIDUAL

There are two goods for you to consider. Both these goods are REAL, and one of them could be yours at the end of the interview:

T1 - Kanga vs Wife Swimming

Text for wives:
One is a new kanga for you. The other is a voucher which gives you the chance to be considered for a local programme that offers swimming lessons. These swimming lessons would be conducted sometime in 2025, and will take place in the sea close to your village. They would be conducted by an organisation called the Panje Project, who are specialised in providing swimming lessons to women, and use female trainers. The voucher ensures the wife of this household is added to a list, which is used by the swimming programme to select its beneficiaries, depending on the resources they have available.

Between free swimming lessons for yourself, or the kanga, which would you choose?

Text for husbands:
One is a new kanga for your wife. The other is a voucher which gives your wife the chance to be considered for a local programme that offers swimming lessons. These swimming lessons would be conducted sometime in 2025, and will take place in the sea close to your village. They would be conducted by an organisation called the Panje Project, who are specialised in providing swimming lessons to women, and use female trainers. The voucher ensures the wife of this household is added to a list, which is used by the swimming programme to select its beneficiaries, depending on the resources they have available.

Between free swimming lessons for your wife, or the kanga, which would you choose?

T2 - Soap vs Wife Swimming

Text for wives:
One is a pack of 1KG of soap. The other is... [repeat text on wife swimming as above in T1.]

Between free swimming lessons for yourself, or 1KG of soap, which would you choose?

Text for husbands:
One is a pack of 1KG of soap. The other is.... [repeat text on wife swimming as above in T1.]

Between free swimming lessons for your wife, or 1KG of soap, which would you choose?

T3 - Books vs Child Swimming

Text for husbands and wives:
One is a pack of 6 exercise books for your child. The other is a voucher which gives one child in the household the chance to be considered for a local programme that offers swimming lessons. These swimming lessons would be conducted sometime in 2025, and will take place in the sea close to your village. They would be conducted by an organisation called the Panje Project, who are specialised in providing swimming lessons to children, and use same-gender trainers. The voucher ensures the child is added to a list, which is used by the swimming programme to select its beneficiaries, depending on the resources they have available.

Between free swimming lessons for your child or exercise books, which would you choose?

T4 - Rice vs Child Swimming

Text for husbands and wives:
One is a pack of 1KG of rice. The other is a voucher which gives one child in the household the chance to be considered for a local programme that offers swimming lessons. These swimming lessons would be conducted sometime in 2025, and will take place in the sea close to your village. They would be conducted by an organisation called the Panje Project, who are specialised in providing swimming lessons to children, and use same-gender trainers. The voucher ensures the child is added to a list, which is used by the swimming programme to select its beneficiaries, depending on the resources they have available.

Between free swimming lessons for your child or the 1KG of rice, which would you choose?

[*Did you tick the choice in the* ***first row*** *of your decision card?*]

BETWEEN BLOCKS: BELIEF ELICITATION

T1 - Kanga vs Wife Swimming

What do you think your spouse would choose if they had the choice between free swimming lessons for the wife of the household, or the kanga?

T2 - Soap vs Wife Swimming

What do you think your spouse would choose if they had the choice between free swimming lessons for the wife of the household, or the 1KG soap?

T3 - Books vs Child Swimming

What do you think your spouse would choose if they had a choice between free swimming lessons for your child, or exercise books?

T4 - Rice vs Child Swimming

What do you think your spouse would choose if they had the choice between free swimming lessons for your child, or the 1KG of rice?

DECISION BLOCK 2: DEFER

We are giving you the chance to let your spouse decide on your behalf. If you allow them to decide on your behalf, their choice will be chosen INSTEAD of your choice when we make the final payment selection. Would you prefer if we ask them to decide on your behalf, or do you want us to consider only your decision?

[*Did you tick the choice in the* ***second row*** *of your decision card?*]

BETWEEN BLOCKS: WILLINGNESS TO CONSULT ELICITATION

We are giving you the chance to share your choice with your spouse. If you say no, your choice will be kept private. Would you like us to share your choice?

[*Did you tick the choice in the* ***third row*** *of your decision card?*]

We are giving you the chance to hear your spouse's choice. If they share their choice, would you like to hear it?

[*Did you tick the choice in the* ***fourth row*** *of your decision card?*]

[You must now pause the interview, and find your enumerator partner to fill in the rest of the decision card.]

DECISION BLOCK 3: CONSULT (only for subgroup of respondents who agree to hear and whose spouses agree to share.)

[*Since your respondent has agreed to hear AND their spouse has agreed to share information, you and your enumerator partner must now fill in the last row of each other's decision cards.*]

[*Explain to the respondent that you will now read out what their spouse chose. Then read out the spouse's choice.*]

Same question for all treatments, adjusting wording for type of Good A and Good B:
You have the chance to either change your choice, or keep the same choice. What is your final answer between [Good A] and [Good B]?

C. Reconciling pre-analysis plan and chapter hypotheses

PAP Hypothesis #	Statement	Paper Hypothesis #	Notes
H1 (PAP)	SPDs differ between T1-T2 and T3-T4	H1 (paper)	Identical
H2 (PAP)	Deferral higher in joint domain	H2 (paper)	Identical
H3 (PAP)	Deferral higher with larger SPDs	H4 (paper)	Renumbered
H4 (PAP)	Consultation higher in joint domain	H3 (paper)	Renumbered
H5 (PAP)	Consultation higher with larger SPDs	H5 (paper)	Identical
H6 (PAP)	Gender differences in H2-H5	H6 (paper)	Identical

D. Summary statistics for key decision block variables

Here, we report the p-values of t-tests for statistical differences in key decision block variables (1) between husbands and wives (Table D.1), and (2) between treatments (Table D.2).

Table D.1: Summary statistics for decision block variables

Variable	(1) Male Mean/(SD)	(2) Female Mean/(SD)	(1)-(2) Pairwise t-test P-value
First choice	0.333	0.262	0.005***
	(0.472)	(0.440)	
Beliefs	0.190	0.244	0.017**
	(0.392)	(0.430)	
Defer	0.503	0.482	0.443
	(0.500)	(0.500)	
Consult	1.682	1.669	0.733
	(0.718)	(0.730)	
PDs	0.369	0.369	1.000
	(0.483)	(0.483)	
ESPDs	0.170	0.166	0.826
	(0.376)	(0.372)	
Number of observations	664	664	1328

Notes: 'First choice' which is a binary variable equal to one if the respondent chose swimming lessons in Decision Block 1, zero otherwise; 'Beliefs' which is a binary variable with value 1 if the respondent believes their spouse chose swimming lessons in Decision Block 1, zero otherwise; 'Defer' which is a dummy equal to 1 if the respondent chose to defer decision-making to their spouse in Decision Block 2, zero otherwise; and 'Consult' which is a dummy equal to 1 if the respondent was open to either sharing their first choice or hearing their spouse's first choice in Decision Block 3, zero otherwise; 'PD' i.e preference difference which is a binary variable equal to 1 if the respondent's first choice is different to their spouse's first choice, zero otherwise (note that PDs are the same values for husbands and wives as these are actual spousal preference differences and thus do not vary by couple); and 'ESPD' which is the expected spousal preference difference and is a dummy equal to 1 if the respondent's belief about their spouse's first choice is different to the respondent's actual first choice, zero otherwise. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: Descriptive statistics for choices and beliefs in decision blocks

	(1)	(2)	(3)	(4)	(1)-(2)	(1)-(3)	(1)-(4)	(2)-(3)	(2)-(4)	(3)-(4)
	T1: Kanga	T2: Soap	T3: Books	T4: Rice			Pairwise t-test			
Variable	Mean/(SD)	Mean/(SD)	Mean/(SD)	Mean/(SD)	P-value	P-value	P-value	P-value	P-value	P-value
First choice (w)	0.163	0.157	0.289	0.440	0.881	0.006***	0.000***	0.004***	0.000***	0.004***
	(0.370)	(0.365)	(0.455)	(0.498)						
First choice (h)	0.211	0.229	0.319	0.572	0.692	0.025**	0.000***	0.065*	0.000***	0.000***
	(0.409)	(0.421)	(0.468)	(0.496)						
Beliefs (w)	0.096	0.139	0.313	0.428	0.234	0.000***	0.000***	0.000***	0.000***	0.031**
	(0.296)	(0.347)	(0.465)	(0.496)						
Beliefs (h)	0.084	0.157	0.145	0.373	0.043**	0.085*	0.000***	0.760	0.000***	0.000***
	(0.279)	(0.365)	(0.353)	(0.485)						
ICBs (w)	0.271	0.271	0.440	0.446	1.000	0.001***	0.001***	0.001***	0.001***	0.912
	(0.446)	(0.446)	(0.498)	(0.499)						
ICBs (h)	0.247	0.241	0.325	0.416	0.899	0.115	0.001***	0.089*	0.001***	0.089*
	(0.433)	(0.429)	(0.470)	(0.494)						
PD (w)	0.277	0.313	0.440	0.446	0.472	0.002***	0.001***	0.017**	0.013**	0.912
	(0.449)	(0.465)	(0.498)	(0.499)						
PD (h)	0.277	0.313	0.440	0.446	0.472	0.002***	0.001***	0.017**	0.013**	0.912
	(0.449)	(0.465)	(0.498)	(0.499)						
PD no swim (w)	0.163	0.193	0.235	0.289	0.474	0.099*	0.006***	0.350	0.040**	0.263
	(0.370)	(0.396)	(0.425)	(0.455)						
PD no swim (h)	0.114	0.120	0.205	0.157	0.865	0.025**	0.263	0.037**	0.342	0.255
	(0.319)	(0.327)	(0.405)	(0.365)						
ESPD (w)	0.175	0.151	0.145	0.193	0.553	0.455	0.672	0.877	0.310	0.242
	(0.381)	(0.359)	(0.353)	(0.396)						
ESPD (h)	0.163	0.108	0.211	0.199	0.150	0.261	0.394	0.011**	0.022**	0.786
	(0.370)	(0.312)	(0.409)	(0.400)						
ESPD no swim (w)	0.120	0.084	0.060	0.102	0.279	0.056*	0.602	0.398	0.573	0.161
	(0.327)	(0.279)	(0.239)	(0.304)						
ESPD no swim (h)	0.145	0.090	0.193	0.199	0.126	0.242	0.191	0.007***	0.005***	0.890
	(0.353)	(0.288)	(0.396)	(0.400)						
Defer (w)	0.500	0.392	0.560	0.476	0.047**	0.273	0.662	0.002***	0.122	0.125
	(0.502)	(0.490)	(0.498)	(0.501)						
Defer (h)	0.560	0.548	0.482	0.422	0.826	0.154	0.011**	0.228	0.021**	0.271
	(0.498)	(0.499)	(0.501)	(0.495)						
Consult (y/n) (w)	0.819	0.807	0.910	0.843	0.779	0.016**	0.559	0.007***	0.387	0.067*
	(0.386)	(0.396)	(0.288)	(0.365)						
Consult (y/n) (h)	0.795	0.867	0.867	0.873	0.079*	0.079*	0.055*	1.000	0.871	0.871
	(0.405)	(0.340)	(0.340)	(0.333)						
N	166	166	166	166	332	332	332	332	332	332

Notes: Summary statistics of the following variables: 'First choice' which is a binary variable equal to one if the respondent chose swimming lessons in Decision Block 1, zero otherwise; 'Beliefs' which is a binary variable with value 1 if the respondent believes their spouse chose swimming lessons in Decision Block 1, zero otherwise; 'PD' i.e preference difference which is a binary variable equal to 1 if the respondent's first choice is different to their spouse's first choice, zero otherwise (note that PDs are the same values for husbands and wives as these are actual spousal preference differences and thus do not vary by couple); 'PD no swim' which is a binary variable equal to 1 if the respondent does not choose swimming and their spouse chooses swimming, zero otherwise; 'ESPD' which is the expected spousal preference difference and is a dummy equal to 1 if the respondent's belief about their spouse's first choice is different to the respondent's actual first choice, zero otherwise; 'ESPD no swim' which is binary variable equal to 1 if the respondent chooses swimming and believes their spouse does not choose swimming, zero otherwise; 'Defer' which is a dummy equal to 1 if the respondent chose to defer decision-making to their spouse in Decision Block 2, zero otherwise; and 'Consult (y/n)' which is a dummy equal to 1 if the respondent was open to either sharing their first choice or hearing their spouse's first choice in Decision Block 3, zero otherwise. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table D.3: Descriptive statistics on behaviour: defer, consult, switch

	Defer			Consult			Switch		
	(1)	(2)	(1)-(2)	(3)	(4)	(3)-(4)	(5)	(6)	(5)-(6)
Variable	Husbands	Wives	Pairwise t-test	Husbands	Wives	Pairwise t-test	Husbands	Wives	Pairwise t-test
Full sample	0.503	0.482	0.021	0.845	0.869	-0.024	0.095	0.135	-0.041*
	(0.500)	(0.500)		(0.356)	(0.362)		(0.293)	(0.342)	
N	664	664	1328	664	664	1328	465	458	923
T1	0.560	0.500	0.060	0.795	0.819	-0.024	0.105	0.107	-0.002
	(0.498)	(0.502)		(0.405)	(0.386)		(0.308)	(0.310)	
N	166	166	332	166	166	332	105	103	208
T2	0.548	0.392	0.157***	0.867	0.807	0.060	0.078	0.115	-0.037
	(0.499)	(0.490)		(0.340)	(0.396)		(0.269)	(0.320)	
N	166	166	332	166	166	332	116	113	229
T3	0.482	0.560	-0.078	0.867	0.910	-0.042	0.104	0.139	-0.035
	(0.501)	(0.498)		(0.340)	(0.288)		(0.306)	(0.348)	
N	166	166	332	166	166	332	125	122	247
T4	0.422	0.476	-0.054	0.873	0.843	0.030	0.092	0.175	-0.083*
	(0.495)	(0.501)		(0.333)	(0.365)		(0.291)	(0.382)	
N	166	166	332	166	166	332	119	120	239

Notes: Summary statistics (means and pairwise t-tests) are presented for three binary variables: 'Defer' (columns 1-3) which takes the value one if the individual defers decision-making to their spouse in Decision Block 2, zero if they do not; 'Consult' (columns 4-6) which takes the value one if the individual is willing to consult decision-making with their spouse in Decision Block 3, zero if they are not; and 'Switch' (columns 7-9) which takes the value one if the individual switches their initial choice in Decision Block 1 to their spouse's choice, zero if they do not. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

E. Beliefs interacted with treatments in all Decision Blocks

Tables E.1 and E.2 present supplementary interaction analyses for Decision Block 1 and Decision Blocks 2–3, respectively. Belief variables are defined as in Section 3.4: *Beliefs* equals one if the respondent believes their spouse chose swimming lessons; *ESPD* equals one when the respondent expects a different choice from their spouse; and *ESPD no swim* equals one when the respondent chose swimming but expects their spouse not to. Since ESPDs lie on the causal pathway from treatment to spousal involvement, we treat them primarily as mediators rather than moderators (Section 3.4); these interaction analyses are exploratory and were not pre-specified. Key findings from Table E.1 are discussed in Section 3.2; key findings from Table E.2 are discussed in Section 4.

Table E.1: Beliefs interacted with treatments: Decision Block 1

	(1) Var = belief		(2) Var = spouse's choice		(3) Var = spouse's belief	
	Wives	Husbands	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	0.036	0.051	-0.010	-0.035	0.036	0.004
	(0.034)	(0.038)	(0.051)	(0.038)	(0.042)	(0.034)
T3 - CS/Books	-0.010	0.118***	0.140**	0.095	0.132**	0.111*
	(0.036)	(0.046)	(0.065)	(0.059)	(0.053)	(0.064)
T4 - CS/Rice	0.081***	0.210***	0.195***	0.273***	0.239***	0.301***
	(0.031)	(0.048)	(0.068)	(0.069)	(0.052)	(0.067)
Var	0.425***	0.778***	0.048	0.054	0.105*	0.148*
	(0.106)	(0.078)	(0.055)	(0.074)	(0.062)	(0.082)
T1 - WS/Kanga × Var	-0.120	-0.151	0.081	0.099	-0.269***	-0.167
	(0.224)	(0.152)	(0.092)	(0.119)	(0.061)	(0.127)
T3 - CS/Books × Var	0.218*	-0.128*	-0.038	-0.041	0.010	-0.149
	(0.119)	(0.076)	(0.104)	(0.119)	(0.092)	(0.105)
T4 - CS/Rice × Var	0.186	-0.095	0.126	0.124	0.058	-0.001
	(0.132)	(0.085)	(0.109)	(0.120)	(0.096)	(0.117)
Predicted outcome under T2	0.201	0.255	0.162	0.235	0.160	0.245

Notes: Seemingly unrelated regressions with the dependent variable 'First choice' in Decision Block 1, equal to 1 if swimming lessons are chosen, 0 otherwise. 'T2 - WS/Soap' is the base treatment. All models report treatment effects where treatment dummies are interacted with 'Var'. In Model (1) 'Var' refers to 'Beliefs'—a binary variable which captures the individual's beliefs about spousal first choices, taking the value 1 if they believed the spouse chose swimming lessons, zero if they believe the spouse chose good B. In Model (2) 'Var' is the 'Spouse's first choice'—a binary variable with value 1 if the respondent's spouse's first choice is swimming, zero otherwise. In Model (3) 'Var' is 'Spouse's beliefs'—a binary variable with value 1 if the respondent's spouse believes the respondent chose swimming lessons in Decision Block 1, zero otherwise. Wife and husband regressions have 664 observations each. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table E.2: Beliefs interacted with treatments: Decision Blocks 2 and 3

	Defer					
	(1) Beliefs		(2) ESPDs		(3) ESPDs no swim	
	Wives	Husbands	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	0.082	0.031	0.135**	-0.008	0.145**	-0.014
	(0.060)	(0.059)	(0.065)	(0.054)	(0.064)	(0.053)
T3 - CS/Books	0.093**	-0.042	0.136***	-0.069	0.169***	-0.072
	(0.046)	(0.054)	(0.049)	(0.058)	(0.050)	(0.059)
T4 - CS/Rice	0.032	-0.030	0.090*	-0.145**	0.106**	-0.153**
	(0.061)	(0.054)	(0.049)	(0.063)	(0.046)	(0.065)
Var	0.005	0.128*	0.097	0.069	0.264**	-0.014
	(0.095)	(0.071)	(0.109)	(0.090)	(0.118)	(0.127)
T1 - WS/Kanga x Var	0.278	-0.120	-0.166	0.101	-0.382**	0.185
	(0.179)	(0.174)	(0.126)	(0.110)	(0.161)	(0.134)
T3 - CS/Books x Var	0.239***	-0.159	0.233*	-0.023	0.107	0.037
	(0.083)	(0.137)	(0.131)	(0.109)	(0.134)	(0.143)
T4 - CS/Rice x Var	0.119	-0.332***	-0.051	0.061	-0.261	0.143
	(0.138)	(0.106)	(0.135)	(0.138)	(0.165)	(0.167)
Predicted outcome under T2	0.392	0.552	0.393	0.552	0.394	0.547

	Consult					
	(1) Beliefs		(2) ESPDs		(3) ESPDs no swim	
	Wives	Husbands	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	0.015	-0.055*	0.013	-0.031	0.004	-0.044
	(0.041)	(0.029)	(0.043)	(0.032)	(0.042)	(0.031)
T3 - CS/Books	0.115**	-0.005	0.095*	0.034	0.090*	0.033
	(0.054)	(0.033)	(0.050)	(0.028)	(0.047)	(0.028)
T4 - CS/Rice	0.072	0.010	0.029	0.014	0.031	0.011
	(0.046)	(0.028)	(0.048)	(0.025)	(0.046)	(0.025)
Var	0.056	0.025	-0.009	0.090	-0.093	0.074
	(0.053)	(0.066)	(0.069)	(0.061)	(0.106)	(0.073)
T1 - WS/Kanga x Var	-0.002	-0.189	-0.004	-0.283**	0.097	-0.221
	(0.103)	(0.130)	(0.114)	(0.139)	(0.125)	(0.139)
T3 - CS/Books x Var	-0.073	0.036	0.050	-0.205*	0.176	-0.212*
	(0.057)	(0.067)	(0.093)	(0.108)	(0.111)	(0.120)
T4 - CS/Rice x Var	-0.122	-0.027	0.040	-0.079	0.070	-0.064
	(0.081)	(0.092)	(0.112)	(0.074)	(0.154)	(0.083)

Notes: Seemingly unrelated regression with the following dependent variables: Panel 1 uses 'Defer' where the choice in Decision Block 2 to defer decision-making to the spouse takes the value 1, and the choice to make the decision alone takes the value 0; Panel 2 uses 'Consult' where the willingness to consult with the spouse in Decision Block 3 takes the value 1 (i.e., if they are open to either hearing their spouse's first choice or sharing their own first choice with their spouse), 0 if they are closed to either hearing or sharing. The base treatment is 'T2 - WS/Soap'. Wife and husband regressions have 664 observations each. All models report treatment effects where treatment dummies are interacted with 'Var'. In Model (1) 'Var' refers to 'Beliefs'—a binary variable which captures the individual's beliefs about spousal first choices, taking the value 1 if they believed the spouse chose swimming lessons, zero if they believe the spouse chose good B. In Model (2) 'Var' is 'ESPD'—a binary variable equal to 1 if the respondent's belief about their spouse's first choice is different to the respondent's actual first choice, zero otherwise. In Model (3) 'Var' is 'ESPD no swim', which is the particular case of ESPD where a value of 1 is assigned if the respondent chooses swimming but believes their spouse does not choose swimming (i.e., they expect no swimming preference from the spouse), and is zero otherwise. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

F. Robustness checks

F.1. Regressions with bootstrapped standard errors

In the main analysis, we account for within-village correlation by clustering standard errors at the village level. However, with only 19 village clusters, conventional cluster-robust standard errors may underestimate true sampling variability and yield overly optimistic p-values. To address this, we bootstrap standard errors clustered at the village level and report the main regression results below using these more reliable estimates as recommended by Cameron, Gelbach and Miller (2008).

Table F.1: Treatment effects for Decision Block 1 (bootstrapped SEs)

	First choice	
	(1)	
	Main effects	
	Wives	Husbands
T1 - WS/Kanga	0.006	-0.018
	(0.042)	(0.032)
T3 - CS/Books	0.133**	0.090**
	(0.052)	(0.046)
T4 - CS/Rice	0.283***	0.343***
	(0.043)	(0.052)
Constant	0.157***	0.229***
	(0.029)	(0.025)
Predicted outcome under T2	0.157	0.229
F-test p-value for T1 vs T3	0.001	0.021
F-test p-value for T3 vs T4	0.002	0.000

Notes: Seemingly unrelated regressions with the dependent variable 'First choice' in Decision Block 1, equal to one if swimming lessons is chosen, zero otherwise. 'T2 - WS/Soap' is the base treatment. Wife and husband regressions have 664 observations each. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F.2: Treatment effects for Decision Blocks 2 and 3

	Defer		Consult	
	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	0.108*	0.012	0.012	-0.072*
	(0.063)	(0.053)	(0.043)	(0.041)
T3 - CS/Books	0.169***	-0.066	0.102***	0.000
	(0.054)	(0.052)	(0.038)	(0.037)
T4 - CS/Rice	0.084**	-0.127**	0.036	0.006
	(0.043)	(0.055)	(0.042)	(0.037)
Predicted outcome under T2	0.392	0.548	0.807	0.867
F-test p-value T1 vs T3	0.318	0.133	0.015	0.077
F-test p-value T3 vs T4	0.058	0.289	0.065	0.870

Notes: Seemingly unrelated regression with the following dependent variables. Model (1): defer in Decision Block 2 where the choice to defer decision-making to the spouse takes the value 1, and the choice to make the decision alone takes the value 0; Model (2): consult in Decision Block 3, where the willingness to consult with the spouse takes the value 1 (i.e., if they are open to either hearing their spouse's first choice or sharing their own first choice with their spouse), 0 if they are closed to either hearing or sharing. The base treatment is 'T2 - WS/Soap'. Wife and husband regressions have 664 observations each. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

F.2. Regression models with fixed effects and controls

Table F.3: Treatment effects for Decision Block 1

	First choice			
	(1) Main effects		(2) with beliefs	
	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	-0.003	-0.023	0.038	0.055*
	(0.043)	(0.029)	(0.034)	(0.031)
T3 - CS/Books	0.106*	0.116***	-0.001	0.139***
	(0.055)	(0.042)	(0.035)	(0.038)
T4 - CS/Rice	0.268***	0.354***	0.075*	0.234***
	(0.050)	(0.051)	(0.040)	(0.050)
Var			0.420***	0.763***
			(0.093)	(0.081)
T1 - WS/Kanga × Var			-0.175	-0.131
			(0.206)	(0.140)
T3 - CS/Books × Var			0.166	-0.127*
			(0.117)	(0.069)
T4 - CS/Rice × Var			0.176	-0.133
			(0.118)	(0.108)
Predicted outcome under T2	0.169	0.221	0.207	0.246
F-test p-value for T1 vs T3	0.002	0.003	0.263	0.043
F-test p-value for T3 vs T4	0.002	0.000	0.087	0.100

Notes: Seemingly unrelated regressions with the dependent variable 'First choice' in Decision Block 1, where the choice of swimming lessons takes the value 1, and the choice of the other good the value 0. 'T2 - WS/Soap' is the base treatment. Model (1) reports main treatment effects, and Model (2) reports treatment effects where treatment dummies are interacted with 'Var'. In Model (2) 'Var' refers to the variable 'Beliefs'—a dummy with value 1 if the respondent believes their spouse chose swimming lessons in Decision Block 1, zero otherwise. Wife and husband regressions have 664 observations each. Wife and husband regressions each have the same set of independent variables, namely, wife-specific and husband-specific village and enumerator fixed effects, wife-specific and husband-specific control variables including age, education, risk preferences, monthly income in a good month, and monthly income in a bad month, and couple-specific control variables including number of years of marriage, number of sons and daughters. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F.4: Treatment effects for Decision Blocks 2 and 3

	Defer		Consult	
	(1)		(2)	
	Wives	Husbands	Wives	Husbands
T1 - WS/Kanga	0.069	0.037	0.009	-0.057
	(0.048)	(0.055)	(0.035)	(0.036)
T3 - CS/Books	0.132***	-0.065	0.080**	-0.003
	(0.043)	(0.051)	(0.036)	(0.033)
T4 - CS/Rice	0.064	-0.148***	0.005	0.008
	(0.041)	(0.051)	(0.039)	(0.034)
Predicted outcome under T2	0.416	0.547	0.822	0.864
F-test p-value for T1 vs T3	0.141	0.052	0.027	0.112
F-test p-value for T3 vs T4	0.095	0.136	0.032	0.725

Notes: Seemingly unrelated regression with the following dependent variables. Model (1): defer in Decision Block 2 where the choice to defer decision-making to the spouse takes the value 1, and the choice to make the decision alone takes the value 0; Model (2): consult in Decision Block 3, where the willingness to consult with the spouse takes the value 1 (i.e., if they are open to either hearing their spouse's first choice or sharing their own first choice with their spouse), 0 if they are closed to either hearing or sharing. The base treatment is 'T2 - WS/Soap'. Wife and husband regressions have 664 observations each. Wife and husband regressions each have the same set of independent variables, namely, wife-specific and husband-specific village and enumerator fixed effects, wife-specific and husband-specific control variables including age, education, risk preferences, monthly income in a good month, and monthly income in a bad month, and couple-specific control variables including number of years of marriage, number of sons and daughters. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G. Demographic correlates of decision-making preferences, by treatment

Table G.1 presents OLS regressions of 'First choice' (= 1 if swimming lessons chosen in Decision Block 1) on demographic characteristics, estimated separately within each treatment cell for wives (columns 1–4) and husbands (columns 5–8).

Table G.1: Demographic correlates of decision-making preferences, by treatment

	Wives				Husbands			
	(1) T1	(2) T2	(3) T3	(4) T4	(5) T1	(6) T2	(7) T3	(8) T4
Age	-0.008**	-0.001	0.009*	0.002	-0.002	-0.006*	0.002	0.003
	(0.004)	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)	(0.007)	(0.007)
Spousal age difference	-0.006	-0.003	-0.001	-0.004	0.003	-0.009*	-0.009	-0.009
	(0.004)	(0.003)	(0.004)	(0.006)	(0.008)	(0.005)	(0.009)	(0.011)
Sons	0.016	-0.007	0.022	0.059**	-0.000	-0.002	0.057***	0.002
	(0.017)	(0.019)	(0.019)	(0.023)	(0.019)	(0.016)	(0.020)	(0.022)
Daughters	-0.041**	0.002	-0.020	-0.008	-0.012	-0.008	-0.036*	0.024
	(0.018)	(0.023)	(0.026)	(0.022)	(0.022)	(0.026)	(0.018)	(0.024)
Swimming skill	0.133	-0.178***	0.023	0.453***	0.047	0.067	0.113	0.104
	(0.126)	(0.042)	(0.157)	(0.121)	(0.079)	(0.086)	(0.096)	(0.080)
Time preference	-0.070	0.009	-0.053	0.042	-0.008	-0.084	-0.140*	-0.038
	(0.050)	(0.059)	(0.071)	(0.094)	(0.068)	(0.058)	(0.072)	(0.081)
Risk aversion	-0.129	-0.090	-0.017	-0.213**	0.126	0.116	0.082	-0.006
	(0.085)	(0.061)	(0.101)	(0.082)	(0.084)	(0.070)	(0.127)	(0.090)
Education	0.065	0.075	-0.135	0.181	0.126*	0.078*	-0.099	0.154
	(0.062)	(0.072)	(0.126)	(0.148)	(0.062)	(0.042)	(0.095)	(0.149)
Years married	0.005**	-0.000	-0.010**	-0.004	-0.000	0.004*	-0.002	0.000
	(0.002)	(0.002)	(0.004)	(0.005)	(0.003)	(0.002)	(0.005)	(0.007)
N	166	166	166	166	166	166	166	166

Notes: OLS regressions with 'First choice' (= 1 if swimming lessons chosen in Decision Block 1, zero otherwise) as the dependent variable. Columns (1) and (3) pool T1 and T2 (wife swimming, wife's domain); columns (2) and (4) pool T3 and T4 (child swimming, joint domain). Variables are defined as follows: 'Age' captures the respondent's age in years; 'Spousal age difference' is the absolute difference in years between the wife's and husband's age; 'Sons' and 'Daughters' give the number of sons and daughters in the household, respectively; 'Swimming skill' is a binary variable equal to 1 if the respondent reports being able to swim in deep water for 10 minutes without stopping, zero otherwise; 'Time preference' is a binary equal to 1 if the respondent chose the earlier, smaller payment option (Tsh 10,000 tomorrow over Tsh 15,000 in one month) in all three rounds of a sequential binary time preference elicitation task, indicating the most impatient type, zero otherwise; 'Risk aversion' is a categorical variable coded as 0 ('Takes risk a lot'), 1 ('Takes risk but not a lot'), 2 ('Avoids risk but not a lot'), and 3 ('Avoids risk a lot'); 'Education' is a binary equal to 1 if the respondent has at least a primary education, zero otherwise; and 'Years married' gives the number of years the respondent has been married. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

H. Uptake by retain vs not-retaining autonomy groups

Table H.1: Uptake of swimming lessons by decision block, for those who retain autonomy

	Decision block 1			Decision block 2			Decision block 3		
	(1)	(2)	(1)-(2)	(3)	(4)	(3)-(4)	(5)	(6)	(5)-(6)
Variable	Husbands	Wives	Pairwise t-test	Husbands	Wives	Pairwise t-test	Husbands	Wives	Pairwise t-test
T1	0.211	0.163	0.048	0.164	0.145	0.020	0.324	0.133	0.190*
	(0.409)	(0.370)			(0.373)	(0.354)		(0.475)	(0.346)
N	166	166	332	73	83	156	34	30	64
T2	0.229	0.157	0.072*	0.213	0.109	0.104*	0.182	0.188	-0.006
	(0.421)	(0.365)		(0.412)	(0.313)		(0.395)	(0.397)	
N	166	166	332	75	101	176	22	32	54
T3	0.319	0.289	0.030	0.326	0.164	0.161**	0.455	0.267	0.188
	(0.468)	(0.455)		(0.471)	(0.373)		(0.510)	(0.458)	
N	166	166	332	86	73	159	22	15	37
T4	0.572	0.440	0.133**	0.615	0.391	0.224***	0.571	0.615	-0.044
	(0.496)	(0.498)		(0.489)	(0.491)		(0.507)	(0.496)	
N	166	166	332	96	87	183	21	26	47

Notes: In Decision Block 1, uptake of swimming corresponds to the swimming preferences of 166 husbands and 166 wives in each treatment. In Decision Block 2, uptake of swimming in columns 3 and 4 correspond to the swimming preferences of husbands and wives, respectively, who chose *not* to defer (i.e., the sample excludes husbands and wives that deferred decision-making). In Decision Block 3, uptake of swimming in columns 5 and 6 corresponds to the swimming preferences of husbands and wives, respectively, who chose *not* to consult (i.e., the sample excludes husbands and wives that were open to consultation). Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table H.2: Uptake of swimming lessons by decision block, for those who do not retain autonomy

	Decision block 1			Decision block 2			Decision block 3		
	(1)	(2)	(1)-(2)	(3)	(4)	(3)-(4)	(5)	(6)	(5)-(6)
Variable	Husbands	Wives	Pairwise t-test	Husbands	Wives	Pairwise t-test	Husbands	Wives	Pairwise t-test
T1	0.211	0.163	0.048	0.183	0.181	0.002	0.174	0.147	0.027
	(0.409)	(0.370)		(0.389)	(0.387)		(0.381)	(0.355)	
N	166	166	332	93	83	176	132	136	268
T2	0.229	0.157	0.072*	0.154	0.246	-0.092	0.188	0.142	0.046
	(0.421)	(0.365)		(0.363)	(0.434)		(0.392)	(0.350)	
N	166	166	332	91	65	156	144	134	278
T3	0.319	0.289	0.030	0.263	0.301	-0.039	0.271	0.258	0.013
	(0.468)	(0.455)		(0.443)	(0.461)		(0.446)	(0.439)	
N	166	166	332	80	93	173	144	151	295
T4	0.572	0.440	0.133**	0.371	0.519	-0.148*	0.531	0.471	0.060
	(0.496)	(0.498)		(0.487)	(0.503)		(0.501)	(0.501)	
N	166	166	332	70	79	149	145	140	285

Notes: In Decision Block 1, uptake of swimming corresponds to the swimming preferences of 166 husbands and 166 wives in each treatment. In Decision Block 2, uptake of swimming in columns 3 and 4 correspond to the swimming preferences of the *spouses* of husbands and wives, respectively, who chose to defer (i.e., column 3 records wives' swimming preferences where husbands chose to defer, and column 4 records husbands' swimming preferences where wives chose to defer). In Decision Block 3, uptake of swimming in columns 5 and 6 corresponds to the swimming preferences of husbands and wives, respectively, who chose *to* consult (i.e., the sample excludes husbands and wives that were open to consultation). Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Spousal Effort: The Role of Confidence and Competition

Abstract

Household welfare is critically determined by how much effort spouses apply to income-earning activities. Yet, the behavioural motivators of spousal effort remain underexplored in economics. This is a notable gap given well-established gender differences in effort motivators, particularly confidence and competition. Moreover, prominent features of the spousal relationship—spousal evaluations and comparisons—arising from close proximity and repeated interaction can further influence spousal effort choices. To improve our understanding of what drives spousal effort, we conduct a field experiment with 1,328 spouses in Tanzania, where spouses must complete two rounds of an incentivised real-effort task. We use a 2 $\times$ 2 treatment design to experimentally vary whether, prior to round two, the individual receives feedback of a spousal evaluation (how well their spouse thinks they will do), a spousal comparison (how well their spouse performed in round one), neither, or both. By experimentally varying spousal evaluation and comparison feedback, we isolate how confidence and competitive motivations, respectively, operate within the household to shape spousal effort. Our results reveal striking differences to empirical evidence on gendered effort motivators. Across treatments, wives exhibit larger timing improvements than husbands. Moreover, wives in the sample are less (more) under(over)confident about their own performance and their spouse's performance than husbands are. Between treatments, when spousal comparisons are prompted, husbands' effort responds more strongly than wives' to being behind, suggesting they are more behind-averse than wives in the spousal setting. When spousal evaluations are combined with spousal comparisons, wives and husbands respond to different motivators. Specifically, wives are more responsive than husbands to positive spousal evaluations, reducing (increasing) effort if they are ahead (behind). On the contrary, husbands are more sensitive than wives to negative spousal evaluations, reducing effort whether they are ahead or behind. Such gender-specific responses of effort to comparisons and evaluations in an intrahousehold context carry policy implications for reducing inequalities in wage and labour outcomes between spouses.

1. Introduction

The spousal relationship is paradoxical in its ability to evoke between spouses both cohesion and division, cooperation and competition, altruism and self-interest (Sanders and Suls, 1982; Sen, 1990). Still, whilst these contrasting interpersonal dynamics have been analysed across a range of intrahousehold economic outcomes such as resource allocation, public goods provision, and consumption decisions, they are notably absent from the literature on effort. Specifically, to our knowledge, no study has examined how the spousal relationship moderates spousal effort and its behavioural motivators, in particular, confidence and competition. This is a critical research gap for three reasons.

First, as the 'grand gender convergence' makes dual-career households common (Goldin, 2014), and assortative mating forms couples with similar interests and careers (Greenwood et al., 2014), we might witness increasingly competitive environments for spouses. However, it is unclear how close proximity and repeated interactions would moderate relative spousal effort in a competitive environment. On the one hand, complete information about achievements could set a reference point for comparisons made on relative earnings, potentially creating an atmosphere of competition. On the other hand, altruism, conflict aversion, or a desire to promote marital stability may suppress the desire to compete with one's spouse on delicate matters such as income, instead motivating spouses to offer each other encouragement and confidence-building. Either way, insights into this moderation process could inform more targeted behavioural interventions which leverage spousal influences to help reduce spousal inequalities in wages, promotions or career progression.

Second, scholarship on the non-pecuniary behavioural motivators of effort in competitive environments find striking disparities along gender lines (Gneezy, Niederle and Rustichini, 2003; Niederle and Vesterlund, 2010, 2011). For example, women's relative performance gap in competitive settings has been linked to their relative underconfidence (Buser, Niederle and Oosterbeek, 2014; Niederle and Vesterlund, 2010), their lower willingness to compete (Gneezy, Niederle and Rustichini, 2003), their stronger preference for cooperation-based effort rewards (Kamas and Preston, 2015), and even their lower responsiveness to luck (Gill and Prowse, 2014). Moreover, in competitive environments of mixed gender groups, these differences become even more pronounced. For instance, Buser, van den Assem and van Dolder (2023) find women more likely to opt out of high-stakes competition when they discover their opponent is a male. In light of these empirical findings, the most gendered of relationships—the spousal relationship—offers a natural testing ground to further examine the link between the interaction of gender with behavioural effort motivators in a competitive environment.

Third, standard models of household behaviour—particularly collective household models—assume allocative efficiency within the household (Chiappori, 1988, 1992). Yet, a growing body of empirical evidence documents substantial inefficiencies in resource allocation within households (Ashraf, 2009; Duflo, 2003; Udry, 1996). For example, Udry (1996) famously evidences that households do not achieve Pareto efficiency in agricultural production where plots controlled by women are farmed less intensively than similar plots controlled by men within the same household. Although later experimental evidence attributes such inefficiencies to drivers such as bargaining power (Iversen et al., 2011), asymmetric information (Ashraf, 2009), or a lack of cooperation (Verschoor et al., 2019), less attention is paid to the possibility that inefficiencies may also reflect gendered distortions in effort. Still, a growing body of evidence points to this channel. For example, Bursztyn, Fujiwara and Pallais (2017) find that women reduce labour market ambition in response to spousal expectations, and Masekesa and Munro (2020) experimentally show that unequal outcomes within couples reduce individual effort. This kind of effort suppression may be exacerbated by feelings of love, altruism, compassion or guilt governing the spousal relationship, explaining why effort choices may deviate from the household optimum, even in environments characterised by shared goals and pooled resources.

In light of these considerations, we provide new experimental evidence on spouses' influence on each other's effort, focusing on the role of confidence and competitiveness as behavioural motivators of effort

provision. We consider these motivators in particular, as they are especially susceptible to influences of spousal comparisons and evaluations, processes inherent in the spousal relationship given spouses' close proximity, regular exchanges, and daily interactions.

To generate this evidence, we constructed a unique dataset that combines a lab-in-the-field experiment and household data collected from a sample of 1,328 spouses (i.e., 664 married or co-habiting couples) from rural coastal villages in the Zanzibar archipelago, off the eastern coast of Tanzania (see panel **(a)** in Figure 1.3 of Chapter 1). This sample is of empirical interest, given previously gendered seascape livelihoods such as fishing (for men) and seaweed-farming (for women) are now becoming less gender segregated, and a confluence of livelihoods requiring similar abilities (e.g., swimming, driving boats) by both men and women creates a naturally competitive environment.

In the experiment, couples performed an incentivised real-effort task which involved separating coloured matchsticks in the quickest time possible, with payments deducted for longer timings. The same task was done twice. In the first round, spouses did not know that the other spouse was performing the same task in the adjacent room, whilst in the second round, this was made known to spouses. Before the first round, we elicited self-confidence beliefs, which captured how a respondent expected to perform. Before the second round of the task, we elicited what we term 'spouse-confidence' beliefs, which captured how a respondent expected *their spouse* to perform. While round one was the same for all participants, we varied the second round using a 2×2 factorial treatment design, with treatments randomly assigned at the couple level.

The four between-subject treatments varied the type of feedback provided to each respondent before round two as being either their spouse's actual performance in round one (spousal comparison), their spouse's spouse-confidence belief about the respondent's performance in round one (spousal evaluation), both (combined) or neither (control). We hypothesised that feedback of the spouse's actual performance would influence a respondent's *competitiveness* through direct spousal comparisons, and that feedback of the spouse's spouse-confidence belief would influence a respondent's *confidence* through a positive or negative spousal evaluation (defined as the relative difference between the individual's self-confidence and the spouse's spouse-confidence).[14]

The latter hypothesis, that spouses adjust their effort when they know what their spouse thinks of their ability, is based on studies from psychology which find that others' expectations of us can affect our performance positively (e.g., the 'Pygmalion effect' (Eden, 1990; Rosenthal and Jacobson, 1968) and the 'Underdog effect' (Xue et al., 2022)), but also negatively (e.g., the 'Golem effect' (Babad, Inbar and Rosenthal, 1982; Davidson and Eden, 2000; Oz and Eden, 1994)). It also relates closely to theories in psychology such as the 'looking-glass self' (Cooley, 1922) and 'reflected appraisal' (Felson, 1985) which postulate that self-image is determined by how others perceive us.

Our study offers three main results. First, descriptive statistics reveal that across all treatments, wives have higher relative timing improvements, even when controlling for starting levels of ability. They also reveal that wives in the sample are less (more) over(under)confident about their own performance and their spouse's performance than husbands are. Second, treatment effects show that when spousal comparisons are made possible, husbands tend to be more behind-averse than wives. Specifically, the more the husband's performance lags behind the wife's, the more their performance improves in the second round. Third, although spousal evaluations do not influence effort on its own, when combined with spousal comparisons, we find wives

[14]Confidence is defined in economics as one's belief about one's own ability where true ability is unknown (Bénabou and Tirole, 2002). The standard economics approach to influence confidence is to feed information about an individual's actual performance rankings in a real-effort task. We depart from this approach. Instead, we feed information about the individual's *spouse's beliefs* about the individual's performance. We apply this novel approach as we posit that spouses are inclined to adjust effort in line with the relative confidence differences between self and spouse (i.e.,either positive or negative spousal evaluations).

and husbands have different responses to relative confidence differences between self and spouse. Specifically, wives are more responsive than husbands to positive spousal evaluations, reducing (increasing) effort if they are ahead (behind). On the contrary, husbands are more sensitive than wives to negative spousal evaluations, reducing effort whether they are ahead or behind. Moreover, husbands are four times less likely than wives to receive a negative spousal evaluation.

Our work makes several important contributions. Broadly, we contribute to the literature examining the motivators of effort. The efficiency of economic production in any setting ultimately depends on individual effort. Thus, the motivators of effort—both monetary and non-monetary—have been the focus of extensive economic research (DellaVigna and Pope, 2018). Beyond financial incentives, behavioural economics has revealed a rich set of psychological and social determinants of effort such as reciprocity (Akerlof, 1984; Fehr and Gächter, 2000), self-confidence (Bénabou and Tirole, 2002), social comparisons (D'Exelle, Habraken and Verschoor, 2024; Frey and Meier, 2004; Gächter and Thöni, 2010; Huet-Vaughn, 2025), and inequality aversion (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999). We build on this literature by examining effort allocation within the under-explored interpersonal context of households, where emotional and economic interdependencies between spouses make behavioural and social motivators particularly relevant. In doing so, we also contribute to four specific strands of literature.

The first is the literature on gender differences in competition.[15] The prevailing consensus is that women are significantly less likely than men to enter into competitive environments, and respond less to competitive incentives, even when their performance under competition is objectively similar to men's performances (Booth and Nolen, 2012; Croson and Gneezy, 2009; Flory, Leibbrandt and List, 2015; Gneezy and Rustichini, 2004; Gneezy, Niederle and Rustichini, 2003; Niederle, 2017; Niederle and Vesterlund, 2007). These gender differences in competitiveness have been linked to broader inequalities in economic outcomes (see Niederle and Vesterlund (2011) for a review). For example, boys' greater willingness to compete leads them to specialise, and even outperform, in more maths- and science-intensive tracks than girls (Buser, Niederle and Oosterbeek, 2014; Buser, Peter and Wolter, 2017; Niederle and Vesterlund, 2010), contributing to gender gaps in later education and labour outcomes. Within this literature, our study relates closest to experiments which examine behavioural responses to competition when the opponent is of the opposite sex. For instance, Datta Gupta, Poulsen and Villeval (2013) find men more likely to choose competition when matched with a man than a woman. Similarly, Booth, Hayashi and Yamamura (2022) find that when the gender of the competitor is revealed, men work harder to outperform a woman but only aim to match the target set by another man. On the contrary, women avoid competing against men (Buser, van den Assem and van Dolder, 2023), and when they do compete, women's performances are lower in mixed-sex than single-sex tournament settings (Booth and Yamamura, 2018), and more prone to mistakes when playing against men (Backus et al., 2023). We extend this literature to the spousal relationship, where the identity of the opponent being one's partner may fundamentally alter competitive dynamics.

The second is the literature on gender differences in (over)confidence, a substrand of the competition literature which looks at key psychological mechanisms of the gender gaps in competitiveness. A person with 'overconfidence' is defined broadly as a low-ability person who holds a belief that they are a high-ability person, where utility of holding such an incorrect belief is interpreted as being derived from ego

[15] A person's competitiveness in economics is measured using several well-established approaches. The most common follows the framework of Niederle and Vesterlund (2010) and Niederle and Vesterlund (2011), which defines competitiveness through an individual's willingness to compete—typically elicited by offering a choice between a piece-rate payment scheme and a competitive (tournament) scheme. Other seminal approaches include measuring competitive preferences through performance responses to tournament incentives (e.g., Gneezy, Niederle and Rustichini (2003)) or through entry decisions into contests with varying stakes and opponents (e.g., Buser, Niederle and Oosterbeek (2014)). In this study, we interpret increases in effort within a competitive setting as reflecting a willingness to compete, consistent with these foundational measures of competitive behaviour.

or self-serving bias (Bénabou and Tirole, 2002; Köszegi, 2006). A consistent finding in this literature is that men tend to overestimate their relative performance, while women underestimate theirs, even with similar abilities (Adamecz-Völgyi and Shure, 2022; Bastani, Giebe and Gürtler, 2023; Buser et al., 2020). Moreover, men and women differ in how they update their beliefs based on performance feedback. For example, Möbius et al. (2022) find men to be more responsive to feedback, whether negative or positive. Our contribution to this literature is methodological and empirical. Methodologically, we introduce a novel approach to influencing confidence. Whereas feedback typically used to influence confidence is that of the individual's actual performance, we use spousal evaluations. We motivate this methodological approach based on empirical evidence from psychology to support the notion that evaluations from peers can influence self-esteem and self-evaluation. Whilst studies have tested this in workplace settings (Donia, O'Neill and Brutus, 2015), classroom learning (Feyli and Ayatollahi, 2016; Ruegg, 2018; Sari, Tajuddin and Sholikah, 2025), and even child development (Cook and Douglas, 1998; Nurra and Pansu, 2009), its relevance within the spousal context remains unexplored in both psychology and economics. Empirically, we offer novel evidence on gender differences in spousal evaluations, and its role in influencing effort.

The third is the literature on comparisons in competitive settings. Seminal work in behavioural economics shows that individuals care not only about their absolute outcomes but also about their position relative to others (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), with experimental research demonstrating how this can affect effort and performance, even in the absence of direct financial incentives (Charness and Kuhn, 2007; Charness et al., 2016; Cohn et al., 2014; D'Exelle, Habraken and Verschoor, 2024; Gächter and Thöni, 2010; Huet-Vaughn, 2025). For instance, applied to the interpersonal relationship of coworkers, Ashraf, Bandiera and Lee (2014) show that salient rankings among health workers in Zambia led to deteriorations in performance. We extend this literature by examining how comparisons operate within spousal relationships. Where emotional and economic interdependencies are highlighted through sustained interactions, it is unclear whether greater salience of relative performance drives or dampens effort between spouses.

Related to the social comparison literature, we also contribute to work around inequality aversion—a social preference arising from social comparison. Prior studies have explored gender differences in inequality aversion primarily through distributional choices—for instance, in dictator or ultimatum games, where women more often choose to equalise payoffs (D'Exelle and Riedl, 2018; Friedl, Pondorfer and Schmidt, 2020; Kamas and Preston, 2015). In contrast, our study focuses on inequality in effort, not outcomes, that is, how individuals adjust their performance in response to their relative position vis-à-vis their spouse. By shifting the focus from outcome allocation to behavioural adjustment, we capture a more dynamic dimension of inequality aversion. Moreover, since competitive environments generate inherent inequalities (i.e., with clear winners and losers), women's aversion to inequality might further explain gender differences in competitiveness. To our knowledge, Dasgupta et al. (2019) is the only other study to examine this link. Our work extends theirs by applying it to the spousal domain. Given women's greater egalitarian preferences and the strong altruistic bonds between spouses, we expect inequality aversion to manifest even more strongly within households, through a suppression of effort in competitive settings.

The rest of this paper is organised as follows. In Section 2 we present the design of our experiment, including experimental procedures, treatments, a simple theoretical model and a series of testable hypotheses, and an empirical strategy (including a brief description of sampling and field implementation). In Section 3 we analyse the data. We first offer some summary statistics of the sample's socio-demographic characteristics and performances in the real-effort task. We then present treatment effects using regression analysis, and explore possible mechanisms and alternative specifications through heterogeneity analyses and robustness checks. A discussion of our results ensues in Section 4 and Section 5 concludes.

2. Design

2.1. Experimental procedures

The experiment was conducted with a sample of 664 married or co-habiting couples following the procedures outlined below and depicted in Figure 3.1.[16,17]

Upon providing consent to participate, spouses in each couple were ushered into separate rooms and interviewed at the same time by same-gender enumerators using a household survey questionnaire. After the interview, each spouse was informed that they would need to perform a real-effort task in two rounds. For the task, the individual was given a small box containing 80 matchsticks of two colours, and asked to separate the matchsticks into two separate stacks, one for each colour. The enumerator demonstrated the task to the respondent using 10 matchsticks. Respondents were told that they would be paid TZS 4,000 (∼USD 1.5) if they finished the task in 30 seconds, with TZS 10 (∼USD 0.004) deducted for every additional second taken thereafter to complete the task. They were also told that they needed to perform the task twice and only one of their two performances would be selected at random for payment (for the full experimental script see Appendix D).

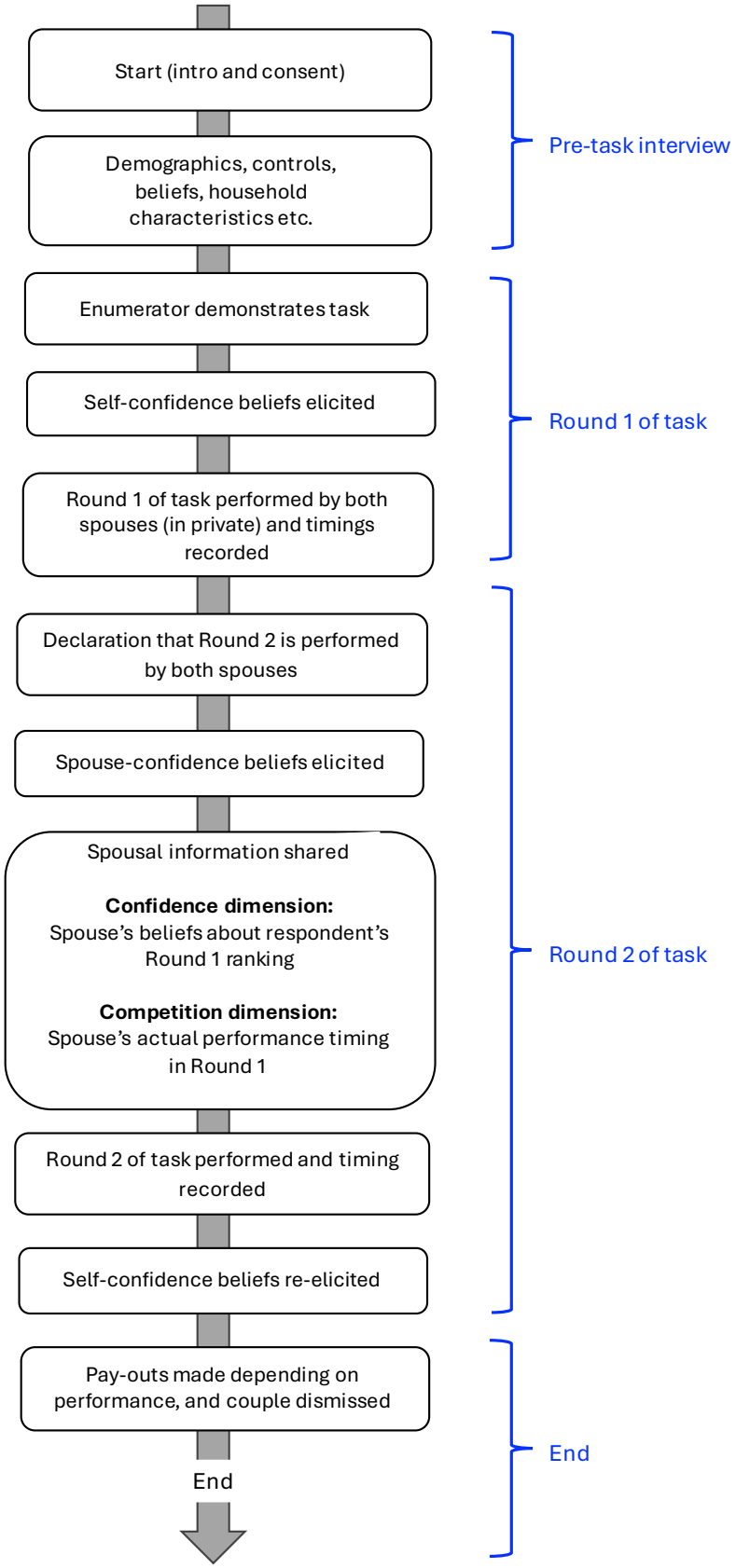
Before Round 1 commenced, enumerators elicited respondents' self-confidence, i.e., their belief about how they would rank in Round 1. This was measured by asking the respondent to state where they thought they would rank, using three consecutive questions on whether they thought their performance timing would be in the top or bottom 50%, 25%, and 12.5%, respectively, of same-gender participants in their village (Appendix E presents the full set of self-confidence questions). After this, participants were asked to perform Round 1 of the task individually and in private. Precautions were taken to ensure that respondents did *not* know that their spouse was doing the same task in the next room. After the task was completed and timings recorded, participants then moved to the second round.

Before Round 2 of the task, enumerators shared with each participant that their spouse was performing the same task in the other room. This was meant to establish a competitive environment since both spouses now knew that the other spouse was performing the same task at the same time. Next, the respondent's belief about their spouse's ability—what we term their 'spouse-confidence' belief—was elicited using the same measures of self-confidence detailed above, but applied to the spouse. After this was collected, each spouse was given information according to the treatment to which the couple had been randomly assigned (details on treatments are provided in Section 2.2). Once both spouses received the relevant information that their treatment dictated, each participant was asked to perform the task a second time. After Round 2, individual self-confidence beliefs were elicited for a second time. Once both spouses completed both rounds and final beliefs were elicited, payments were made and the couple was dismissed. Note that neither self-confidence beliefs nor spouse-confidence beliefs were elicited with monetary incentives. Whilst incentivised belief elicitation is standard practice, we deliberately chose not to incentivise predictions in this setting to avoid introducing interactions between belief reporting and subsequent effort provision in the performance task. The principal concern is the hedging problem: when participants have a financial stake in the predicted action, incentivised belief elicitation can induce participants to hedge between their stated beliefs and their subsequent choices (Armantier and Treich, 2013; Blanco et al., 2010; Rutström and Wilcox, 2009). In our

[16]The full sample was subject to two lab-in-the-field experiments, one of which is discussed in Chapter 2 of this thesis, the other which is the focus of this Chapter 3. To ensure the two experiments did not affect the outcomes of each other, beyond the randomisation of treatment assignment in each experiment, the order in which the two experiments were conducted was randomised at the couple level. Please take note of this when reading the full experimental script presented in Appendix D.

[17]The pre-analysis plan of this study was registered in the OSF Registries on the 14th March 2025 and can be accessed here: https://osf.io/g9xpu

Figure 3.1: Experimental procedures for spousal effort



design, this risk was particularly acute—participants who were incentivised to report accurate predictions about their own performance may have tempered their actual effort to bring outcomes in line with stated beliefs, or vice versa, directly confounding the behavioural responses of interest (Blanco et al., 2010; Gneezy and Rustichini, 2000; Gächter and Renner, 2010; Rutström and Wilcox, 2009). Our aim was therefore to capture beliefs as they naturally inform effort decisions, rather than as objects of optimisation.

2.2. Treatments

We used a 2 × 2 between-subject treatment design, where couples were randomly assigned to one of four treatments which varied information along two dimensions. The **confidence** dimension experimentally varied whether, prior to Round 2, each spouse in the couple received information that could either undermine or boost their confidence. That is, they were told *their spouse's answer* to the spouse-confidence questions. We expect this to moderate confidence as spouses could compare how well they themselves expect to perform (their self-confidence) with how well their spouse expects them to perform (their spousal evaluation).

The **competition** dimension experimentally varied whether, prior to Round 2, each spouse in the couple received information that could induce competition. That is, they were told their spouse's actual completion time in Round 1. We expect this to spark competition as spouses are given the necessary information to make comparisons of their performance with their spouse's performance. The combination of both dimensions leads to four treatment cells (Figure 3.2). Note that in treatment T4, spousal evaluations were shared first, then the spouses' timings.

Figure 3.2: Experimental treatments for spousal effort

		Spousal evaluation (confidence)	
		No information shared	Spouse's beliefs shared
Spousal comparison (competition)	No information shared	T_1	T_2
	Spouse's timing shared	T_3	T_4

2.3. Theory and hypotheses

We formulate some hypotheses for how we expect timing improvements between both rounds to vary across treatments. To structure our arguments, we first put forth a basic model of individual behaviour in the effort task.

Simple model

To begin, take an individual i who must choose their effort level e_i as input into their concave production function $p(e_i)$ where $p' > 0$ and $p'' < 0$. The effort they choose also incurs a cost determined by the convex cost function $c(e)$ where $c' > 0$, $c'' > 0$. The aim of the individual is to maximise their payoff Q where:

$$Q_i = p(e_i) - c(e_i) \tag{3.1}$$

which is simply the earnings from effort minus the costs from effort. Note, we assume their production function is equivalent to their earnings function. Solving the first order condition (FOC) which equates to zero the first derivative of the payoff Q_i with respect to effort e_i, we get:

$$\frac{\partial Q_i}{\partial e_i} = p'(e_i) - c'(e_i) = 0 \implies p'(e_i^*) = c'(e_i^*) \tag{3.2}$$

In other words, the individual's optimal effort e_i^* equalises the marginal revenue of effort and the marginal cost of effort.

Model extension with spousal evaluations and spousal comparisons

Now consider the possibility that an individual i's effort is influenced not only by their own payoff but also by spousal evaluations from, and spousal comparisons with, their spouse j. To capture these external social influences arising from spousal evaluations and comparisons, we modify the utility term U_i so that:

$$U_i = p(e_i) - c(e_i) + \Theta(e_i, e_j) \tag{3.3}$$

where U_i depends on the individual's payoff Q_i and Θ, a function of *both spouses'* effort choices e_i and e_j, such that:

$$\Theta(e_i, e_j) = \underbrace{[d_i \cdot \gamma_P(e_i) + (1 - d_i) \cdot \gamma_N(e_i)]}_{\text{spousal evaluation}} + \underbrace{[\beta \cdot g(e_i, e_j) - \alpha \cdot f(e_i, e_j)]}_{\text{spousal comparison}} \tag{3.4}$$

The terms in the first square bracket capture the effect of spousal evaluations on utility, and the terms in the second square bracket capture the effect of spousal comparisons on utility. In sum, Θ represents the *additive* shift in benefits from two distinct psychological channels—spousal evaluations and comparisons. We now define the square bracket terms in more detail.

Spousal evaluation (first bracket): This captures how individual i responds to the gap between their evaluation of their own ability i.e., their self-confidence, and their spouse j's evaluation of i's ability i.e., the spouse's spouse-confidence. The indicator d_i equals one if the spouse's evaluation is positive (i.e., j's spouse-confidence is greater than i's self-confidence and so effort is determined by the function γ_P), and zero if the spouse's evaluation is negative (i.e., j's spouse-confidence is lower than i's self-confidence and so effort is determined by the function γ_N). We assume that the function $\gamma_P(e_i) > 0$ and that its functional form is such that it has a positive first-order derivative with respect to e_i i.e., $\gamma_P'(e_i) > 0$. In words, when the spouse is more optimistic than the individual about i's ability, there is a marginal utility gain or 'motivational boost' from an *encouragement effect* that occurs. This is in line with the notable psychological channel known as the 'Pygmalion effect', which dictates that positive expectations can lead to improved performance (Daido and Itoh, 2005; Eden, 1990; Rosenthal and Jacobson, 1968). Conversely, we assume that $\gamma_N(e_i) < 0$, to capture the effect of pessimistic spousal evaluations on utility, but that its functional form results in a *positive* marginal utility i.e., $\gamma_N'(e_i) > 0$, to capture a *defiance effect* or a 'prove-them-wrong' incentive that leads the individual to exert more effort when the spouse is more pessimistic as a way of contesting their negative evaluation. This is in line with new research on the 'underdog' effect, where low expectations would increase effort (Nurmohamed, 2020; Xue et al., 2022).[18]

[18] We also assume that $\gamma_P''(e_i) < 0$ and $\gamma_N''(e_i) < 0$, i.e., that positive evaluation and negative evaluation functions are concave.

Spousal comparison (second bracket): This captures how the individual i responds to learning about the relative standing of their own earnings in relation to their spouse's earnings. $\alpha > 0$ captures behind-aversion, where individuals increase effort to close an earnings gap; and $\beta > 0$ captures ahead-seeking motivation, where individuals increase effort to widen their lead. Both are in line with competitive preferences. Specifically, the utility reducing effect of being 'behind' is similar to a dislike for disadvantageous inequity as modelled by Fehr and Schmidt (1999) and Demougin, Fluet and Helm (2006). However, we deviate from the Fehr and Schmidt (1999) model when people are 'ahead': whilst they anticipate ahead-averse behaviour that is motivated by guilt or compassion (Bartling and von Siemens, 2010), i.e., a dislike for advantageous inequity, we assume that spouses are competitively ahead-seeking and that being ahead *increases* rather than decreases utility.[19] This is more in line with the model of Roels and Su (2014) and supported by empirical evidence of Gill et al. (2019). Thus, $\beta.g(e_i, e_j)$ represents the utility gain from being ahead in earnings and $-\alpha.f(p(e_i), p(e_j))$ represents the utility loss from being behind in earnings, where:

$$f(e_i, e_j) = \begin{cases} \big(p(e_j) - p(e_i)\big) & if \quad p(e_i) < p(e_j) \\ 0 & if \quad p(e_i) \geq p(e_j) \end{cases}$$

$$g(e_i, e_j) = \begin{cases} 0 & if \quad p(e_i) \leq p(e_j) \\ \big(p(e_i) - p(e_j)\big) & if \quad p(e_i) > p(e_j) \end{cases}$$

Expected outcomes under experimental treatments

In order to make predictions around spousal effort outcomes in each of our treatments, we modify Θ according to each treatment, and then solve the FOC of the corresponding utility function. For instance, in the control group (T1), with no spousal evaluations or comparisons being made known, $\Theta(e_i, e_j) = 0$, and so the FOC collapses to equation 3.2. In all other treatments, the FOC can be generalised to equation A.2 (and specified by each of the separate FOCs derived in Appendix A). Using these final FOCs, we can formulate hypotheses on how timing improvements between Round 1 and Round 2 compare between T1 (where there is no feedback between rounds) and T2 (where spousal evaluations are shared), between T1 and T3 (where timings are shared and thus inequalities are made known), and between T1 and T4 (where both timings and evaluations are shared).[20]

Hypothesis 3.1: *Timing improvements between rounds are larger in T2 than in T1.*

In T1, where spousal evaluations nor comparisons are shared, we have $\Theta = 0$ and thus $p'(e_i^*) = c'(e_i^*)$. In T2, receiving a spousal evaluation increases the marginal benefit of effort ($\Theta > 0$), both when a respondent receives a positive or negative evaluation, increasing the optimal effort e^* relative to T1 (see equations A.4 and A.5 in Appendix A).

Hypothesis 3.2: *Timing improvements between rounds are larger in T3 than in T1.*

In T1, where spousal evaluations nor comparisons are shared, we have $\Theta = 0$ and thus $p'(e_i^*) = c'(e_i^*)$. In T3, hearing spousal comparisons increases the marginal benefit of effort ($\Theta > 0$), both when a respondent is ahead or behind, increasing the optimal effort e^* relative to T1 (see equations A.7 and A.8 in Appendix A).

[19] We revisit the implications of this departure from standard fairness assumptions, and its relationship to alternative fairness-based models, in the discussion section.

[20] Note that we ignore higher-level reasoning, where one spouse anticipates that the other spouse might also adjust their behaviour in Round 2, and therefore adjusts to the *expected change* of the spouse.

Hypothesis 3.3: *Timing improvements between rounds are larger in T4 than in T1, T2 or T3.*

While T2 and T3 isolate the effects of evaluations and comparisons respectively, T4 captures their joint influence. In T4, Θ incorporates both an evaluation and a comparison effect. Because $\gamma'_P(e_i), \gamma'_N(e_i), \alpha, \beta > 0$, the aggregate shift in the marginal benefit of effort is strictly greater in T4 than in any other treatment. Given the concavity of $p(e)$ and convexity of $c(e)$, this combined effect results in the highest level of optimal effort e^*, leading to the largest timing improvements relative to all other treatments.

Hypothesis 3.4: *Timing improvements in Hypotheses 3.1-3.3 are larger for husbands than wives.*

Given empirical evidence that men are more competitive than women (Niederle and Vesterlund, 2007; Sarsons and Xu, 2015), we expect competitive motivational boosts arising from evaluations and comparisons to be stronger amongst men than women.

Hypothesis 3.5: *Couples' timing improvements are larger in T2, T3, and T4, compared to T1.*

Given that on an individual level, both spouses are expected to improve their timings, we also expect couples' timings improvements to be larger under each of the treatments T2, T3, and T4, compared to T1.

Individual effort within stylised household models

Our simple theoretical model on effort presented above assumes that effort choices are made individually by spouses, regardless of household-level objectives. However, spouses in this experiment form real couples, where individual effort choices translate to outcomes relevant for the *household*. To situate our design within the household economics literature, we briefly consider how two canonical models—the unitary and collective models—differ in their implications for effort provision (formal derivations are presented in Appendix B).

Under the unitary model, the household maximises total income as a single decision-making unit, so each spouse's effort is determined entirely by individual returns and production technology (Appendix B, equation B.2). Thus, this benchmark predicts no variation in effort across treatments in our experiment. However, empirical evidence broadly rejects the unitary model in favour of richer frameworks such as the collective model (see, e.g., Browning et al., 1994; Lundberg and Pollak, 1993). The collective model relaxes the assumption of a single unified objective. Specifically, it allows for spouses to have heterogeneous preferences and bargaining over outcomes, such that the household objective becomes a weighted sum of individual utilities where the weights represent the relative bargaining power of each spouse (Browning and Chiappori, 1998). In our formulation, individual utilities within this framework incorporate private costs of effort and the psychological responses to social comparisons and evaluations captured by Θ (Appendix B, section B.2). Thus, the collective model predicts the same spousal behaviour as our theoretical model: in treatments that 'activate' the psychological channel through Θ, optimal effort is higher than in the control group where $\Theta = 0$, while the unitary benchmark, where Θ plays no role, predicts no such variation.

2.4. Regression specifications

To test Hypotheses 3.1-3.3, we estimate the following regressions with OLS, on wives and husbands separately:

$$Y_{ic} = \beta_1 + \beta_2 \text{T2}_{ic} + \beta_3 \text{T3}_{ic} + \beta_4 \text{T4}_{ic} + \beta_5 \mathbf{X}_{ic} + \mu_e + \sigma_v + \epsilon_{ic} \tag{3.5}$$

where for each spouse i of couple c, Y_{ic} is the dependent variable which is either the absolute or relative difference in timings of individual i in couple c. The first is measured as the difference in timings between Round 1 and Round 2. The second divides this difference by the timings in Round 1. In this model, T1 is the reference group, and β_1 is the mean outcome under T1. T2, T3, and T4 are dummy variables for

the respective treatment groups, where coefficients β_2, β_3 and β_4 estimate the difference between T2, T3, and T4 respectively, and T1. To increase efficiency of the estimates we include a vector of control variables $\mathbf{X}_{ic}$ (e.g., the respondent's age, education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with their spouse), fixed effects for the specific enumerator who conducts the interview (μ_e for enumerator e) and village fixed effects (σ_v for village v). ϵ_{ic} is the error term. We estimate the regressions using OLS with village clustered standard errors, to handle non-independencies between respondents of the same village which could underestimate standard errors.

To test Hypotheses 3.1-3.3, we examine each of the coefficients β_2, β_3 and β_4, respectively. To further test Hypothesis 3.3 that the combined effect on effort is larger than the isolated effects, we use an F-test to examine the null hypotheses $H_0 : \beta_4 = \beta_2$ and $H_0 : \beta_4 = \beta_3$.

To examine whether our hypothesised treatment effects arise from the relevant psychological channels, we run the following regressions:

$$Y_{ic} = \beta_1 + \sum_{j=2}^{4} \beta_j \mathrm{T}_j + \theta \mathrm{I}_{ic} + \sum_{j=2}^{4} \gamma_j (\mathrm{T}_j \times \mathrm{I}_{ic}) + \delta \mathbf{X}_{ic} + \mu_e + \epsilon_{ic} \tag{3.6}$$

Here, our two dependent variables are regressed separately on interactions of I_{ic} with our treatment indicators T_j for $j = 2, 3, 4$, where I_{ic} is specific to each psychological channel.

For example, to examine whether T3 treatment effects work through behind-averse and ahead-seeking motivations, we interact treatment indicators with $\mathrm{I}_{ic} = Behind_{ic}$ and $\mathrm{I}_{ic} = Ahead_{ic}$ in separate regressions, where *Behind* is a continuous variable that indicates by how much a respondent's performance is lagging behind their spouse's performance if they are behind, and is zero if they are not behind, and *Ahead* is a continuous variable that indicates by how much a respondent's performance is better than their spouse's performance if they are ahead, and is zero if they are not ahead. A significant and positive coefficient on $\mathrm{T}_3 \times \mathrm{I}_{ic}$ would confirm ahead-seeking and behind-averse motivations driving larger improvements in timings in T3 (where spousal timings are shared) relative to T1.

Similarly, to examine whether T2 treatment effects work through the encouragement effect and the defiance effect from positive and negative evaluations, respectively, we interact treatment indicators with $\mathrm{I}_{ic} = Poseval_{ic}$ and $\mathrm{I}_{ic} = Negeval_{ic}$ in separate regressions, where *Poseval* is a binary that takes the value one if a respondent's self-confidence is higher than their spouse's spouse-confidence, zero otherwise, and *Negeval* is a binary that takes the value one if a respondent's self-confidence is lower than their spouse's spouse-confidence, zero otherwise.[21] Again, a significant and positive coefficient on $\mathrm{T}_2 \times \mathrm{I}_{ic}$ would confirm that encouragement and defiance motivations are driving larger improvements in timings in T2 (where spousal evaluations are shared) relative to T1.

[21] We make use of a few specific measures to construct *Poseval* and *Negeval*: self-confidence, spouse-confidence, and spousal evaluations. These are defined in the following way. 'Self-confidence' is a binary variable which takes the value one if a respondent answers with 'Not top person' consistently to the suite of self-confidence questions (see Appendix E), which effectively adds up to the question "If we compare your timing in Round 1 with eight other men/women in the village, will you be the top 1 of 8?". The binary variable takes the value zero if their suite of answers effectively adds up to 'Top person'. Note that throughout the chapter, a self-confidence value of one corresponds to lower self-confidence, specifically, a belief that one will not be the top performer among same-gender village peers. We accordingly describe respondents with this value as 'less confident' or 'lower self-confidence'. 'Spouse-confidence' is a binary variable with value one if the respondent answers with 'Not top person' to the suite of questions in Appendix E which effectively asks the same self-confidence question, but about their spouse's performance. 'Spousal evaluation' is simply the respondent's spouse's answer to the spouse-confidence question i.e., their spouse's spouse-confidence. Finally, spousal evaluations are delineated as positive and negative: *Poseval* is a dummy that takes the value one if the self-confidence dummy is one (not top person) whilst the spouse evaluation dummy is zero (top person), zero otherwise, and vice versa for 'Negeval'.

To test Hypothesis 3.4 that improvement gains as per Hypotheses 3.1-3.3 are larger for husbands than wives, we run the same regression of equation 3.6 with $I_{ic} = Gender_{ic}$, a binary variable equal to one if the respondent is female, and examine the interaction term coefficients. Last, to test Hypothesis 3.5, we run the same regression as in equation 3.5, but with the unit of analysis being the couple instead of the individual, and thus the dependent variable on timing improvements being the sum of timing improvements of both husband and wife.

Finally, using the same empirical model of equation 3.6, we explore heterogeneity in the main regression findings across three spousal characteristics with the potential to moderate the spousal relationship, and hence treatment effects: competitive preferences, marital trust, and conflict aversion. To do this, we re-run the treatment analysis using ahead and behind as interaction terms with the treatment indicators, for subgroups defined as being above or below median scores on each respective index.

2.5. Sampling and implementation

We recruited a sample of 664 married or co-habiting couples from coastal villages of Unguja, Zanzibar archipelago's biggest and most populous island, using the following sampling approach. First, we sampled a list of 19 coastal villages covering 5 of the 7 districts in Unguja (see Figure A.1 in Appendix A of Chapter 2 for village locations). Our implementing partners worked with each of the village's respective leaders (a local government official known as the *Sheha*) to select a sample of 40 married couples per village. The only eligibility criterion we enforced was that the couple had at least one child of the age of six years or older but younger than 18 years.[22]

To recruit our sample, the village leader worked together with local mobilisers, calling or visiting each selected couple and inviting them to participate in the study on a particular date and time, at a specified location. During the recruitment discussion, the only information participants received about the study was that it was for research aiming to improve their livelihoods, that they would be compensated TZS 4,000 (∼USD 1.6) for participation, and would have the chance to win some additional money.

On the day of the interview, we arranged for all selected couples who confirmed their participation to assemble at a specific location on a specified date and a specific time. Upon arrival, each couple signed next to their name on an attendance sheet and were given a unique ID. The ID consisted of a number that was the same for both spouses belonging to the same couple, followed by a letter (M for male; F for female). Each spouse was given a sticker with the ID number, which they needed to stick on their chest. It was explained that this ID was unique and allowed us to identify them during the exercise.

Next, they were asked to take a seat and wait for their number to be called for the interview. Spouses of the same couple were interviewed individually and at the same time by a same-gender enumerator in a private space. After the interview, the spouses moved to the experimental stages outlined in Figure 3.1. Between rounds 1 and 2, enumerators wrote down the information that needed to be shared with the other spouse depending on the couple's assigned treatment. This information was passed between enumerators on slips of paper. At the end of both rounds, one of the rounds was randomly selected for implementation. Each respondent was paid in private depending on their performance in the randomly selected round, after which the couple were dismissed.[23] On average, husbands earned TZS 3,960 and wives earned TZS 3,970 (both ∼USD 1.6) in this experiment (the minimum for husbands and wives were TZS 3,000 (∼USD 1.2) and TZS 3,500 (∼USD 1.4), respectively). In addition, each respondent received a show-up fee of TZS 4,000 (∼USD 1.6).

[22] This was relevant for Chapter 2.

[23] Only half the sample was paid based on performances in this experiment. The other half was paid based on choices made in experiment 1, which is presented in Chapter 2.

3. Results

In this section, we present the results of our data analysis. We first offer some summary statistics of socio-demographic characteristics and variables related to spousal performances in the real-effort task (section 3.1). We then present treatments effects (section 3.2), heterogeneity analyses (section 3.3) and robustness checks (section 3.4), respectively.

3.1. Sample

Socio-demographic characteristics

We collected data from a sample of 1,328 respondents. Table 3.1 presents some gender-disaggregated summary statistics of key socio-demographic indicators, indicating p-values of t-tests that compare characteristics between husbands and wives.

The average ages of husbands and wives are 49 and 41 years, respectively. Over 86% of spouses have received at least a primary education, and this does not differ by gender. Husbands on average earn significantly more than women, with TZS 400,000 (∼USD 163) compared to TZS 80,000 (∼USD 32) in a good month, and TZS 158,000 (∼USD 64) compared to TZS 37,000 (∼USD 15) in a bad month. Husbands also contribute the majority of costs (over 80%) for child, personal, and household expenses. Although we would expect to see the same mean years of marriage/cohabitation given the sample is of couples, this is slightly different for husbands (20 years) and wives (19 years). In line with the overall finding in the literature that men are less risk-averse than women, husbands score lower on our risk aversion scale than women. Similarly in line with the competition literature, husbands have (slightly) stronger competition preferences than wives as they score higher on our competitiveness index. Husbands also score higher than wives on our self-efficacy index and marital trust scale. However, contrary to the literature, we find that both spouses appear to have the same average scores on our conflict aversion index.

Since couples were randomly assigned to experimental treatments, we should expect differences in characteristics between treatments to be due to chance, for both husbands and wives. Appendix C presents balancing tables on a range of relevant indicators, including age, education, risk aversion, years of marriage, etc. For wives, we fail to reject the null that the treatments are balanced on all characteristics. For husbands, we fail to reject the null that treatments are balanced on all characteristics except on their responses to how much income is earned in a good month, with those in T4 showing a lower mean monthly income than other treatments. To address this, we include income in the set of controls in the regressions.

Performance statistics

Although there were 1,328 wives and husbands for which data was collected, we removed some observations. First, as timings under 15 seconds or above 150 seconds were likely to be incorrect data entries, we dropped couples with at least one spouse whose Round 1 timing was less than 15 seconds or greater than 150 seconds. We dropped these at the couple-level, as we needed both spouses' Round 1 timings to be able to calculate spousal time differences in Round 1. Together, 28 observations were dropped from the sample. Second, we dropped individuals (not couples) whose Round 2 timings were less than 15 seconds or greater than 150 seconds. This resulted in a further 10 observations being dropped. Third, we dropped 3 observations where timing improvements were greater(less) than 200(-200) seconds and % time improvements were greater(less) than 80%(-80%), again, as this would have been implausible given the task and thus a possible data entry error. With this smaller sample of 1,287 respondents, we present below some gender-disaggregated statistics.

Table 3.1: Summary statistics of socio-demographic characteristics

Variable	(1) Husbands Mean/(SD)	(2) Wives Mean/(SD)	(1)-(2) Pairwise t-test P-value
Age	48.831 (13.291)	41.161 (11.099)	0.000***
Primary education	0.886 (0.319)	0.860 (0.347)	0.162
Monthly income in good month	400.339 (791.455)	81.306 (118.762)	0.000***
Monthly income in bad month	158.142 (359.415)	36.968 (77.877)	0.000***
Years of marriage/cohabitation	20.253 (12.697)	19.017 (12.497)	0.074*
% of personal costs covered	83.554 (24.801)	56.536 (27.092)	0.000***
% of household costs covered	84.834 (16.171)	38.614 (22.247)	0.000***
% of child costs covered	80.648 (17.950)	42.937 (23.256)	0.000***
Attitude to risk	1.125 (1.091)	2.273 (0.933)	0.000***
Marital trust	8.827 (1.542)	7.697 (2.378)	0.000***
Competitiveness index	21.593 (4.777)	19.931 (5.031)	0.000***
Self-efficacy index	17.051 (3.160)	15.074 (2.699)	0.000***
Conflict aversion index	23.407 (4.051)	23.345 (3.456)	0.765
Number of observations	664	664	1328

Notes: Summary statistics of the following variables: 'Age' which captures the respondent's age; 'Primary education' which is a dummy taking the value 1 if they have at least a basic primary education; 'Income in a good/bad month' which is the respondent's estimate of personal monthly income in a good/bad month given in Tanzanian shillings (in 000's); 'Years of marriage' gives the number of years the respondent has been married; '% of child/personal/household costs covered' is the respondent's estimate between 0-100 of how much they contribute to each expense item; 'Attitude to risk' which is a categorical variable capturing how willing the respondent is to take risks, coded as 0='Takes risk a lot', 1='Takes risk but not a lot', 2='Avoids risk but not a lot', and 3='Avoids risk a lot'; 'Marital trust' which is a continuous variable of responses to the question 'On a scale of 1 to 10, where 1 is "not at all" and 10 is "completely", how much do you trust your spouse's judgment when making important decisions that affect both of you?'; 'Competitiveness index' is the sum of the respondent's answers to a set of six Likert-scale questions on their preference for competition, with 6=lowest score for competitive preferences and 30=highest score for competitive preferences; 'Self-efficacy index' is the sum of the respondent's answers to a set of six Likert-scale questions on their self-efficacy beliefs, with 6=lowest self-efficacy score and 30=highest self-efficacy score; 'Conflict aversion index' is the sum of the respondent's answers to a set of six Likert-scale questions on their attitudes to conflict, with 6=lowest conflict aversion score and 30=highest conflict aversion score. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Specifically, in Figure 3.3, we present the gender-disaggregated distributions of timings in Round 1 and Round 2, and in Table 3.2 we present summary statistics on variables related to the task.

In both rounds, the distributions for both wives and husbands approximates a normal distribution when restricting to timings below 60 seconds, with the highest concentration of timings occurring between approximately 25 and 40 seconds (Figure 3.3). In Round 1, the distributions for wives and husbands show a high degree of overlap. In Round 2, both distributions appear to shift slightly to the left (faster timings), suggesting a possible learning effect. Moreover, the wives' distribution appears slightly more concentrated in the 30-second range, while the husbands' distribution shows a marginally higher density than wives above the 40-second mark. Put differently, the husbands' distribution in Round 2 maintains a slightly fatter right tail compared to the wives', indicating a higher frequency of slower completion times within that group.

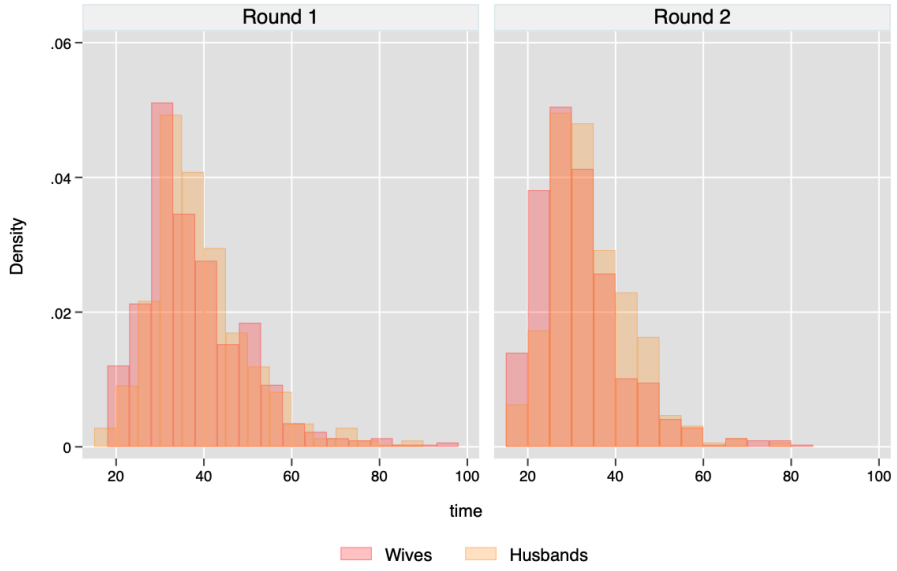


Figure 3.3: Distribution of timings (seconds) in rounds 1 and 2, by gender

This is confirmed by looking at mean performance times in Table 3.2: wives perform better than husbands, with faster average timings in both Round 1 (34 seconds compared to 40 seconds) and Round 2 (31 seconds compared to 38 seconds). It is important to note that a substantial share of observations lies around the 30-second mark in both rounds, suggesting that individuals are potentially targeting the maximum payoff offered at 30 seconds. The potential role of this 30-second threshold is examined more formally in robustness checks below.

Our main outcome variable of interest is the difference in individual performances between rounds i.e., timing improvements. Figure 3.4 presents the gender-disaggregated distributions of timing improvements, calculated as the difference in completion seconds between Round 1 and Round 2, so that a positive value on the x-axis indicates a faster performance in the second round (improvement), while a negative value indicates a slower performance.

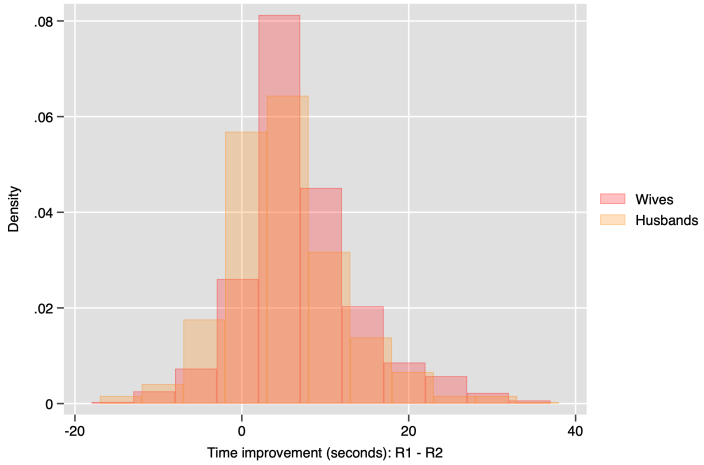


Figure 3.4: Timing improvements between rounds (in seconds), by gender

Table 3.2: Summary statistics from real-effort task

Variable	(1) Male Mean/(SD)	(2) Female Mean/(SD)	(1)-(2) Pairwise t-test P-value
Round 1 (total seconds)	39.586	37.519	0.007***
	(14.377)	(13.177)	
Round 2 (total seconds)	34.366	30.958	0.000***
	(10.658)	(10.650)	
Time improvement	5.219	6.561	0.003***
	(8.673)	(7.741)	
% time improvement	10.988	15.878	0.000***
	(15.831)	(16.064)	
Ahead	4.694	6.841	0.000***
	(8.993)	(12.213)	
Behind	6.930	4.737	0.000***
	(12.283)	(9.054)	
Self-confidence	0.334	0.170	0.000***
	(0.472)	(0.376)	
Overconfidence	0.773	0.809	0.107
	(0.419)	(0.393)	
Underconfidence	0.082	0.045	0.007***
	(0.274)	(0.208)	
Spouse-confidence	0.372	0.139	0.000***
	(0.484)	(0.346)	
Spouse overconfidence	0.736	0.823	0.000***
	(0.441)	(0.382)	
Spouse underconfidence	0.105	0.037	0.000***
	(0.307)	(0.190)	
Positive spousal evaluation	0.263	0.106	0.000***
	(0.440)	(0.308)	
Negative spousal evaluation	0.070	0.308	0.000***
	(0.255)	(0.462)	
Revised self-confidence	0.276	0.151	0.000***
	(0.447)	(0.358)	
Upward revision of self-confidence	0.216	0.124	0.000***
	(0.412)	(0.329)	
Downward revision of self-confidence	0.060	0.027	0.004***
	(0.237)	(0.162)	
Number of observations	647	640	1287

Notes: Summary statistics of the following variables: 'Round 1' gives the total completion time in seconds in Round 1; 'Round 2' gives the total completion time in seconds in Round 2 of the real-effort task; 'Time improvement' gives the absolute time difference in seconds between Round 1 and Round 2; '% time improvement' gives the relative time difference of between Round 1 and Round 2 ; 'Ahead' and 'Behind' are continuous variables capturing the time difference between spouse and respondent when the respondent is faster, and between respondent and spouse when the respondent is slower, respectively (so that they both contain non-negative values), and zero otherwise; 'Self-confidence' is equal to one if a respondent answers with 'Not top person' to the question "If we compare your timing in Round 1 with eight other men/women in the village, will you be the top 1 of 8?", and zero otherwise; 'Overconfidence' is equal to one if a respondent incorrectly believes they are in a particular octile but their performance in Round 1 puts them in a *higher* octile, and zero otherwise; 'Underconfidence' is equal to 1 if a respondent incorrectly believes they are in a particular octile but their performance in Round 1 puts them in a *lower* octile, and zero otherwise; 'Spouse-confidence' is equal to one if a respondent answers with 'Not top person' to the question "If we compare your spouse's timing in Round 1 with eight other men/women in the village, will they be the top 1 of 8?", zero otherwise; 'Spouse over/underconfidence' is equal to one if a respondent incorrectly believes their spouse is in a particular octile but their spouse's performance in Round 1 puts them in a *lower* octile, and zero otherwise; 'Positive spousal evaluation' is equal to one if self-confidence is 1 and the respondent's spouse's response to the spouse-confidence question is 0, zero otherwise; 'Negative spousal evaluation' is equal to one if self-confidence is 0 and the respondent's spouse's spouse-confidence is 1, zero otherwise. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Looking at Figure 3.4, the majority of both wives and husbands sit to the right of zero. This is in line with our expectations that individuals perform better in Round 2 than Round 1, due to both a 'learning effect' as the individual gets better at the task after trying it the first time, and a 'spousal' effect as spouses are made to think about how their performance compares with their spouse's (without observing it). While both

groups peak in the 0 to 10-second improvement range, on average, wives appear to have higher improvements than husbands. Conversely, husbands appear to be closer to 0 or in the negative range than wives. This is confirmed by looking at Table 3.2, where on average, wives also have higher absolute and relative timing improvements than husbands: $\sim$ 6.5 seconds faster, or 16% time improvements between rounds for wives, compared to $\sim$ 5 seconds faster, or 11% time improvements for husbands.

We also report on our variables capturing psychological channels of effort provision: 'Behind' and 'Ahead', which capture the time differences between spouse and respondent, and our confidence measures (please see notes under Table 3.2 for definitions of these variables). Looking at the participants who are ahead, wives are ahead by more ($\sim$ 7 seconds) than husbands are ($\sim$ 5 seconds). For participants who are behind, wives are behind by less ($\sim$ 5 seconds) than husbands are ($\sim$ 7 seconds). Looking at 'self-confidence', which captures respondents' beliefs about their performance ranking in Round 1, a higher proportion of husbands (33%) than wives (17%) report lower self-confidence, rating themselves as unlikely to be the top performer among same-gender peers. The 'underconfidence' variable, which takes into account actual performance rankings, shows that a lower proportion of wives than husbands are underconfident, i.e., they inaccurately believe their performance ranking in Round 1 is worse than it actually is.

Looking at the spouse-confidence variable, we find that a higher proportion of husbands (37%) than wives (14%) say that their spouse will *not* be the top person of eight same-gender participants in their session. In other words, husbands have *less spouse-confidence* in their spouse than wives. Looking at spouse over(under)confidence, which measures whether they inaccurately believe that their spouse will be better(worse) than they actually are, we find that husbands have lower(higher) spouse over(under)confidence than wives.

Looking at spouse evaluations which compare self-confidence with spouse-confidence, we see that wives are less likely to receive a positive spousal evaluation than husbands (10% vs. 26%) and more likely to receive a negative evaluation (30% vs. 7%). In other words, wives are four times more (2.6 times less) likely than husbands to receive a negative (positive) spousal evaluation.

Last, we look at how self-confidence beliefs are revised. As we elicit self-confidence beliefs before Round 1, and then again at the end of Round 2, we are able to track whether there are gender differences between self-confidence revisions, if any. Looking at 'Revised self-confidence', we see that wives are half as likely to revise their belief as husbands (28% vs. 15%). We also find that the revisions for husbands and wives are driven by *upward* revisions, i.e., respondents who change their belief are more likely to believe they rank higher than what they believed before Round 1.

3.2. Treatment comparisons

For the analyses that follow, regression results are presented on wives and husbands, separately. We exclude respondents with a spousal time difference larger than 60 seconds, as we expect treatment differences to be larger when competition is stronger, and we expect competition to be stronger if time differences are not too large, consistent with evidence that competitive effort is greatest when rivals are closely matched (Berger and Pope, 2011; Kilduff, Elfenbein and Staw, 2010; Lazear and Rosen, 1981).[24] This removes 18 observations from the previously cleaned sample of 1,287 respondents, leaving a total of 1,269 respondents, including 631 wives and 638 husbands.Treatment comparisons with the sample with respondents who are ahead/behind their spouse by up to 120 seconds instead of 60 seconds are reported in Appendix G (section G.4) as robustness checks, and discussed in section 3.4.

[24]This also mitigates concerns about regression to the mean—a point we discuss in Section 3.4.

Average treatment effects

To test Hypotheses 3.1-3.4, the regressions follow the empirical model laid out in equation 3.6. Table 3.3 presents the regression results.

Table 3.3: Average treatment effects (gender-disaggregated sample with cutoff=60)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.337	0.790	0.152	0.835	1.215	1.486	0.477	1.167
	(0.729)	(0.790)	(0.677)	(0.780)	(1.572)	(1.968)	(1.444)	(1.924)
T3: Times	0.654	0.557	0.537	0.386	1.870	1.362	1.884	0.654
	(0.823)	(0.944)	(0.771)	(0.883)	(1.649)	(2.295)	(1.608)	(2.173)
T4: Both	0.022	1.270	-0.315	0.964	0.172	3.132	-0.642	2.441
	(0.844)	(0.927)	(0.821)	(0.775)	(1.808)	(2.054)	(1.733)	(1.788)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.26	4.08	6.42	4.19	15.16	9.17	15.55	9.59
p-value: T2 = T4	0.687	0.602	0.549	0.868	0.556	0.400	0.557	0.463
p-value: T3 = T4	0.352	0.334	0.115	0.358	0.299	0.317	0.079	0.249

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Wife and husband regressions have 631 and 638 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Looking at the predicted means under T1 in Table 3.3, we see that the average time improvements between rounds for wives (columns 1 and 3) and husbands (columns 2 and 4) are 6 seconds and 4 seconds, respectively, and average percentage time improvements are 15% for wives (columns 5 and 7) and 9% for husbands (columns 6 and 8). This indicates that there is an important learning effect, which is stronger for wives than for husbands. However, we do not find average treatment effectsn (ATEs) for husbands nor wives, using both absolute and relative timing improvements. Thus, we cannot confirm Hypotheses 3.1, 3.2 or 3.3. We note that an F-test comparing coefficients of T3 and T4 produces $p = 0.079$ for wives (col 7, Table 3.3), which, given the negative T4 coefficient implies that wives' relative timing improvements are *lower* in T4 than T3, contrary to what we expect. We find the same results when we use the sample with the 120 seconds cutoff (Table F.2, Appendix F). Finally, running the same regressions on the pooled sample (Table F.1, Appendix F) yields no effects.

To test Hypothesis 3.4 that treatment effects expected in 3.1–3.3 are larger for husbands than for wives, we interact our treatment indicators with the variable *Wives* which is equal to one if the respondent is the wife, and zero if they are the husband. Table 3.4 presents the results.

Although wives generally exhibit significantly higher time improvements than husbands (cols 1 and 2, significant coeff on *Wives*), the interaction terms for all three treatments are statistically insignificant. This suggests that the timing improvements between T1 and each of the treatments do not differ significantly by gender, thus providing no empirical support for Hypothesis 3.4. To complement the interaction analysis, we examine marginal predictions by treatment and gender (Figure 3.5). We see that wives consistently maintain higher timing improvements across all treatments. Tests of simple effects (contrasts of marginal linear predictions) also reveal that these higher performances are statistically significant in T1 ($p = 0.011$), T3 ($p = 0.026$) and T2 ($p = 0.072$), but are not statistically different from husbands in T4 ($p = 0.124$). These results demonstrate that while absolute gender differences remain statistically significant, they are attenuated in T4.

Last, we also examine ATEs at the couple-level, using the total time improvements of both husband and wife as the dependent variable (Appendix F, Table F.3). As we do not find significant treatment effects, we are unable to confirm Hypothesis 3.5.

Table 3.4: Average treatment effects (pooled sample with cutoff=60)

	(1) Time improvements	(2) % Time improvements
T2: Beliefs	0.790	1.486
	(0.790)	(1.969)
T3: Times	0.557	1.362
	(0.944)	(2.296)
T4: Both	1.270	3.132
	(0.928)	(2.055)
Wives	2.181**	5.993***
	(0.773)	(1.670)
T2: Beliefs × Wives	-0.454	-0.271
	(1.113)	(2.699)
T3: Times × Wives	0.097	0.509
	(1.321)	(2.926)
T4: Both × Wives	-1.248	-2.960
	(1.071)	(2.383)
N	1269	1269

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Model 1) and % timing improvements between rounds (Model 2). N=1,269. Enumerator fixed effects along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

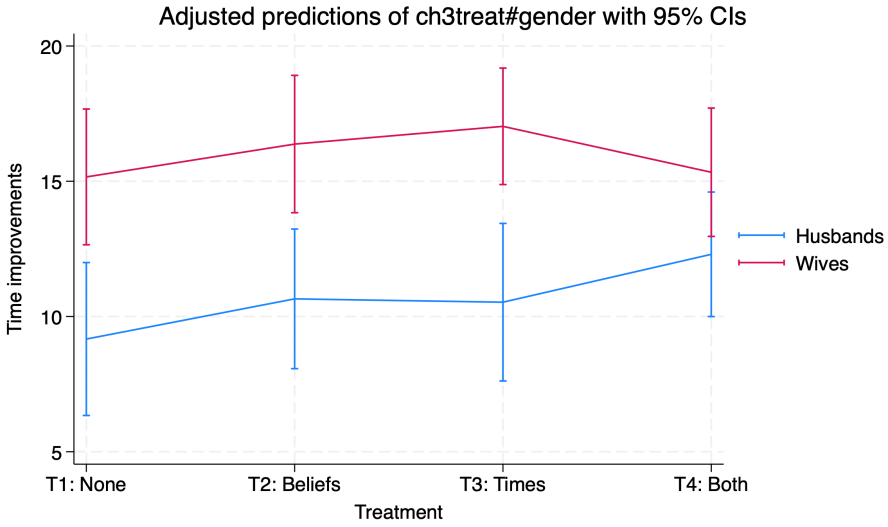


Figure 3.5: Predictive margins in each treatment, by gender

Although we find no average treatment effects, our initial hypotheses were predicated on specific assumptions regarding the psychological channels of increased effort i.e., motivations related to ahead-seeking, behind-aversion, encouragement, and defiance. Thus, it might still be that these channels remain active, but operate in opposing directions, thereby cancelling one another in the aggregate. Furthermore, it may be that only a subset of these theorised channels is active. For example, our initial Hypothesis 3.2 assumes a cumulative effect on increased effort from being behind and ahead. However, if only behind-seeking motivations are strong enough, the total effect size would be too small to be identified. To explore this possibility, the following sections disaggregate the results by relative standing (ahead versus behind) and relative evaluations (positive versus negative).

Spousal comparisons: decomposing ahead and behind channels

In this section, we examine how treatment effects in T3—where comparisons are made possible—vary by how much one lags behind, and by how much one is ahead, of their spouse. To do so, we regress time improvements on the treatment indicators interacted with *Behind* (Table 3.5) and *Ahead* variables (Table 3.6), as defined in section 2.4.

Table 3.5: Treatment effects on timing improvements when behind

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.126	0.550	-0.146	0.416	1.521	1.091	0.606	0.449
	(0.731)	(1.033)	(0.627)	(1.052)	(1.730)	(2.455)	(1.655)	(2.551)
T3: Times	-0.142	-0.543	-0.251	-0.686	1.341	-0.545	1.223	-1.322
	(0.831)	(0.995)	(0.738)	(0.920)	(1.933)	(2.624)	(1.863)	(2.352)
T4: Both	-0.274	0.555	-0.747	0.417	-0.417	2.451	-1.595	2.142
	(0.770)	(0.902)	(0.697)	(0.686)	(1.934)	(2.271)	(1.848)	(1.895)
Behind	0.317***	0.248***	0.298***	0.249***	0.383**	0.310***	0.384**	0.312***
	(0.080)	(0.047)	(0.081)	(0.045)	(0.171)	(0.088)	(0.171)	(0.096)
T2: Beliefs × Behind	0.058	0.030	0.060	0.030	-0.056	0.052	-0.040	0.066
	(0.123)	(0.086)	(0.116)	(0.091)	(0.227)	(0.175)	(0.212)	(0.186)
T3: Times × Behind	0.089	0.242***	0.099	0.227**	0.029	0.403**	0.060	0.402**
	(0.111)	(0.076)	(0.108)	(0.079)	(0.216)	(0.152)	(0.208)	(0.150)
T4: Both × Behind	0.117	0.089	0.142	0.064	0.199	0.079	0.285	0.019
	(0.125)	(0.069)	(0.129)	(0.076)	(0.238)	(0.119)	(0.253)	(0.148)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.28	4.08	6.45	4.25	15.18	9.17	15.59	9.68
p-value: T2 × Behind = T4 × Behind	0.602	0.559	0.482	0.733	0.208	0.877	0.170	0.792
p-value: T3 × Behind = T4 × Behind	0.821	0.117	0.724	0.103	0.502	0.033	0.369	0.013

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are absolute timing improvements between rounds (calculated as timing in Round 1 minus timing in Round 2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We first examine the effect of being behind on time improvements in T3 (Table 3.5). Looking at the coefficients on 'T3 × Behind', we find statistically significant positive coefficients for husbands (columns 2, 4, 6, and 8), which means that the effect of being behind is statistically different between T3 and T1 (e.g., if the husband is behind his wife by 10 seconds in Round 1, he will improve his Round 2 performance by 2.4 seconds or by 40%). We also note the positive and significant *Behind* coefficient, which indicates that even in the control group T1, the improvement between both rounds is larger among respondents (husbands and wives) who are behind. Finally, we also note that there is a statistically significant *decrease* in time improvements for husbands who are behind in T4 compared to those behind in T3 (Table 3.5, cols 7 and 8, two-sided F-tests $p = 0.033$ and $p = 0.013$).

Next, we look at Table 3.6 to examine whether the effect of being ahead differs across treatments. Given that none of the coefficients on 'T3 × Ahead' are significant for husbands nor wives, we cannot confirm ahead-seeking motivations in T3 relative to T1. However, we note marginally significant positive coefficients on 'T2 × Ahead' for wives (columns 1, 2, and 7), even though they do not hear their spouse's timings in this treatment. We also note positive and significant coefficients on treatment indicator T4 in all husband columns, indicating that husbands who are *not ahead* have a larger timing improvement in T4 than in T1. Moreover, we find that the effect of being ahead differs between treatments T4 and T2 for both wives (col 1, two-sided F-test $p = 0.094$) and husbands (col 4, two-sided F-test $p = 0.073$). Specifically, the sign and size of the coefficients indicate that, for wives, the negative effect of being ahead in T2 disappears in T4, while

Table 3.6: Treatment effects on timing improvements when ahead

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.020	0.965	1.020	0.957	2.638	1.959	2.431	1.401
	(0.985)	(1.090)	(0.975)	(0.928)	(2.090)	(2.562)	(2.019)	(2.200)
T3: Times	1.143	1.446	1.029	1.266	3.008	3.648	2.995	2.739
	(1.255)	(1.080)	(1.194)	(1.106)	(2.704)	(2.436)	(2.606)	(2.513)
T4: Both	-0.043	1.920**	-0.002	1.774***	0.324	4.458**	0.706	4.018***
	(1.164)	(0.774)	(1.067)	(0.608)	(2.439)	(1.605)	(2.292)	(1.363)
Ahead	-0.148**	-0.049	-0.135**	-0.049	-0.226	-0.069	-0.178	-0.067
	(0.058)	(0.109)	(0.052)	(0.098)	(0.133)	(0.258)	(0.130)	(0.232)
T2: Beliefs × Ahead	-0.120*	-0.042	-0.138*	-0.025	-0.247	-0.111	-0.316*	-0.047
	(0.067)	(0.149)	(0.069)	(0.127)	(0.159)	(0.373)	(0.166)	(0.309)
T3: Times × Ahead	-0.112	-0.152	-0.096	-0.141	-0.244	-0.401	-0.206	-0.348
	(0.128)	(0.141)	(0.112)	(0.137)	(0.318)	(0.340)	(0.286)	(0.320)
T4: Both × Ahead	0.019	-0.174	-0.039	-0.212**	-0.010	-0.349	-0.200	-0.408
	(0.085)	(0.103)	(0.078)	(0.100)	(0.228)	(0.285)	(0.227)	(0.275)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.28	4.08	6.40	4.16	15.19	9.16	15.50	9.55
F-test: T2 × Ahead=T4 × Ahead	0.094	0.266	0.218	0.073	0.251	0.442	0.560	0.153
F-test: T3 × Ahead=T4 × Ahead	0.191	0.827	0.566	0.442	0.404	0.834	0.982	0.790

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are absolute timing improvements between rounds (calculated as timing in Round 1 minus timing in Round 2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

for husbands the negative effect in T2 becomes stronger in T4. We summarise the findings reported so far in a first result.

Result 1: *Husbands who lag behind their wives in Round 1 improve more in T3 than in T1. Husbands who are not ahead show a stronger performance improvement in T4 than T1. Husbands (wives) who are ahead have stronger (lesser) reductions in time improvements in T4 relative to T2.*

Spousal evaluations: decomposing positive and negative evaluation channels

Next, we examine the influence of positive and negative evaluations of spouses. Recall that these are simply the relative difference between a respondent's self-confidence and their spouse's 'spouse-confidence'. Regression results of time improvements on treatment indicators interacted with *Poseval* and *Negeval* variables (as defined in section 2.4) are presented in Table 3.7.

We first examine the effect of positive evaluations (columns 1-4 of Table 3.7) and negative evaluations (columns 5-8 of Table 3.7) on timing improvements in T2, where just spousal evaluations are shared. Looking at the interaction terms on T2, we do not see any significant effects for husbands or wives. Thus, we cannot confirm that a positive (negative) evaluation would increase (decrease) timing improvements in T2 relative to T1. Moreover, examining the two-sided p-values of F-tests for the equality of interactions with T4 and T3, and T4 and T2, (reported in the lower panel of Table 3.7), we do not find any effects of hearing negative or positive spousal evaluations.

Still, we note that for husbands (columns 6 and 8), interactions of *Negeval* with T3 and T4 are *negative* and statistically significant, even though respondents only hear negative evaluations in the latter treatment. In both cases, the size of the coefficient (together with the predicted mean under T1) indicate that the increase in timings between rounds in T1 is completely undone (and even reversed) in T3 and T4, i.e.,

Table 3.7: Treatment effects on timing improvements from positive/negative evaluations

	Positive evaluations				Negative evaluations			
	Time improvements		% Time improvements		Time improvements		% Time improvements	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.148	0.148	0.610	0.417	-0.078	0.911	-0.006	1.478
	(0.677)	(0.677)	(1.433)	(2.235)	(0.760)	(0.663)	(1.806)	(1.685)
T3: Times	0.249	0.249	1.487	1.725	0.095	1.069	1.305	2.311
	(0.594)	(0.594)	(1.394)	(2.461)	(1.056)	(0.763)	(2.341)	(1.881)
T4: Both	-0.241	-0.241	-0.343	1.308	-0.072	1.308	-0.385	3.368*
	(0.759)	(0.759)	(1.610)	(1.861)	(0.866)	(0.798)	(1.974)	(1.929)
Evaluation	-2.063	-2.063	-3.503	-1.621	0.193	4.369*	-0.329	10.563*
	(2.047)	(2.047)	(4.879)	(3.542)	(1.123)	(2.095)	(2.591)	(5.038)
T2 × Evaluation	-0.572	-0.572	-2.790	1.814	0.816	-4.803	1.693	-9.493
	(2.582)	(2.582)	(5.732)	(3.431)	(1.535)	(2.982)	(3.098)	(6.093)
T3 × Evaluation	-0.154	-0.154	0.228	-1.884	0.336	-8.152**	0.561	-21.943***
	(1.969)	(1.969)	(4.760)	(4.640)	(1.692)	(3.093)	(3.680)	(7.139)
T4 × Evaluation	1.495	1.495	0.974	3.602	-0.108	-6.939**	-0.078	-16.297**
	(2.460)	(2.460)	(6.086)	(3.880)	(1.298)	(2.696)	(3.383)	(6.802)
Predicted mean under T1	6.42	6.42	15.54	9.64	6.44	4.26	15.58	9.71
p-value: T2 × Eval = T4 × Eval	0.381	0.381	0.479	0.557	0.558	0.476	0.622	0.223
p-value: T3 × Eval = T4 × Eval	0.352	0.352	0.862	0.170	0.729	0.594	0.828	0.284

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are absolute timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(2) and (5)-(6)), and % timing improvements between rounds (Models (3)-(4) and (7)-(8)). Relative confidence is a dummy variable for positive evaluations in Models (1)-(4), and a dummy for negative evaluations in Models (5)-(8). The positive evaluation dummy takes the value 1 if the self-confidence dummy is 1 (respondent believes they are not top person) whilst the spousal evaluation dummy is 0 (spouse believes the respondent is the top person), zero otherwise. The negative evaluation dummy takes the value 1 if the self-confidence dummy is 0 (respondent believes they are the top person) whilst the spousal evaluation dummy is 1 (spouse believes the respondent is not top person), zero otherwise. Enumerator fixed effects are included in all models, along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. We use the 'Ahead' and 'Behind' continuous variables as further controls. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

husbands who receive a negative evaluation in both treatments actually do *worse* in Round 2 than in Round 1. We summarise these new findings in a second result.

Result 2: *When wives evaluate them negatively, husbands improve less in Round 2 in T3 and T4 than in T1. Wives show no such effect if husbands evaluate them negatively. We find no evidence of positive evaluations as a motivational channel for either spouse.*

Combined effects: decomposing ahead/behind and positive/negative evaluation channels

In a final step, we examine the combined effects of respondents learning in T4 that they are ahead or behind and have received a positive or negative evaluation from their spouse. To do so, we create a new set of variables for the different combinations, namely, *Ahead and Poseval*, *Ahead and Negeval*, *Behind and Poseval*, and *Behind and Negeval* (for definitions please refer to section 2.4). Note, given we hypothesise that all effects present a motivational boost on effort, we would expect reinforcing effects in T4. For example, if a respondent learns they are ahead and receives a positive evaluation (*Ahead and Poseval*), we would expect them to increase their timing improvements in T4 compared to T1, as they would be incentivised by both ahead-seeking and encouragement motivations.

Table 3.8 presents the regression results of *Ahead and Poseval* and *Ahead and Negeval* interacted with our treatment indicators. Looking at their interactions with T4—where respondents receive both types of feedback—we find *negative* and significant coefficients on T4 × *Ahead and Poseval* for wives (columns 1, 3, 5 and 7). This indicates that, contrary to expected, wives who are told they are ahead and who also receive a positive evaluation in T4 have *smaller* timing improvements than those who are ahead and have a positive evaluation (but who do not know this) in T1. For husbands, whilst there are no significant effects for T4 ×

Ahead and Poseval, we find significant negative interaction terms on T4 × *Ahead and Negeval* (last row in columns 2, 4, and 8). That is, husbands who are told they are ahead and who receive a negative evaluation in T4 are likely to have *smaller* timing improvements than those who are ahead with a negative evaluation (but who do not know this) in T1. This indicates that, not only do negative evaluations affect husbands' effort in the opposite direction to what is expected, but that the *Negeval* effect is *stronger* than ahead-seeking motivations. Put differently, hearing a negative evaluation appears to increase the marginal cost of effort more than being ahead increases the marginal utility of higher effort. It is worth noting that in the case of wives, there are also significant and negative coefficients on T3 × *Ahead and Positive* (cols 1, 3, and 7) and on T2 × *Ahead and Positive* (cols 1 and 3), even though they only receive one type of information in these treatments. We return to this point in the discussion section.

Table 3.8: Treatment effects on effort, with interactions of 'Ahead and Positive/Negative'

	Time improvements			% Time improvements				
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.549	0.961	0.458	1.039	1.704	1.900	1.216	1.617
	(0.773)	(0.874)	(0.764)	(0.835)	(1.677)	(2.098)	(1.631)	(2.007)
T3: Times	0.859	0.974	0.745	0.878	2.383	2.518	2.487	1.921
	(0.921)	(0.895)	(0.826)	(0.862)	(1.912)	(2.113)	(1.779)	(2.042)
T4: Both	0.143	1.697*	-0.272	1.387*	0.650	4.019**	-0.137	3.292*
	(0.853)	(0.859)	(0.857)	(0.730)	(1.654)	(1.874)	(1.708)	(1.656)
Ahead and Positive	0.021	0.148	-0.050	0.109	0.168***	0.358	0.060	0.271
	(0.023)	(0.207)	(0.037)	(0.197)	(0.052)	(0.482)	(0.075)	(0.473)
T2: Beliefs × Ahead and Positive	-0.352**	-0.087	-0.394*	-0.113	-0.363	-0.220	-0.369	-0.222
	(0.165)	(0.205)	(0.205)	(0.190)	(0.456)	(0.538)	(0.529)	(0.501)
T3: Times × Ahead and Positive	-0.255***	-0.256	-0.388***	-0.242	-0.320	-0.651	-0.609**	-0.572
	(0.057)	(0.238)	(0.086)	(0.232)	(0.254)	(0.597)	(0.234)	(0.600)
T4: Both × Ahead and Positive	-0.526***	-0.271	-0.598***	-0.228	-1.687***	-0.508	-1.899***	-0.391
	(0.165)	(0.197)	(0.156)	(0.190)	(0.472)	(0.470)	(0.490)	(0.460)
Ahead and Negative	-0.113	-0.035	-0.112	0.027	-0.230	-0.084	-0.190	0.141
	(0.077)	(0.128)	(0.069)	(0.147)	(0.211)	(0.296)	(0.170)	(0.329)
T2: Beliefs × Ahead and Negative	-0.073	-0.098	-0.120	-0.108	-0.176	-0.221	-0.290	-0.331
	(0.076)	(0.151)	(0.080)	(0.170)	(0.210)	(0.354)	(0.216)	(0.370)
T3: Times × Ahead and Negative	-0.147	-0.062	-0.140	-0.163	-0.394	-0.294	-0.394	-0.634
	(0.160)	(0.183)	(0.140)	(0.190)	(0.429)	(0.401)	(0.385)	(0.409)
T4: Both × Ahead and Negative	0.051	-0.296*	0.046	-0.422**	0.166	-0.683	0.081	-0.968**
	(0.133)	(0.161)	(0.110)	(0.185)	(0.410)	(0.403)	(0.311)	(0.432)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.31	4.10	6.51	4.20	15.24	9.22	15.69	9.66

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead and Positive' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9 presents the regression results of the interactions with the *Behind and Poseval* and *Behind and Negeval* variables. Looking at our T4 interaction terms, a few results stand out. We find positive and significant coefficients of T4 × *Behind and Positive* for wives (columns 1, 3, 5, and 7) and husbands (columns 2 and 4), which, in line with what we would expect, tells us that both wives and husbands who learn they are behind and receive a positive evaluation from their spouse significantly improve their timings. However looking at T4 × *Behind and Negative*, we find no significant effects for either gender, which suggests that when they are behind, they are not affected by negative evaluations in the same way as they are by positive evaluations. One surprising effect worth noting is the significant coefficient on T3 × *Behind and Negative* for husbands (cols 2, 4, 6 and 8), even though husbands only receive timing comparisons and not evaluations in this treatment. We return to this point in the discussion section.

Table 3.9: Treatment effects on effort, with interactions of 'Behind and Positive/Negative'

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.023	0.645	-0.151	0.578	0.478	1.204	-0.089	0.658
	(0.761)	(0.836)	(0.725)	(0.788)	(1.668)	(2.073)	(1.591)	(1.968)
T3: Times	0.028	0.170	-0.057	-0.108	0.940	1.092	0.974	0.095
	(0.990)	(0.901)	(0.939)	(0.838)	(1.982)	(2.287)	(1.897)	(2.118)
T4: Both	-0.424	1.179	-0.630	0.902	-0.633	3.121	-1.195	2.384
	(0.906)	(0.893)	(0.895)	(0.683)	(1.963)	(2.044)	(1.910)	(1.680)
Behind and Positive	0.003	0.188**	-0.048	0.143	-0.135	0.218	-0.074	0.103
	(0.096)	(0.074)	(0.121)	(0.090)	(0.188)	(0.127)	(0.231)	(0.158)
T2: Beliefs × Behind and Positive	0.309	0.184	0.271	0.200	0.494	0.368	0.242	0.394*
	(0.252)	(0.130)	(0.268)	(0.120)	(0.506)	(0.216)	(0.503)	(0.192)
T3: Times × Behind and Positive	0.265	0.327***	0.155	0.349***	0.553	0.424**	0.312	0.515***
	(0.299)	(0.102)	(0.324)	(0.112)	(0.636)	(0.153)	(0.705)	(0.175)
T4: Both × Behind and Positive	0.582***	0.162*	0.590***	0.175**	0.951***	0.192	0.850***	0.229
	(0.200)	(0.079)	(0.193)	(0.078)	(0.326)	(0.155)	(0.285)	(0.148)
Behind and Negative	0.231**	0.436***	0.257**	0.416***	0.226	0.858***	0.341*	0.752**
	(0.094)	(0.102)	(0.092)	(0.122)	(0.176)	(0.235)	(0.178)	(0.267)
T2: Beliefs × Behind and Negative	0.341	-0.417*	0.225	-0.400	0.685	-0.875*	0.566	-0.708
	(0.232)	(0.229)	(0.281)	(0.278)	(0.409)	(0.433)	(0.511)	(0.502)
T3: Times × Behind and Negative	0.160	-1.234***	0.141	-1.059***	0.254	-2.830***	0.232	-2.362***
	(0.124)	(0.134)	(0.124)	(0.163)	(0.201)	(0.342)	(0.201)	(0.382)
T4: Both × Behind and Negative	0.040	-0.359	-0.054	-0.374	0.147	-0.677	0.019	-0.602
	(0.304)	(0.227)	(0.344)	(0.231)	(0.592)	(0.573)	(0.675)	(0.517)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.23	4.06	6.40	4.19	15.10	9.14	15.52	9.59

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind and Positive' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 3: *Wives who are ahead (behind) and receive a positive evaluation show lower (higher) performance improvements in T4 relative to T1. Husbands who are ahead (behind) and receive a negative (positive) evaluation show lower (higher) performance improvements in T4 relative to T1. Finally, relative to T1, husbands who are behind in T3 will show a larger (smaller) improvements if their wife evaluates them positively (negatively), even if they do not hear these evaluations. Similarly, wives who are ahead with positive evaluations will have lower timing improvements in T2 and T3, relative to T1, even though they only receive one type of information.*

3.3. Heterogeneity

We next explore heterogeneity in the main regression findings across three spousal characteristics: competitive preferences, marital trust, and conflict aversion. These three characteristics are selected because they map directly onto the mechanisms central to our design: competitive preferences capture individual propensity to respond to relative performance information; marital trust captures whether spouses process evaluative feedback as supportive or threatening; and conflict aversion captures sensitivity to interpersonal friction arising from outperforming one's partner.

For the heterogeneity analysis, we re-run the treatment analysis using *Ahead* and *Behind* interacted with the treatment indicators, for subgroups defined as being above or below median scores on each respective index. We focus on the ahead/behind distinction instead of positive/negative evaluations, as we are particularly interested in examining how these characteristics modulate the level of spousal competition.

Table 3.10: Timings interacted with competitiveness index

	Ahead				Behind			
	More competitive		Less competitive		More competitive		Less competitive	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.646	-0.236	0.719	1.650	0.905	-0.126	-0.417	1.142
	(1.362)	(1.277)	(1.556)	(1.507)	(0.869)	(1.377)	(0.932)	(1.661)
T3: Times	0.413	0.251	1.463	1.720	-0.785	-0.949	0.246	-0.531
	(1.826)	(1.172)	(1.734)	(1.397)	(1.385)	(1.302)	(1.108)	(1.453)
T4: Both	-1.853	-0.027	0.634	3.204**	-2.855***	-0.069	0.969	0.566
	(1.659)	(1.207)	(1.491)	(1.198)	(0.988)	(1.016)	(1.124)	(1.399)
Ahead/Behind	-0.135	-0.309***	-0.164*	0.059	0.359***	0.271***	0.305***	0.212**
	(0.083)	(0.076)	(0.081)	(0.131)	(0.089)	(0.043)	(0.102)	(0.096)
T2 × Ahead/Behind	-0.186*	0.238*	-0.083	-0.206	-0.086	0.096	0.141	-0.069
	(0.103)	(0.125)	(0.104)	(0.196)	(0.120)	(0.096)	(0.152)	(0.179)
T3 × Ahead/Behind	-0.060	0.068	-0.124	-0.217	0.177	0.232*	0.059	0.245*
	(0.176)	(0.117)	(0.158)	(0.188)	(0.181)	(0.113)	(0.149)	(0.135)
T4 × Ahead/Behind	0.020	0.040	0.050	-0.250*	0.297*	-0.005	0.051	0.208*
	(0.122)	(0.118)	(0.089)	(0.129)	(0.168)	(0.079)	(0.126)	(0.106)
N	226	290	405	348	226	290	405	348

Notes: OLS regressions. Dependent variable is timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(8)). Models (1)-(4) interact treatments with 'Ahead'—a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), zero otherwise. Models (5)-(8) interact treatments with 'Behind'—a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), zero otherwise. To explore heterogeneity by competitiveness, we split the sample at the median of the 'Competitiveness index' constructed as the sum of six Likert-scale items on preference for competition (ranging from 6 = lowest to 30 = highest). The median is calculated over the full sample. Models (1) and (5) include wives above the median; Models (2) and (6) include husbands above the median; Models (3) and (7) include wives below the median; and Models (4) and (8) include husbands below the median. No enumerator fixed effects or controls are used in any models. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10 presents interactions of treatments with *Ahead* and *Behind* for subgroups which are above/below median competitiveness measured using a 'competitiveness index'. This is equal to the sum of the respondent's answers to a set of six Likert-scale questions on their preference for competition, with 6 being lowest score for competitive preferences and 30 being the highest score. We first look at respondents who are ahead (columns 1-4), and compare these results with results reported in Table 3.6. We observe that the reduced time improvements in T4 for husbands who are ahead, appear to be driven by husbands who are below median competitiveness (column 4). For wives, the reduced time improvements in T2 are driven by wives who are above median competitiveness (column 1).

Next, we look at respondents who are behind (columns 5-8), and compare these results with the results in Table 3.5. We note that the increased time improvements in T3 for husbands are observed in husbands who are both above and below median competitiveness, suggesting that treatment effects are similar for both groups of husbands (columns 6 and 8). The reduced time improvements in T4 appear to be driven by husbands who are below median competitiveness (column 8). For wives, we note that the increased time improvements in T2 are driven by wives who are above median competitiveness (column 5). Moreover, column 5 tells us that for wives who are more competitive, being in T4 when not behind leads to lower time improvements relative to T1, and this T4 treatment effect is reduced if she is behind. We summarise the effects of competitiveness in the following way:

Result 4: *For wives, the impact of competition is primarily driven by those above median competitiveness, showing reduced (increased) time improvements in T2 (T4) when ahead (behind). For husbands, treatment effects show mixed dependence on competitiveness, with reduced (increased) time improvements driven by those below median competitiveness and ahead (not ahead) in T4.*

Table 3.11: Timings interacted with marital trust index

	Ahead				Behind			
	High trust		Low trust		High trust		Low trust	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.662	1.860	0.785	0.008	0.877	1.842	-0.321	-0.676
	(1.959)	(1.318)	(1.199)	(1.172)	(1.282)	(1.592)	(1.126)	(1.212)
T3: Times	1.881	3.632**	0.810	-0.473	0.953	0.508	-0.749	-1.766
	(1.642)	(1.718)	(1.464)	(1.312)	(1.696)	(1.472)	(1.118)	(1.380)
T4: Both	-0.775	3.234**	0.329	0.733	0.060	2.203*	-0.524	-0.961
	(1.863)	(1.151)	(1.275)	(1.420)	(1.472)	(1.077)	(0.945)	(1.314)
Ahead/Behind	-0.162**	-0.014	-0.136	-0.088	0.413***	0.240***	0.254**	0.231***
	(0.072)	(0.060)	(0.086)	(0.176)	(0.093)	(0.068)	(0.106)	(0.065)
T2 × Ahead/Behind	-0.106	-0.009	-0.138	-0.068	-0.009	-0.051	0.101	0.114
	(0.121)	(0.167)	(0.085)	(0.156)	(0.164)	(0.155)	(0.158)	(0.090)
T3 × Ahead/Behind	-0.202	-0.170	-0.097	-0.124	0.006	0.265**	0.139	0.260**
	(0.206)	(0.108)	(0.155)	(0.216)	(0.204)	(0.118)	(0.143)	(0.104)
T4 × Ahead/Behind	0.068	-0.154*	-0.013	-0.198	-0.010	0.019	0.191	0.162*
	(0.114)	(0.085)	(0.102)	(0.178)	(0.245)	(0.082)	(0.126)	(0.083)
N	210	287	421	351	210	287	421	351

Notes: OLS regressions. Dependent variables is timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(8)). Models (1)-(4) interact treatments with 'Ahead'—a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), zero otherwise. Models (5)-(8) interact treatments with 'Behind'—a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), zero otherwise. To explore heterogeneity by marital trust, we split the sample at the median of the 'Marital trust index' which is a continuous variable of responses to the question 'On a scale of 1 to 10, where 1 is "not at all" and 10 is "completely," how much do you trust your spouse's judgment when making important decisions that affect both of you?'. The median is calculated over the full sample. Models (1) and (5) include wives above the median; Models (2) and (6) include husbands above the median; Models (3) and (7) include wives below the median; and Models (4) and (8) include husbands below the median. No enumerator fixed effects or controls are used in any models. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11 presents interactions of treatments with *Ahead* and *Behind* for subgroups which score above/below the median of 'Marital trust', a continuous variable of responses to the question 'On a scale of 1 to 10, where 1 is "not at all" and 10 is "completely", how much do you trust your spouse's judgment when making important decisions that affect both of you?'. For respondents who are ahead (columns 1-4), comparing these results with Table 3.6, we see that the reduced time improvements in T4 for husbands who are ahead, appear to be driven by husbands who score above median marital trust (column 2). For respondents who are behind (columns 5-8), comparing these results with Table 3.5, we see that the increased time improvements in T3 for husbands are found in both groups where husbands are above and below median marital trust (columns 6 and 8)

Result 5: *Reduced time improvements in T4 for husbands who are ahead appear to be driven by husbands with above-median trust scores. Marital trust does not seem to moderate treatment effects for wives.*

Last, we examine Table 3.12 which presents interactions of treatments with *Ahead* and *Behind* for subgroups which score above/below the median of a 'Conflict aversion index', constructed as the sum of the respondent's answers a set of six Likert-scale questions on their attitudes to conflict, with 6 being the lowest conflict aversion score and 30 being the highest conflict aversion score. The most striking result is found among wives who have above median conflict aversion (column 5). For this subgroup, wives who are not behind in T4 show a considerable increase in their timing improvements, while those who are behind in T4 see these treatment effects reduce significantly. The opposite is true for wives who are below median conflict aversion (column 7): here, the wives are not behind their spouse have *lower* timing improvements in T4 relative to T1. This lowering effect is reduced if wives are behind. Put differently, among wives who are highly conflict averse (column 5), being in treatment T4 seems to motivate greater effort, but only when they are not behind their spouse. However, when they are behind, the same treatment effect diminishes sharply,

Table 3.12: Timings interacted with conflict aversion index

	Ahead				Behind			
	High conflict aversion		Low conflict aversion		High conflict aversion		Low conflict aversion	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	4.892**	1.170	-0.799	0.977	3.385***	1.198	-1.242	0.134
	(2.103)	(1.188)	(0.937)	(1.515)	(0.908)	(1.058)	(1.058)	(1.471)
T3: Times	2.340	1.206	0.592	1.809	1.994	-1.170	-1.170	-0.289
	(2.503)	(1.414)	(1.334)	(1.339)	(1.160)	(1.025)	(1.137)	(1.408)
T4: Both	1.566	1.835**	-0.813	1.955	3.555***	0.528	-2.152**	0.524
	(1.933)	(0.871)	(1.385)	(1.238)	(0.988)	(0.727)	(0.870)	(1.517)
Ahead/Behind	-0.204**	-0.151*	-0.114*	0.011	0.576***	0.235***	0.225**	0.255***
	(0.093)	(0.077)	(0.054)	(0.141)	(0.112)	(0.059)	(0.081)	(0.072)
T2 × Ahead/Behind	-0.211	-0.115	-0.089	-0.064	-0.087	-0.022	0.029	0.053
	(0.162)	(0.140)	(0.057)	(0.174)	(0.113)	(0.140)	(0.139)	(0.099)
T3 × Ahead/Behind	-0.113	-0.235	-0.129	-0.181	-0.142	0.365**	0.167	0.187
	(0.207)	(0.173)	(0.116)	(0.154)	(0.209)	(0.141)	(0.125)	(0.110)
T4 × Ahead/Behind	0.078	-0.151	-0.013	-0.197	-0.364***	0.128**	0.350**	0.064
	(0.120)	(0.105)	(0.093)	(0.132)	(0.122)	(0.056)	(0.148)	(0.116)
N	195	246	436	392	195	246	436	392

Notes: OLS regressions. Dependent variables is timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(8)). Models (1)-(4) interact treatments with 'Ahead'—a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), zero otherwise. Models (5)-(8) interact treatments with 'Behind'—a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), zero otherwise. To explore heterogeneity by conflict aversion, we split the sample at the median of the 'Conflict aversion index' constructed as the sum of six Likert-scale items on preference for competition (ranging from 6 = lowest to 30 = highest). The median is calculated over the full sample. Models (1) and (5) include wives above the median; Models (2) and (6) include husbands above the median; Models (3) and (7) include wives below the median; and Models (4) and (8) include husbands below the median. No enumerator fixed effects or controls are used in any models. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

suggesting that falling behind may make them withdraw effort or avoid further competition to reduce the risk of interpersonal friction. In contrast, among wives who are less conflict averse (column 7), the pattern flips. When these less conflict-averse wives are not behind, they actually slow down (timing improvements drop relative to the control), suggesting they are less driven by concern over outperforming their spouse and perhaps more comfortable with competitive tension. When they are behind, this negative effect lessens, implying they might even enjoy catching up or see it as a challenge rather than a source of discomfort.

Result 6: *For wives, increased time improvements in T4 for wives who are not behind (who are behind) are driven by wives above (below) median conflict aversion.*

3.4. Robustness checks

To verify that the main results are not driven by specific modelling choices or sample restrictions, we conducted a series of robustness checks. Reported in this section are a summary of our results from these robustness checks (presented in Appendix G).

The first of these checks are the regressions where separate interaction terms are combined in the same model (Appendix G, section G.1). First, we combine both 'ahead' and 'behind' in one model, that is, combining both interaction terms from Tables 3.6 and 3.5 (reported in Table G.1). We note that the observed treatment effects in T3 for husbands who are behind in Table 3.5 are still found with this combined model. Second, we combine into one model all interaction terms from Tables 3.8 and 3.9: *Ahead and Positive*, *Ahead and Negative*, *Behind and Positive*, and *Behind and Negative* (reported in Table G.2). Here too, the most significant treatment effects from Tables 3.8 and 3.9 remain significant. For instance, wives (husbands) in T4 who are behind and receive a positive (negative) evaluation still show increased (reduced) time improvements, relative to T1.

In a second set of checks, we use two alternative variables instead of relative spousal evaluations between self and spouse. The first variable is respondents' own answers to spouse-confidence beliefs (Appendix G, section G.2). That is, instead of what the spouse says about the respondent, we use the variable which captures what the respondent says about the spouse. We use this variable as a robustness check, as it might be the case that instead of focusing on what their spouse says, respondents instead anchor their performance on what they themselves stated. Using this variable, we detect (previously undetected) treatment effects in T2 for wives: those who believe their spouse is *not* the top person have higher timing improvements in T2 relative to those in T1 (columns 1 and 3 of Table G.3). The second variable is respondents' absolute spousal evaluations which is simply what the spouse says about the respondent (Appendix G, section G.3). Looking at Table G.4, we find the same negative T3 treatment effects for husbands that we detected in Table 3.7.

Next, we look at regressions with the unrestricted sample of respondents whose Round 1 performance was less than 120 seconds, instead of less than 60 seconds (Appendix G, section G.4), and regressions where we use a binary version of the 'behind' and 'ahead' variables (Appendix G, section G.5). Findings for 'T3 × Behind' found in Table 3.5 for husbands are no longer significant in both checks (Tables G.6 and G.13), although combined effects are still consistently recorded for wives in T4 (positive 'T4 × Behind and Positive' and husbands in T3 (negative 'T3 × Behind and Negative') in Tables G.10, G.11, G.16 and G.17.

We also run the same regressions but where standard errors are estimated with bootstrapping (Appendix G, section G.6). Here, we follow Cameron, Gelbach and Miller (2008) who show that bootstrapping gives more accurate standard errors than cluster-robust standard errors if the number of clusters is low (less than 50), as in our case. As a last robustness check, we also report regressions estimated by fitting seemingly unrelated regression (SUREG) models with village cluster standard errors (Appendix G, section G.7). This approach allows us to simultaneously estimate the choices of wives and husbands, taking into account possible correlations in the unobserved factors that might influence the choices of both partners. Both these specifications show that the results using OLS remain unchanged across alternative specifications.

Because individuals participated in two experiments, we also test for potential order effects and cross-experiment spillovers. To examine order effects, we estimate a regression which interacts our Chapter 3 treatment indicators with *CH3 second*, a binary variable which takes the value one if the experiment of Chapter 3 was administered after the experiment of Chapter 2 (Appendix G, Table G.28, cols 1-4). Spillover effects are tested by regressing the outcome indicator (effort) in Chapter 3 on the treatment indicators of Chapter 2 (Appendix G, Table G.28). Across all specifications, we find no statistically significant interaction terms using our *CH3 second* variable, nor do we find any of the Chapter 2 treatment indicators to be statistically significant. This confirms that the results reported in Chapters 2 and 3 can be interpreted as independent behavioural responses.

As a final robustness check, we examine whether the 30-second incentive threshold—where the marginal earnings to effort drop to zero—distorts or moderates treatment effects. First, we construct an alternative binary outcome equal to one if Round 2 of the task is completed within 30 seconds. Re-estimating our main specifications using this binary variable (while controlling for an equivalent binary measure for Round 1 performance), we find that the treatment effects are largely insignificant (see Appendix G, Table G.29). The only exception is a significant interaction for husbands in T3 who were behind their wives in Round 1, who become slightly more likely to finish under 30 seconds. The general lack of significance in these binary models suggests that the informational treatments did not function by pushing marginal participants across the payoff threshold. This interpretation is supported by a visual inspection of the timing distributions across rounds and treatments (G, Table G.1). In Round 1, whilst a baseline 'spike' at the 30-second mark is visible, the distributions are relatively dispersed. Following the intervention, the Round 2 distributions exhibit two key changes: a general shift toward the 30-second threshold and a marked increase in concentration (taller peaks) around this target.

Crucially, this tightening of the distribution is evident both near the threshold and across the tails of the distribution. This suggests that the informational feedback facilitated greater timing precision and calibration across the entire sample. The fact that participants moved toward the incentive in a continuous, precise manner—rather than a simple binary jump—supports the idea that the treatments improved task-specific information and effort calibration, rather than triggering a narrow response to the threshold itself.

We also re-run our original regressions restricting the sample to those who took over 30 seconds to complete the task in Round 1 (see Appendix G, Table G.30). By excluding participants who had already reached the maximum payoff at baseline, we isolate the group for whom the 30-second incentive remained a relevant target. Our main findings remain robust in this sub-sample. Taken together, these results indicate that our findings represent a genuine response to the informational feedback and are not driven by idiosyncratic effort responses to the 30-second threshold.

It is also worth discussing whether there was a possibility of time improvements displaying regression to the mean. That is, whether treatment effects were driven by instances where large positive/negative time improvements between rounds were the result of extremely low/high values due to chance in Round 1—i.e., random outliers—and values closer to the mean in Round 2. We are less concerned with this possibility for a few reasons. For one, we removed extreme outliers in timings of both rounds 1 and 2, to ensure that time improvements remained meaningful within the context of the task. Of course, it is still possible that Round 1 performances even within our limits were due to random chance. However, as we randomised assignment to treatments, we would expect this to occur across *all* treatments at the same rate. Thus, we are confident that any observed treatment differences would be due to the treatments themselves and hence, that the results represent genuine gendered behavioural responses to spousal comparison and evaluation rather than statistical regression to the mean.

4. Discussion

In this section, we revisit our results in relation to our predictions from the simple model we developed, and in relation to alternative mechanisms. We also comment on the overall external validity of the experiment.

4.1. Revisiting treatment effects

Spousal evaluations

Our first hypothesis was that spousal evaluations (defined as the relative difference between the respondent's self-confidence and their spouse's spouse-confidence) would affect effort if made known to spouses. Specifically, positive and negative evaluations would lead to higher timing improvements in T2 relative to T1 (albeit through two different psychological channels). However, we do not find average treatment effects of spousal evaluations (Table 3.3), nor of positive and negative evaluations specifically (Table 3.7), in T2 relative to T1.

One possible explanation for the absence of treatment effects from spousal evaluations might be that the respondent focuses more on what they themselves predicted about their spouse's performance, rather than what their spouse predicted about them. For example, if a respondent declares that they believe their spouse is *not* the top person, expressing this spouse-confidence belief might set a reference point that motivates their own performance. Evidence to support this explanation is found in the regressions run as robustness checks, where the respondent's own spouse-confidence beliefs are interacted with treatment indicators (Table G.3). Here, spouse-confidence beliefs are correlated with timing improvements in T2. Specifically, wives in T2 who state that their spouse is *not* the top person have higher timing improvements in T2 relative to those in T1 (Table G.3, cols 1 and 3).

It is important to note that although we do not find treatment effects of evaluations on timings in T2, we do find them in T3 (where they do not receive spouse evaluations). Specifically, husbands with a negative evaluation in T3 have significantly *lower*[25] time improvements relative to T1. One explanation for this effect appearing in the wrong treatment relates to the link between competition and evaluation. As spousal comparisons in T3 induce a competitive environment, it may be that husbands form beliefs about what their spouse thinks of them. These beliefs are also likely to be accurate, as spouses know how they evaluate each other in other instances, which they might extrapolate to the task. For example, if a man's wife is generally negative about his ability, he likely knows that she will evaluate him negatively in this task, even though it is not communicated to him. This explanation is supported by looking at the subgroup analysis of 'Behind and Positive/Negative' (Table 3.9). Here, the negative coefficient on T3 for husbands remains only for the subgroup of *Behind and Negative*, whilst the coefficient is *positive* for the *Behind and Positive* group in T3 (columns 2, 4, 6, 8). Additionally, we find a statistically significant difference between these two T3 interaction term coefficients (F=93.82, $p = 0.0000$). Taken together, these results suggest that although husbands do not hear spousal evaluations in T3, hearing that they are behind their wife prompts them to form a belief around what their partner thinks of them. Moreover, these beliefs appear to be accurate (since effort moves in the expected direction to receiving a negative or positive evaluation if we assume the Golem effect[26] and encouragement effect, respectively).

Spousal comparisons

Our second hypothesis was that respondents who received spousal comparison feedback would have larger timing improvements in treatment T3, compared to those in T1. This was based on the expectation that both ahead-seeking and behind-averse motivations would increase the optimal effort compared to T1. Although we do not find average treatment effects of spousal comparisons on effort (Appendix F), we do find some evidence of its psychological channels, which we examine in turn.

On behind-averse behaviour in T3, where spousal comparisons become effective, we find significant results for its influence on spousal effort, but only for husbands (Table 3.5, columns 2, 4, 6 and 8). Not finding evidence of behind aversion in wives contrasts with the evidence of Dasgupta et al. (2019) that behind-averse women are more willing to compete than behind-averse men. We posit that the spousal relationship may be a key factor explaining the lack of behind-aversion in wives, i.e., *who* the respondent lags behind seems to matter in an individual's willingness to compete: even if a wife is fundamentally behind-averse, when the opponent is their spouse, she may be less inclined to ramp up her effort in the competition compared to a behind-averse husband.

On ahead-seeking behaviour, our results do not detect its influence in T3 for either gender. However, we do find evidence of *ahead-averse* behaviour in wives in T1 and T2, even though they do not receive feedback of being ahead in either treatment (Table 3.6 columns 1 and 3). There may be two explanations for this. First, recall that spouses in all treatments learn that their spouse is doing the same task in the other room prior to Round 2. It might then be that merely knowing this already creates a (weak) competitive environment between spouses, even in treatments where timings are not shared. In this situation, it is plausible that wives adjust effort in other treatments as if they are in competition with their spouse. Second, although we would expect competition to increase ahead-seeking behaviour, observing ahead-averse behaviour in wives might be explained by a wider finding in the literature on social preferences, which observes that wives display

[25]The direction of negative evaluations on effort runs counter to what we expect. However, this might be explained by a well-known psychological channel known as the 'Golem effect', where low expectations placed on an individual can lead to decreased motivation and performance (Babad, Inbar and Rosenthal, 1982; Davidson and Eden, 2000; Oz and Eden, 1994).

[26]See footnote 25.

stronger inequality aversion compared to husbands (Engel, 2011). In a setting where the opponent is the spouse, a wife who is ahead (even though she does not know this[27]) might reduce effort when she knows that performances are comparable at the end of the game, in order to reduce potential inequality between spouses. This possibility is also supported by Bursztyn, Fujiwara and Pallais (2017)'s finding that women reduce ambition when observed by potential partners.

Combined effects

Looking at T4, where information is shared on both comparisons and evaluations, we do not find average treatment effects to support Hypothesis 3.4 that the combined effect leads to higher timing improvements in T4 than T1 (Table 3.3).

Still, when we examine the individual mechanisms which we expect to be driving comparisons and evaluations, we find four interesting results highlighting how husbands and wives differ in their responses to both pieces of information.

First, *Ahead and Poseval* wives are more likely to *reduce* their timing improvements in T4 compared to T1 (Table 3.8, columns 1, 3, 5, and 7), whereas *Behind and Poseval* wives are more likely to *increase* their timing improvements in T4 than in T1 (Table 3.9, columns 1, 3, 5, and 7). Taken together, this yields further support for the notion that the competitor being the husband is a significant factor in effort for wives. Positive evaluations appear to act as encouragement only when the wife is behind her husband, potentially reducing the disutility associated with competing against him. It also reflects support of the 'Pygmalion effect', where high expectations lead to improved performance, but again, only if she is behind her husband (Daido and Itoh, 2005; Rosenthal and Jacobson, 1968).

Second, although we observe similar positive effects of *Poseval* for husbands who are behind in T4 (Table 3.9, columns 2 and 4), these are not as strong as they are for wives. We interpret this as suggestive evidence that wives have more to gain from a positive evaluation than husbands do. This explanation is also supported by results from the distribution of positive and negative evaluations between husbands and wives (Table 3.2), where we find that husbands are over twice as likely as wives to receive a positive evaluation from their wives, and wives are four times more likely than husbands to receive a negative evaluation.[28]

Third, *Ahead and Negeval* husbands are more likely to decrease their timing improvements in T4, relative to T1 (Table 3.8, columns 2, 4, and 8). Interestingly, this effect is not present for wives, which suggests that perhaps husbands are more sensitive to negative evaluations from their spouses than wives are. Once again, this is plausible given that husbands are much less likely to receive a negative evaluation from their spouse than wives are, which might make it more effective when it is received. It is also worth noting that the reduced time improvements for husbands in T4 with a negative spousal evaluation noted in Table 3.7 appear to be driven by husbands who are ahead (significant T4 × *Ahead and Negeval* coeff in Table 3.8, cols 2, 4, 6, and 8) than those who are behind (no significant T4 × *Behind and Negeval* coeff in Table 3.9). However, this might also be due to these regressions being run separately. Indeed, examining Table G.2 which combines all subgroups, we find that both T4 × *Ahead and Negeval* and T4 × *Behind and Negeval* to be significant for husbands (columns 2 and 4).

[27]Note that the wife should still hold a belief that her performance is better than her spouse's—we would assume that this belief is accurate as spouses have full information of each other's abilities).

[28]Wives might also have more to gain from positive evaluations if, as the literature often finds, women have relatively lower self-confidence than men (e.g., see Möbius et al. (2022); Sarsons and Xu (2015)). However, using our measure of self-confidence described in Appendix E, and comparing its mean between wives and husbands, we do not find statistical differences between genders.

Finally, it is worth discussing why we might have significant combined effects for husbands and wives in treatments where only one type of information is shared. These are noted for *Ahead and Poseval* wives in T2 and T3 (cols 1 and 3, Table 3.8), and for *Behind and Poseval* as well as *Behind and Negeval* husbands in T3 (cols 2, 4, 6 and 8, Table 3.9). In all cases, we assume that, since spouses know each other's ability and nature of evaluations outside of the game, they form accurate beliefs around whether they are ahead or behind, and whether they have received a positive or negative evaluation, thus adjusting effort accordingly.

4.2. Alternative behavioural motivations

This section benchmarks our results against three alternative motivations—fairness, altruism, and conflict aversion—distinguishing between mechanisms operating through realised outcomes (relative performance gaps) and those operating through beliefs (relative confidence gaps). Table 3.13 summarises the predicted effort responses under each motivation and their consistency with our results.

Table 3.13: Other behavioural motivations and predicted spousal responses

Motivation	**If ahead/behind compared to spouse**	**If positively/negatively evaluated by spouse**	**Consistent with results**
Fairness (inequality aversion)	**Behind:** ↑ effort to reduce inequality (driven by envy). **Ahead:** ↓ effort to limit advantage.	N/A: fairness concerns realised outcomes, not beliefs.	High
Altruism	**Behind:** ↑ effort, but attenuated by concern for reducing spouse's utility. **Ahead:** ↓ or muted effort to avoid increasing spouse's disadvantage	N/A: evaluations do not link one spouse's effort to the other's utility.	Moderate
Conflict aversion	**Behind:** ↑ effort to reduce disparity and avoid tension. **Ahead:** ↓ effort to avoid creating conflict.	**Negeval:** ↓ effort (perceived discouragement from competing). **Poseval:** ↑ effort (perceived approval to compete).	Moderate

1. **Fairness (inequality aversion):** Following Fehr and Schmidt (1999), if a spouse dislikes inequality, their response depends on their relative position. When ahead, they may experience 'guilt' and reduce effort to limit the disparity. When behind, they may experience 'envy' (Grund and Sliwka, 2005; Kragl and Schmid, 2009), and increase effort to 'catch up' or eliminate a partner's advantage and equalise earnings. This generates predictions similar to behind-aversion in our model, though the motivation is distributional rather than competitive. Crucially, because fairness relates to realised outcomes, spousal evaluations in T2 (which do not reveal outcome gaps, only confidence gaps) would not trigger fairness concerns.

2. **Altruism:** An altruistic spouse attaches positive weight to their partner's utility, dampening the marginal incentives for competition created by spousal comparisons (see Appendix A, Section A.4). Thus, when behind, the competitive drive to catch up is partially offset by concern for reducing a partner's advantage; when ahead, the incentive to extend the lead is tempered by concern for increasing the partner's disadvantage. Critically, if altruistic concerns are sufficiently strong, the model predicts that spousal comparisons could actually lead to a *decrease* in effort in T3 relative to T1.

3. **Conflict aversion**: A conflict-averse individual would prioritise household harmony over individual earnings. They may therefore withdraw effort if they perceive that winning causes the other spouse disutility. Conflict aversion further predicts a strong response to spousal evaluations in T2, as the individual might feel 'validated' or 'permitted' to compete if positively evaluated by their partner, and 'discouraged' from competing if negatively evaluated.

Our results are most consistent with a gender split in which husbands respond primarily through competitive and fairness channels—increasing effort when behind in T3 (Table 3.5), consistent with envy-driven behind-aversion—while wives respond primarily through relational channels. Wives reduce effort when ahead (Table 3.8, cols 1, 3, and 7), which is less consistent with envy and more consistent with fairness, altruistic, and conflict-averse motivations, reflecting a reluctance to maintain or widen an advantage over their spouse. This is reinforced by our heterogeneity analysis (Table 3.12), which shows that highly conflict-averse wives withdraw effort even when behind—suggesting that for many wives, the relational cost of a competitive interaction outweighs the marginal benefit of higher productivity. The altruism channel adds a further nuance for husbands: those with above-median marital trust reduce time improvements in T4 when ahead, consistent with the prediction that concern for the spouse's utility can override competitive incentives. Together, these results suggest that the spousal context influences effort responses through both relational and individual motivations that differ markedly across genders.

4.3. External validity

It is necessary to address common concerns over external validity levelled at lab-in-the-field experiments such as ours, where it is unclear to what extent spouses' game behaviour maps onto real-world economic behaviour. In our context, the task itself requires sustained concentration and speed under pressure, capturing fundamental non-cognitive skills rewarded in the labour market. However, there are two caveats in bridging spousal competition in a timed task to competitive dynamics of a dual-career household.

First, one might argue that spouses only tolerate competition within the 'safe' confines of an abstract task. Put differently, although the real-effort task is incentivised, spouses may not explicitly equate 'out-performing' in the effort task with 'out-earning' in real life. Even if they did, it is possible that the collaborative nature of a marriage might suppress such competitive behaviours when the financial stakes are high and 'winning' could cause genuine household friction. However, our experiment focuses on the relative strength of spousal comparisons and evaluations as fundamental behavioural motivators. The emergence of systematic, gender-differentiated responses in this low-stakes environment suggests an underlying sensitivity to performance hierarchies. If these dynamics manifest when rewards are negligible, they likely serve as a conservative baseline for the more potent performance comparisons found in higher-stakes domains, such as income generation or career progression. Indeed, this sensitivity is echoed in studies examining behavioural responses to relative income shifts within couples, which find that beyond a threshold, outperforming leads to compensatory gender-role behaviour, diminished marital stability, or even a deliberate reduction in female labour supply (Bertrand, Kamenica and Pan, 2015; Bittman et al., 2003; Booth and Van Ours, 2009; Bursztyn, Fujiwara and Pallais, 2017; Cheng and Sheng, 2023; Fisman et al., 2006; Folke and Rickne, 2020; Zhao, Ye and Shi, 2022). Ultimately, establishing whether these evaluative pressures persist in naturally occurring settings would require longitudinal data linking protracted intra-household comparisons to long-term economic outcomes—a promising avenue for future research.

Second, generalisability of experimental results to real-life settings rests on the assumption that proximity and observability are core features of dual-career households. While such competitive pressures are often associated with highly-educated, urban dual-career households, we argue they are equally salient in rural settings characterised by high occupational overlap. In our sample, this overlap is substantial: of the 519

couples where both spouses work, 21% of couples report working in the same primary activity, and 41% operate within the same broad sector (i.e., agriculture, trade, or marine livelihoods). In such contexts, spousal productivity is frequently observable, public, and thus a conceivably common subject of household discourse. This relevance is further sharpened by ongoing structural changes in Zanzibar, particularly the expansion of the tourism and value-added marine sectors, which have increased cross-gender participation in similar economic activities. Consequently, the experimental environment is not a mere abstraction; it captures a stylised version of a rapidly evolving reality where spouses' economic performances are becoming increasingly visible and comparable.

5. Conclusion

In this chapter, we use results from a lab-in-the-field experiment conducted with 1,328 spouses (664 married or cohabiting couples) to examine what drives spousal effort. We focus on two behavioural motivators of effort that have gained significant attention in the literature on gender differences in effort: confidence and competition (Buser et al., 2020; Buser, van den Assem and van Dolder, 2023; Buser, Niederle and Oosterbeek, 2014; Gneezy, Niederle and Rustichini, 2003; Möbius et al., 2022; Niederle and Vesterlund, 2010, 2011). In the experiment, spouses perform an incentivised real-effort task twice. Following the first round and prior to the second round of the task, we use a 2×2 treatment design to experimentally vary which of two types of information respondents are provided with: (1) their spouse's belief about the respondent's ability i.e., the spouse's spouse-confidence belief, hypothesised to influence effort through a positive or negative spousal evaluation (defined as the relative confidence difference between self and spouse); (2) their spouse's actual Round 1 performance timing, used to induce a competitive environment through direct spousal comparisons.

We formulate a simple model to predict differences in spousal effort improvements between the control group (T1) where spouses received no information, and each of the other treatment groups, T2 ('Beliefs'), T3 ('Timings'), and T4 ('Both'). In particular, we hypothesise that where spouse-confidence beliefs are shared in T2, respondents would increase their time improvements whether they received a positive or negative spousal evaluation, as an 'encouragement effect' and 'defiance effect', respectively, would raise the optimal effort compared to T1, where these effects are absent. Where timings are shared in T3, we hypothesise that ahead-seeking and behind-averse motivations would increase effort in T3 relative to T1. We also hypothesise that the effects of receiving both pieces of information as opposed to none or one of them, would lead to effort changes that are in line with the predicted directions of the earlier hypotheses, and that all hypothesised effects would be stronger for husbands than wives, owing to empirical evidence around gender differences in competition and confidence.

Our results offer novel insights into the role of the spousal relationship in driving effort, and how this special case of gendered dyads generates unexpected gender differences in the motivators of effort. For one, our summary statistics underscore the importance of examining gender differences in spousal effort. Notably, married couples do not mirror the behaviours documented in individual or mixed-gender settings: for example, we find that across all treatments, wives have larger improvements than men. Between treatments, we observe gender differences in spousal responses to different types of feedback. When spousal comparisons are prompted, we find husbands to be more behind-averse than wives in the spousal setting. When spousal evaluations are made known, we find that wives are more responsive than husbands to positive spousal evaluations, reducing (increasing) effort if they are ahead (behind). On the contrary, husbands are more responsive than wives to negative spousal evaluations, reducing effort whether they are ahead or behind.

Taken together, the findings from this experiment reveal that the unique dynamics of a marital relationship can meaningfully moderate confidence and competition as drivers of effort, through spousal evaluations and comparisons. This, in our view, has two main policy implications.

The fist relates to a future where dual career households become more widespread, and assortative mating forms couples with similar interests and skills. A natural corollary to this could be that spousal comparisons become more common: where spouses perform similar tasks, work in comparable occupations, and pursue increasingly convergent career paths, each other's effort and performances would become a more salient reference point for comparison. In this scenario, our results show that both spouses are susceptible to comparisons, and appear to moderate effort accordingly. Where larger inequalities exist, this dynamic may offer an unexpected channel for narrowing gaps in wages or labour outcomes between spouses: if competitive individuals increase their effort in response to being comparatively behind, spousal comparisons could serve as an internal household mechanism that nudges performance upward. However, this same mechanism could entrench disparities if the more advantaged spouse responds by increasing effort even further. Understanding these dynamics is therefore key for policies that rely on household-level production, entrepreneurship, or shared labour decisions.

A second implication relates to sectors in which spouses' livelihoods begin to converge—a trend that may accelerate as labour markets become less gender-segregated due to environmental change or shifting economic demands. In such settings, confidence-building (or confidence-undermining) evaluations within the household can meaningfully shape effort and labour supply. For example, in Zanzibar, where environmental pressures and economic transitions are pushing households toward tourism, fisheries, and seaweed, spouses with sector-specific expertise (often the husbands) may be well-positioned to encourage skill acquisition among their wives. Positive spousal evaluations could therefore support women's entry into emerging sectors, helping to narrow gender gaps in earnings and economic opportunities. This suggests that policies promoting joint training, shared performance feedback, or household-level labour interventions may generate behavioural spillovers that amplify their impact on gender equality.

Whilst this chapter offers novel evidence on gender differences in spouses' effort responses, it also opens avenues for future work. First, the study relies on a measure of general confidence relative to same-gender peers. Subsequent work could explore how confidence assessments shift when the comparison group is mixed-gender, other-gender, or even the spouse themselves. Such variations may reveal additional heterogeneity in self- and spouse-confidence belief formation within couples. Moreover, although the step-wise design of the beliefs questions sought to elicit granular data on perceived ranking distributions, this approach was not as effective as intended: the majority of respondents consistently selected the top category. Whilst this is expected given the well-documented effect of overconfidence bias in self-reported beliefs (Logg, Haran and Moore, 2018; Moore and Healy, 2008; Svenson, 1981), this also reflects the limitation in our design that beliefs were not monetarily incentivised. Although this design choice was made to avoid influencing subsequent effort provision, it may have contributed to the high concentration of responses in the top category, reflecting overconfidence or imprecise reporting. Future work could therefore explore alternative elicitation methods that balance incentive compatibility with minimising behavioural interference, to mitigate overconfident self-reporting and generate greater variation in reported beliefs.

Second, the observed patterns around spousal effort in a competitive environment may also be context-specific. For example, as shown by Günther et al. (2010), women can be equally or more competitive than men in gender-neutral or 'female-type' tasks, indicating that stereotype threat and task framing can meaningfully shape a willingness to compete. Future research could therefore replicate the experiment with alternative real-effort tasks to examine whether the spousal dynamics documented here generalise across contexts.

Last, it would also be interesting to extend this lab-in-the-field setting to real-world behaviours which require effort provision. Future research could explore whether similar gendered patterns arise in other economically meaningful settings, such as joint financial or labour allocation decisions, which would help assess the broader implications of intra-household influences of confidence and competition for efficiency and welfare.

References

Adamecz-Völgyi, Anna, and Nikki Shure. 2022. "The gender gap in top jobs–the role of overconfidence." *Labour Economics*, 79: 102283.

Akerlof, George A. 1984. "Gift exchange and efficiency-wage theory: Four views." *The American Economic Review*, 74(2): 79–83.

Armantier, Olivier, and Nicolas Treich. 2013. "Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging." *European Economic Review*, 62: 17–40.

Ashraf, Nava. 2009. "Spousal control and intra-household decision making: An experimental study in the Philippines." *American Economic Review*, 99(4): 1245–1277.

Ashraf, Nava, Oriana Bandiera, and Scott S Lee. 2014. "Awards unbundled: Evidence from a natural field experiment." *Journal of Economic Behavior & Organization*, 100: 44–63.

Babad, Elisha Y, Jacinto Inbar, and Robert Rosenthal. 1982. "Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers." *Journal of educational psychology*, 74(4): 459.

Backus, Peter, Maria Cubel, Matej Guid, Santiago Sánchez-Pagés, and Enrique López Mañas. 2023. "Gender, competition, and performance: Evidence from chess players." *Quantitative Economics*, 14(1): 349–380.

Bartling, Björn, and Ferdinand A. von Siemens. 2010. "Equal Sharing Rules in Partnerships." *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft*, 166(2): 299–320.

Bastani, Spencer, Thomas Giebe, and Oliver Gürtler. 2023. "Overconfidence and gender equality in the labor market." CESifo Working Paper No. 10339.

Berger, Jonah, and Devin Pope. 2011. "Can losing lead to winning?" *Management Science*, 57(5): 817–827.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *The Quarterly Journal of Economics*, 130(2): 571–614.

Bittman, Michael, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. 2003. "When does gender trump money? Bargaining and time in household work." *American Journal of sociology*, 109(1): 186–214.

Blanco, Mariana, Dirk Engelmann, Alexander K Koch, and Hans-Theo Normann. 2010. "Belief elicitation in experiments: is there a hedging problem?" *Experimental economics*, 13(4): 412–438.

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 91(1): 166–193.

Booth, Alison, and Eiji Yamamura. 2018. "Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in Japan." *Review of Economics and Statistics*, 100(4): 581–593.

Booth, Alison, and Patrick Nolen. 2012. "Choosing to compete: How different are girls and boys?" *Journal of Economic Behavior & Organization*, 81(2): 542–555.

Booth, Alison L, and Jan C Van Ours. 2009. "Hours of work and gender identity: Does part-time work make the family happier?" *Economica*, 76(301): 176–196.

Booth, Alison, Ryohei Hayashi, and Eiji Yamamura. 2022. "Gender differences in tournament-performance over time in single-sex and mixed-sex environments." *Labour Economics*, 76: 102173.

Browning, Martin, and Pierre-André Chiappori. 1998. "Efficient intra-household allocations: A general characterization and empirical tests." *Econometrica*, 1241–1278.

Browning, Martin, Francois Bourguignon, Pierre-André Chiappori, and Valerie Lechene. 1994. "Income and outcomes: A structural model of intrahousehold allocation." *Journal of political Economy*, 102(6): 1067–1096.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting wife': Marriage market incentives and labor market investments." *American Economic Review*, 107(11): 3288–3319.

Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. 2020. "Overconfidence and gender gaps in redistributive preferences: Cross-Country experimental evidence." *Journal of Economic Behavior & Organization*, 178: 267–286.

Buser, Thomas, Martijn J. van den Assem, and Dennie van Dolder. 2023. "Gender and willingness to compete for high stakes." *Journal of Economic Behavior & Organization*, 206: 350–370.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, competitiveness, and career choices." *The Quarterly Journal of Economics*, 129(3): 1409–1447.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *The American Economic Review*, 107(5): 125–130.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics*, 117(3): 871–915.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics*, 90(3): 414–427.

Charness, Gary, and Peter Kuhn. 2007. "Does Pay Inequality Affect Worker Effort? Experimental Evidence." *Journal of Labor Economics*, 25(4): 693–723.

Charness, Gary, Ramón Cobo-Reyes, Juan A. Lacomba, Francisco Lagos, and Jose Maria Perez. 2016. "Social comparisons in wage delegation: experimental evidence." *Experimental Economics*, 19(2): 433–459.

Cheng, Yawen, and Yan Sheng. 2023. "Relative income within households, gender identity, and entrepreneurship." *Economic Analysis and Policy*, 80: 90–108.

Chiappori, Pierre-André. 1988. "Rational household labor supply." *Econometrica: Journal of the Econometric Society*, 63–90.

Chiappori, Pierre-André. 1992. "Collective labor supply and welfare." *Journal of Political Economy*, 100(3): 437–467.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2014. "Social Comparison and Effort Provision: Evidence from a Field Experiment." *Journal of the European Economic Association*, 12(4): 877–898.

Cook, William L., and Emily M. Douglas. 1998. "The looking-glass self in family context: A social relations analysis." *Journal of Family Psychology*, 12(3): 299–309. Place: US.

Cooley, Charles Horton. 1922. *Human nature and the social order.* New York: Scribner.

Croson, Rachel, and Uri Gneezy. 2009. "Gender differences in preferences." *Journal of Economic literature*, 47(2): 448–474.

Daido, Kohei, and Hideshi Itoh. 2005. "The Pygmalion effect: An agency model with reference dependent preferences." CESifo Working Paper.

Dasgupta, Utteeyo, Subha Mani, Smriti Sharma, and Saurabh Singhal. 2019. "Can gender differences in distributional preferences explain gender gaps in competition?" *Journal of Economic Psychology*, 70: 1–11.

Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval. 2013. "Gender matching and competitiveness: Experimental evidence." *Economic Inquiry*, 51(1): 816–835.

Davidson, Oranit B, and Dov Eden. 2000. "Remedial self-fulfilling prophecy: Two field experiments to prevent Golem effects among disadvantaged women." *Journal of Applied Psychology*, 85(3): 386.

DellaVigna, Stefano, and Devin Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts." *The Review of Economic Studies*, 85(2): 1029–1069.

Demougin, Dominique, Claude Fluet, and Carsten Helm. 2006. "Output and Wages with Inequality Averse Agents." *The Canadian Journal of Economics / Revue canadienne d'Economique*, 39(2): 399–413.

Donia, Magda, Thomas A O'Neill, and Stephane Brutus. 2015. "Peer feedback increases team member performance, confidence and work outcomes: A longitudinal study." Vol. 2015, 12560. Academy of Management Briarcliff Manor, NY 10510.

Duflo, Esther. 2003. "Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa." *The World Bank Economic Review*, 17(1): 1–25.

D'Exelle, Ben, and Arno Riedl. 2018. "Gender inequality and resource sharing: Evidence from rural Nicaragua." *Journal of Development Perspectives*, 2(1-2): 62–87.

D'Exelle, Ben, Rik Habraken, and Arjan Verschoor. 2024. "Effort and Social Comparison: Experimental Evidence from Uganda." *Economic Development and Cultural Change*, 72(4).

Eden, Dov. 1990. *Pygmalion in management: Productivity as a self-fulfilling prophecy.* Lexington Books/DC Heath and Com.

Engel, Christoph. 2011. "Dictator games: A meta study." *Experimental Economics*, 14(4): 583–610.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." *The Quarterly Journal of Economics*, 114(3): 817–868.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." *Journal of Economic Perspectives*, 14(3): 159–182. ISBN: 0895-3309.

Felson, Richard B. 1985. "Reflected appraisal and the development of self." *Social Psychology Quarterly*, 71–78.

Feyli, Sara, and Mohammad Ali Ayatollahi. 2016. "The effects of classroom peer-feedback on advanced students' self-confidence." *Journal of Modern Research in English Language Studies*, 3(1): 87–67.

Fisman, Raymond, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. 2006. "Gender differences in mate selection: Evidence from a speed dating experiment." *The Quarterly Journal of Economics*, 121(2): 673–697.

Flory, Jeffrey A, Andreas Leibbrandt, and John A List. 2015. "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions." *The Review of Economic Studies*, 82(1): 122–155.

Folke, Olle, and Johanna Rickne. 2020. "All the single ladies: Job promotions and the durability of marriage." *American Economic Journal: Applied Economics*, 12(1): 260–287.

Frey, Bruno S., and Stephan Meier. 2004. "Social Comparisons and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment." *American Economic Review*, 94(5): 1717–1722.

Friedl, Andreas, Andreas Pondorfer, and Ulrich Schmidt. 2020. "Gender differences in social risk taking." *Journal of Economic Psychology*, 77: 102182.

Gill, David, and Victoria Prowse. 2014. "Gender differences and dynamics in competition: The role of luck." *Quantitative Economics*, 5(2): 351–376. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE309.

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse. 2019. "First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision." *Management Science*, 65(2): 494–507.

Gneezy, Uri, and Aldo Rustichini. 2000. "Pay Enough or Don't Pay at All." *The Quarterly Journal of Economics*, 115(3): 791–810.

Gneezy, Uri, and Aldo Rustichini. 2004. "Gender and competition at a young age." *American Economic Review*, 94(2): 377–381.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics*, 118(3): 1049–1074.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American economic review*, 104(4): 1091–1119.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry your like: Assortative mating and income inequality." *American Economic Review*, 104(5): 348–353.

Grund, Christian, and Dirk Sliwka. 2005. "Envy and Compassion in Tournaments." *Journal of Economics & Management Strategy*, 14(1): 187–207. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1430-9134.2005.00039.x.

Gächter, Simon, and Christian Thöni. 2010. "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis." *Journal of Economic Behavior & Organization*, 76(3): 531–543.

Gächter, Simon, and Elke Renner. 2010. "The effects of (incentivized) belief elicitation in public goods experiments." *Experimental Economics*, 13(3): 364–377.

Günther, Christina, Neslihan Arslan Ekinci, Christiane Schwieren, and Martin Strobel. 2010. "Women can't jump?—An experiment on competitive attitudes and stereotype threat." *Journal of Economic Behavior & Organization*, 75(3): 395–401.

Huet-Vaughn, Emiliano. 2025. "Do social comparisons motivate workers? A field experiment on relative earnings, labor supply and the inhibitory effect of pay inequality." *Experimental Economics*, 28(1): 155–180.

Iversen, Vegard, Cecile Jackson, Bereket Kebede, Alistair Munro, and Arjan Verschoor. 2011. "Do Spouses Realise Cooperative Gains? Experimental Evidence from Rural Uganda." *World Development*, 39(4): 569–578.

Kamas, Linda, and Anne Preston. 2015. "Can social preferences explain gender differences in economic behavior?" *Journal of Economic Behavior & Organization*, 116: 525–539.

Kilduff, Gavin J, Hillary Anger Elfenbein, and Barry M Staw. 2010. "The psychology of rivalry: A relationally dependent analysis of competition." *Academy of management journal*, 53(5): 943–969.

Kragl, Jenny, and Julia Schmid. 2009. "The impact of envy on relational employment contracts." *Journal of Economic Behavior & Organization*, 72(2): 766–779.

Köszegi, Botond. 2006. "Ego utility, overconfidence, and task choice." *Journal of the European Economic Association*, 4(4): 673–707.

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of political Economy*, 89(5): 841–864.

Logg, Jennifer M, Uriel Haran, and Don A Moore. 2018. "Is overconfidence a motivated bias? Experimental evidence." *Journal of Experimental Psychology: General*, 147(10): 1445.

Lundberg, Shelly, and Robert A Pollak. 1993. "Separate spheres bargaining and the marriage market." *Journal of political Economy*, 101(6): 988–1010.

Masekesa, Faith, and Alistair Munro. 2020. "Intra-household inequality, fairness and productivity. Evidence from a real effort experiment." *World Development*, 127: 104763.

Moore, Don A, and Paul J Healy. 2008. "The trouble with overconfidence." *Psychological review*, 115(2): 502.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2022. "Managing self-confidence: Theory and experimental evidence." *Management Science*, 68(11): 7793–7817.

Niederle, Muriel. 2017. "A gender agenda: a progress report on competitiveness." *American Economic Review*, 107(5): 115–119.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122(3): 1067–1101.

Niederle, Muriel, and Lise Vesterlund. 2010. "Explaining the gender gap in math test scores: The role of competition." *Journal of Economic Perspectives*, 24(2): 129–144.

Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and competition." *Annual Review of Economics*, 3(1): 601–630.

Nurmohamed, Samir. 2020. "The underdog effect: When low expectations increase performance." *Academy of Management Journal*, 63(4): 1106–1133.

Nurra, Cécile, and Pascal Pansu. 2009. "The impact of significant others' actual appraisals on children's self-perceptions: What about Cooley's assumption for children?" *European Journal of Psychology of Education*, 24(2): 247–262.

Oz, Sasson, and Dov Eden. 1994. "Restraining the Golem: Boosting performance by changing the interpretation of low scores." *Journal of Applied Psychology*, 79(5): 744.

Roels, Guillaume, and Xuanming Su. 2014. "Optimal design of social comparison effects: Setting reference groups and reference points." *Management Science*, 60(3): 606–627.

Rosenthal, Robert, and Lenore Jacobson. 1968. "Pygmalion in the classroom." *The Urban Review*, 3(1): 16–20.

Ruegg, Rachael. 2018. "The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy." *The Language Learning Journal*, 46(2): 87–102.

Rutström, E., and Nathaniel Wilcox. 2009. "Stated Beliefs versus Inferred Beliefs: A Methodological Inquiry and Experimental Test." *ESI Publications.*

Sanders, Glenn S., and Jerry Suls. 1982. "Social Comparison, Competition and Marriage." *Journal of Marriage and Family*, 44(3): 721–730.

Sari, Hesty Puspita, Azza Jauhar Ahmad Tajuddin, and Miftachus Sholikah. 2025. "The Effectiveness of Peer Feedback in Collaborative Writing Viewed from Self-Confidence." *Proceedings International Conference on Sustainable Innovation (ICoSI)*, 2(1).

Sarsons, Heather, and Guo Xu. 2015. "Confidence men? Gender and confidence: Evidence among top economists." *Harvard University, Department of Economics, Littauer Center*, 1–26.

Sen, Amartya. 1990. "Gender and Cooperative Conflicts." In *Persistent Inequalities.* 123–148. New York:Oxford University Press.

Svenson, Ola. 1981. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica*, 47: 143–148.

Udry, Christopher. 1996. "Gender, Agricultural Production, and the Theory of the Household." *Journal of Political Economy*, 104(5): 1010–1046.

Verschoor, Arjan, Bereket Kebede, Alistair Munro, and Marcela Tarazona. 2019. "Spousal control and efficiency of intra-household decision-making: Experiments among married couples in India, Ethiopia and Nigeria." *The European Journal of Development Research*, 31(4): 1171–1196.

Xue, Jinli, Mengting Zhu, Yanan Guo, and Demin Kong. 2022. "The Double-Edged Sword of Underdog Expectations in Organizations in Shanghai, China: The Mediating Role of Feedback-Avoiding Behaviors and Proving Others Wrong and the Moderating Role of Regulatory Focus." *Psychology Research and Behavior Management*, 15: 2205–2218.

Zhao, Yucong, Bing Ye, and Jinchuan Shi. 2022. "Gender identity, preference, and relative income within households." *China Economic Review*, 71: 101741.

Appendix

A. First-order condition derivations for theoretical model

Given that spouses maximise utility U_i where:

$$U_i = p(e_i) - c(e_i) + \Theta_i(e_i, e_j) \tag{A.1}$$

we can find the optimal level of effort provision that maximises utility by solving the following first-order condition (FOC):

$$\frac{\partial U_i}{\partial e_i} = p'(e_i^*) - c'(e_i^*) + \frac{\partial \Theta_i}{\partial e_i} = 0 \implies p'(e_i^*) + \frac{\partial \Theta_i}{\partial e_i} = c'(e_i^*) \tag{A.2}$$

The FOC specific to each treatment is then derived by modifying Θ as defined by equation 3.4 (repeated here for ease of reference):

$$\Theta(e_i, e_j) = \underbrace{[d_i \cdot \gamma_P(e_i) + (1 - d_i) \cdot \gamma_N(e_i)]}_{\text{spousal evaluation}} + \underbrace{[\beta \cdot g(e_i, e_j) - \alpha \cdot f(e_i, e_j)]}_{\text{spousal comparison}}$$

A.1. FOCs under T2 (spousal evaluations)

Where only spousal evaluations are shared (T2), the FOC is derived from the first square bracket in equation 3.4 (the second square bracket is zero as they do not receive spousal comparison information), and becomes;

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + d_i.\gamma'_P(e_i) + (1 - d_i).\gamma'_N(e_i) = 0 \tag{A.3}$$

However, as a person can only receive a negative or positive evaluation (not both), for a person who receives a positive evaluation i.e., $(d_i = 1)$, the FOC from equation A.3 is reduced to:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \gamma'_P(e_i) = 0 \implies p'(e_i^*) + \gamma'_P(e_i) = c'(e_i^*) \tag{A.4}$$

Similarly, for a person who receives a negative evaluation i.e., $(d_i = 0)$, the FOC from equation A.3 is reduced to:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \gamma'_N(e_i) = 0 \implies p'(e_i^*) + \gamma'_N(e_i) = c'(e_i^*) \tag{A.5}$$

A.2. FOCs under T3 (spousal comparisons)

Where only spousal comparisons are shared (T3), the FOC is derived from the second square bracket in equation 3.4 (the first square bracket is now zero as they do not receive spousal evaluation information), and becomes;

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) - \alpha \frac{f(e_i, e_j)}{\partial e_i} + \beta \frac{\partial g(e_i, e_j)}{\partial e_i} = 0 \tag{A.6}$$

where

$$\frac{\partial f(e_i, e_j)}{\partial e_i} = \begin{cases} -p'(e_i) & if \quad p(e_i) < p(e_j) \\ 0 & if \quad p(e_i) \geq p(e_j) \end{cases}$$

$$\frac{\partial g(e_i, e_j)}{\partial e_i} = \begin{cases} 0 & if \quad p(e_i) \leq p(e_j) \\ p'(e_i) & if \quad p(e_i) > p(e_j) \end{cases}$$

However, as a person can only be ahead or behind (not both), for a person who is behind, the FOC from equation A.6 is reduced to:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \alpha p'(e_i) = 0 \implies (1+\alpha)p'(e_i^*) = c'(e_i^*) \tag{A.7}$$

and for a person who is ahead, the FOC from equation A.6 is reduced to:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \beta p'(e_i) = 0 \implies (1+\beta)p'(e_i^*) = c'(e_i^*) \tag{A.8}$$

A.3. FOCs under T4 (both evaluations and comparisons)

Depending on the feedback received, the FOC from equation A.3 collapses into four specific states:

Behind and Negeval: For an individual who is behind ($p(e_i) < p(e_j)$) and receives a negative evaluation ($d_i = 0$), the FOC becomes:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \gamma'_N(e_i) + \alpha p'(e_i) = 0 \implies (1+\alpha)p'(e_i) + \gamma'_N(e_i) = c'(e_i) \tag{A.9}$$

Behind and Poseval: For an individual who is behind ($p(e_i) < p(e_j)$) but receives a positive evaluation ($d_i = 1$), the FOC becomes:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \gamma'_P(e_i) + \alpha p'(e_i) = 0 \implies (1+\alpha)p'(e_i) + \gamma'_P(e_i) = c'(e_i) \tag{A.10}$$

Ahead and Negeval: For an individual who is ahead ($p(e_i) > p(e_j)$) but receives a negative evaluation ($d_i = 0$), the FOC becomes:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \gamma'_N(e_i) + \beta p'(e_i) = 0 \implies (1+\beta)p'(e_i) + \gamma'_N(e_i) = c'(e_i) \tag{A.11}$$

Ahead and Poseval: For an individual who is ahead ($p(e_i) > p(e_j)$) and receives a positive evaluation ($d_i = 1$) the FOC becomes:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \gamma'_P(e_i) + \beta p'(e_i) = 0 \implies (1+\beta)p'(e_i) + \gamma'_P(e_i) = c'(e_i) \tag{A.12}$$

This can be summarised by Table A.1.

Table A.1: Summary of FOCs under T4

Scenario	Comparison component	Evaluation component	Equilibrium condition
Behind & Negeval	$+\alpha p'(e_i)$	$+\gamma'_N(e_i)$	$(1+\alpha)p'(e_i^*) + \gamma'_N(e_i) = c'(e_i^*)$
Behind & Poseval	$+\alpha p'(e_i)$	$+\gamma'_P(e_i)$	$(1+\alpha)p'(e_i^*) + \gamma'_P(e_i) = c'(e_i^*)$
Ahead & Negeval	$+\beta p'(e_i)$	$+\gamma'_N(e_i)$	$(1+\beta)p'(e_i^*) + \gamma'_N(e_i) = c'(e_i^*)$
Ahead & Poseval	$+\beta p'(e_i)$	$+\gamma'_P(e_i)$	$(1+\beta)p'(e_i^*) + \gamma'_P(e_i) = c'(e_i^*)$

A.4. FOCs under T3, with altruism

If we assume that spouses are altruistic, the individual utility is re-defined as:

$$U_i = p(e_i) - c(e_i) + \Theta_i(e_i, e_j) + \rho U_j$$

where $\rho \in [0, 1)$ is the individual's degree of altruism toward the spouse, with a higher ρ placing more weight on the spouse's welfare. Utility can thus be rewritten as:

$$U_i = p(e_i) - c(e_i) + \Theta_i(e_i, e_j) + \rho\big[p(e_j) - c(e_j) + \Theta_j(e_j, e_i)\big]$$

Taking the FOC, we get:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \frac{\partial \Theta_i}{\partial e_i} + \rho\left[\frac{\partial \Theta_j}{\partial e_i}\right]$$

But since in T3 only comparisons matter:

$$\Theta_i = \beta g(e_i, e_j) - \alpha f(e_i, e_j) \qquad \text{and} \qquad \Theta_j = \beta g(e_j, e_i) - \alpha f(e_j, e_i)$$

Thus:

$$\frac{\partial U_i}{\partial e_i} = p'(e_i) - c'(e_i) + \underbrace{\left[\beta\frac{\partial g(e_i, e_j)}{\partial e_i} - \alpha\frac{\partial f(e_i, e_j)}{\partial e_i}\right]}_{\text{own comparison}} + \rho\underbrace{\left[\beta\frac{\partial g(e_j, e_i)}{\partial e_i} - \alpha\frac{\partial f(e_j, e_i)}{\partial e_i}\right]}_{\text{spouse comparison}} = 0$$

Case 1: Individual is behind i.e., $p(e_i) < p(e_j)$. In this case, the partial derivative of the individual's own comparison component is still $+\alpha p'(e_i)$. If i is behind, then spouse j is ahead, which makes the partial derivative of the spouse's comparison component $\rho \cdot \beta(-p'(e_i))$. Thus, the FOC (behind, with altruism) is:

$$p'(e_i) - c'(e_i) + \alpha p'(e_i) - \rho\beta p'(e_i) = 0 \Longrightarrow (1 + \alpha - \rho\beta)p'(e_i^*) = c'(e_i^*)$$

Case 2: Individual is ahead i.e., $p(e_i) > p(e_j)$. In this case, the partial derivative of the individual's own comparison component is still $+\beta p'(e_i)$. If i is ahead, then spouse j is behind, which makes the partial derivative of the spouse's comparison component $\rho \cdot (-\alpha p'(e_i))$. Thus, the FOC (ahead, with altruism) is:

$$p'(e_i) - c'(e_i) + \beta p'(e_i) - \rho\alpha p'(e_i) = 0 \Longrightarrow (1 + \beta - \rho\alpha)p'(e_i^*) = c'(e_i^*)$$

Essentially, altruism reduces the effective strength of comparison incentives. When behind, the incentive to catch up (α) is offset by concern for reducing the spouse's advantage ($\rho\beta$). When ahead, the incentive to extend the lead (β) is offset by concern for increasing the spouse's disadvantage ($\rho\alpha$). Note, we can also have that $(1 + \beta - \rho\alpha < 1)$ or $(1 + \alpha - \rho\beta < 1)$ i.e., with sufficiently strong altruism, comparison feedback may reduce rather than increase effort, relative to the baseline.

B. Household models

Here, we consider how the first-order conditions from our model—and hence the model predictions—might change under different household models.

B.1. Income-maximising household

In the income-maximising (unitary) model, the household is assumed to choose effort levels to maximise total monetary income generated through individual earnings. Since income in our setting is derived solely from effort-dependent earnings, the household objective can be written as:

$$\max_{e_i, e_j}; W(e_i, e_j) = p(e_i) + p(e_j). \tag{B.1}$$

This formulation abstracts from any direct role of effort costs or psychological components of utility (such as Θ) in the household objective. Instead, such considerations are assumed not to enter the collective decision problem, consistent with a unitary benchmark in which only income generation matters.

Assuming that $p(e)$ is increasing and concave, the first-order conditions are:

$$p'(e_i^*) = 0 \qquad \text{and} \qquad p'(e_j^*) = 0. \tag{B.2}$$

Thus, effort is chosen up to the point at which the marginal product of effort is zero. Therefore, under the income-maximising household, effort is determined purely by the production technology and is independent of psychological influences captured by Θ in our individual model. As a result, we would expect no variation in spousal behaviour across our treatments under .

B.2. Collective household model

In the collective framework, spouses jointly choose effort levels and the allocation of resources to maximise a Pareto-weighted household objective H:

$$H = \lambda U_i + (1 - \lambda) U_j, \tag{B.3}$$

where $\lambda \in [0, 1]$ represents the bargaining power of spouse i.

Individual utility depends on private consumption (x), psychological influences on effort (Θ), and the private cost of effort (c):

$$U_k = u_k(x_k) + \Theta_k(e_i, e_j) - c_k(e_k) \quad \text{for } k \in i, j. \tag{B.4}$$

The household faces a budget constraint where total consumption equals total income generated by individual effort:

$$x_i + x_j = Y = p(e_i) + p(e_j). \tag{B.5}$$

Following Browning and Chiappori (1998), the household problem can be decomposed into two stages. In the first stage, for any given level of income Y, the household allocates consumption to maximise the weighted sum of utilities from private consumption:

$$\max_{x_i, x_j} [\lambda u_i(x_i) + (1 - \lambda) u_j(x_j)] \quad \text{s.t.} \quad x_i + x_j = Y \tag{B.6}$$

We can set up the Lagrangian for this sub-problem:

$$\mathcal{L} = \lambda u_i(x_i) + (1-\lambda)u_j(x_j) + \mu[Y - x_i - x_j]$$

where the FOCs are:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \lambda u_i'(x_i) - \mu = 0 \implies \lambda u_i'(x_i) = \mu$$

$$\frac{\partial \mathcal{L}}{\partial x_j} = (1-\lambda)u_j'(x_j) - \mu = 0 \implies (1-\lambda)u_j'(x_j) = \mu$$

which implies:

$$\lambda u_i'(x_i) = (1-\lambda)u_j'(x_j) = \mu \tag{B.7}$$

The solution to this first-stage problem defines the optimal allocation of income, which we denote as the sharing rules $x_i = \phi_i(Y, \lambda)$ and $x_j = \phi_j(Y, \lambda)$. These rules determine how any level of total household income is distributed between spouses at the margin.

In the second stage, the household chooses effort levels. By substituting the sharing rules from the first stage, we can express the household objective H solely as a function of effort and bargaining power:

$$\max_{e_i, e_j} H = \underbrace{\lambda u_i(\phi_i(Y,\lambda)) + (1-\lambda)u_j(\phi_j(Y,\lambda))}_{\text{Allocated consumption}} + \underbrace{\lambda\Theta_i(e_i,e_j) + (1-\lambda)\Theta_j(e_i,e_j)}_{\text{Psychological influences on utility}} - \underbrace{\lambda c_i(e_i) - (1-\lambda)c_j(e_j)}_{\text{Weighted effort costs}}. \tag{B.8}$$

To derive the first-order condition for spouse i, we differentiate H with respect to e_i.

The consumption part of the household objective is:

$$V(Y) = \lambda u_i(x_i(Y)) + (1-\lambda)u_j(x_j(Y))$$

which we can differentiate with respect to e_i using the Chain Rule:

$$\frac{\partial V}{\partial e_i} = \left[\lambda u_i'(x_i)\frac{\partial x_i}{\partial Y} + (1-\lambda)u_j'(x_j)\frac{\partial x_j}{\partial Y}\right] \cdot \frac{\partial Y}{\partial e_i}$$

However our first derivation shown in equation B.7, we know that $\lambda u_i'(x_i) = \mu$ and $(1-\lambda)u_j'(x_j) = \mu$. Substituting μ into the bracket:

$$\frac{\partial V}{\partial e_i} = \left[\mu\frac{\partial x_i}{\partial Y} + \mu\frac{\partial x_j}{\partial Y}\right] \cdot \frac{\partial Y}{\partial e_i}$$

and we can factor out the μ:

$$\frac{\partial V}{\partial e_i} = \mu\left[\frac{\partial x_i}{\partial Y} + \frac{\partial x_j}{\partial Y}\right] \cdot \frac{\partial Y}{\partial e_i}$$

Because of the budget constraint ($x_i + x_j = Y$), we can rewrite this:

$$\frac{\partial x_i}{\partial Y} + \frac{\partial x_j}{\partial Y} = \frac{\partial (x_i + x_j)}{\partial Y} = \frac{\partial Y}{\partial Y} = 1$$

So, the whole complex bracket simply collapses to $\mu \cdot 1$. Thus:

$$\frac{\partial V}{\partial e_i} = \mu \cdot \frac{\partial Y}{\partial e_i}$$

Defining $U'(Y) \equiv \mu$ and since $\frac{\partial Y}{\partial e_i}$ is just the marginal product $p'(e_i)$, we get:

$$\mathbf{U'(Y)p'(e_i)}$$

Thus,

$$\frac{\partial H}{\partial e_i} = \underbrace{U'(Y)p'(e_i)}_{\text{Marginal income}} + \underbrace{\lambda\frac{\partial \Theta_i}{\partial e_i} + (1-\lambda)\frac{\partial \Theta_j}{\partial e_i}}_{\text{Psychological spillover}} = \underbrace{\lambda c'(e_i)}_{\text{Weighted cost}} \tag{B.9}$$

The FOC illustrates that optimal effort choices are determined by three factors: the marginal product of effort, the psychological externalities of spouses' performances imposed on one other, and the marginal cost of effort. Crucially, these factors are moderated by bargaining power. For instance, if spouse i has very little bargaining power ($\lambda \to 0$), the household objective places almost no weight on their private cost of effort, potentially pushing them toward higher effort levels to increase total income for the other spouse.

In the context of our treatments, the collective model predicts the same spousal behaviour that our theoretical model predicts. That is, in cases where the treatment 'activates' the psychological channel through Θ, optimal effort is higher than in the control group, where $\Theta = 0$.

C. Balancing tests

Table C.1: Balancing tests for CH3 treatment groups, for women

Variable	(1) T1: None Mean/(SE)	(2) T2: Beliefs Mean/(SE)	(3) T3: Times Mean/(SE)	(4) T4: Both Mean/(SE)	(1)-(2) Mean difference	(1)-(3) Mean difference	(1)-(4) Mean difference	(2)-(3) Mean difference	(2)-(4) Mean difference	(3)-(4) Mean difference
							Pairwise t-test			
Age	42.072	40.657	40.072	41.843	1.416	2.000	0.229	0.584	-1.187	-1.771
	(0.895)	(0.901)	(0.855)	(0.789)						
Education (y/n)	0.837	0.855	0.843	0.904	-0.018	-0.006	-0.066*	0.012	-0.048	-0.060*
	(0.029)	(0.027)	(0.028)	(0.023)						
Swim skills (y/n)	0.072	0.030	0.042	0.042	0.042*	0.030	0.030	-0.012	-0.012	0.000
	(0.020)	(0.013)	(0.016)	(0.016)						
Risk averse (y/n)	0.747	0.759	0.765	0.723	-0.012	-0.018	0.024	-0.006	0.036	0.042
	(0.034)	(0.033)	(0.033)	(0.035)						
Income in a good month (TZS '000s')	78.849	78.886	87.066	80.422	-0.036	-8.217	-1.572	-8.181	-1.536	6.645
	(9.472)	(8.426)	(10.003)	(8.965)						
Income in a bad month (TZS '000s')	33.654	38.479	36.398	39.342	-4.825	-2.744	-5.689	2.081	-0.863	-2.945
	(5.706)	(6.083)	(5.394)	(6.930)						
Years lived in the village	32.313	32.277	31.319	32.813	0.036	0.994	-0.500	0.958	-0.536	-1.494
	(1.400)	(1.312)	(1.277)	(1.293)						
Years of marriage	19.524	18.837	18.283	19.422	0.687	1.241	0.102	0.554	-0.584	-1.139
	(1.085)	(1.011)	(0.925)	(0.849)						
No. of sons	1.916	1.922	2.145	1.970	-0.006	-0.229	-0.054	-0.223	-0.048	0.175
	(0.126)	(0.145)	(0.135)	(0.117)						
No. of daughters	1.819	1.735	1.849	1.861	0.084	-0.030	-0.042	-0.114	-0.127	-0.012
	(0.119)	(0.100)	(0.123)	(0.128)						
% of child costs contributed	44.880	42.892	41.566	42.410	1.988	3.313	2.470	1.325	0.482	-0.843
	(1.773)	(1.926)	(1.672)	(1.845)						
% of personal costs contributed	56.867	55.904	55.000	58.373	0.964	1.867	-1.506	0.904	-2.470	-3.373
	(2.034)	(2.172)	(2.112)	(2.101)						
% of household costs contributed	39.398	38.072	37.289	39.699	1.325	2.108	-0.301	0.783	-1.627	-2.410
	(1.693)	(1.739)	(1.706)	(1.777)						
Number of observations	166	166	166	166	332	332	332	332	332	332

Table C.2: Balancing tests for CH3 treatment groups, for men

Variable	(1) T1: None Mean/(SE)	(2) T2: Beliefs Mean/(SE)	(3) T3: Times Mean/(SE)	(4) T4: Both Mean/(SE)	(1)-(2) Mean difference	(1)-(3) Mean difference	(1)-(4) Mean difference	(2)-(3) Mean difference	(2)-(4) Mean difference	(3)-(4) Mean difference
							Pairwise t-test			
Age	50.422	47.223	49.012	48.669	3.199**	1.410	1.753	-1.789	-1.446	0.343
	(1.002)	(1.080)	(1.065)	(0.970)						
Education (y/n)	0.867	0.873	0.910	0.892	-0.006	-0.042	-0.024	-0.036	-0.018	0.018
	(0.026)	(0.026)	(0.022)	(0.024)						
Swim skills (y/n)	0.795	0.765	0.783	0.753	0.030	0.012	0.042	-0.018	0.012	0.030
	(0.031)	(0.033)	(0.032)	(0.034)						
Risk averse (y/n)	0.319	0.259	0.253	0.271	0.060	0.066	0.048	0.006	-0.012	-0.018
	(0.036)	(0.034)	(0.034)	(0.035)						
Income in a good month (TZS '000s')	443.542	441.609	427.228	288.976	1.933	16.314	154.566	14.380	152.633**	138.252**
	(92.429)	(56.699)	(54.132)	(19.323)						
Income in a bad month (TZS '000s')	177.488	156.398	171.241	127.440	21.090	6.247	50.048	-14.843	28.958	43.801*
	(42.879)	(26.044)	(22.091)	(10.626)						
Years lived in the village	45.946	43.102	45.434	44.627	2.843	0.512	1.319	-2.331	-1.524	0.807
	(1.260)	(1.344)	(1.307)	(1.256)						
Years of marriage	21.133	19.705	19.614	20.560	1.428	1.518	0.572	0.090	-0.855	-0.946
	(1.064)	(1.002)	(0.979)	(0.893)						
No. of sons	1.916	1.922	2.145	1.970	-0.006	-0.229	-0.054	-0.223	-0.048	0.175
	(0.126)	(0.145)	(0.135)	(0.117)						
No. of daughters	1.819	1.735	1.849	1.861	0.084	-0.030	-0.042	-0.114	-0.127	-0.012
	(0.119)	(0.100)	(0.123)	(0.128)						
% of child costs contributed	80.000	79.578	82.530	80.482	0.422	-2.530	-0.482	-2.952	-0.904	2.048
	(1.475)	(1.384)	(1.437)	(1.270)						
% of personal costs contributed	84.819	81.807	84.277	83.313	3.012	0.542	1.506	-2.470	-1.506	0.964
	(1.887)	(1.988)	(1.972)	(1.859)						
% of household costs contributed	85.241	83.313	85.482	85.301	1.928	-0.241	-0.060	-2.169	-1.988	0.181
	(1.261)	(1.291)	(1.334)	(1.129)						
Number of observations	166	166	166	166	332	332	332	332	332	332

D. Experimental script

The following text is the experimental script used for this study. Instructions for the enumerator are between brackets. Text in bold was emphasised by enumerators.

[Introduce yourself using the following script.]

Hello! My name is [insert name]. Thank you very much for coming. On behalf of our research team, I would like to welcome you to today's session. This session is being conducted as part of our research into coastal communities in Zanzibar. I am part of a research team that is working with a University. We have invited you here, today, because we want to learn about how people in this area take decisions. You are going to be asked to take decisions that allow you to earn some money. The money that results from your decisions will be yours to keep. What you need to do will be explained fully. But first we want to make a few of things clear.

First of all, this is not our money. We belong to a university, and this money has been given to us for research. Second, participation is voluntary. You may still choose not to participate in the exercise. If you decide to not to participate, you will still be compensated for your time. If you participate, make sure that you listen carefully to us. It is important that you follow our instructions. Your interview will last around 45 minutes. For participating in this interview, you will receive TSH 4,000 for sure, in addition to any money you might win during the interview. I will now explain the procedures in more detail.

Today's interview will be in THREE PARTS

In PART 1 we would like to ask you some questions about you and your household. Please remember that there is no right or wrong answer, and your answers will be anonymised for any research purposes. This means that your answers will not be able to be personally linked to you or your household, and you will not be personally identifiable by anyone who uses this information for research. Therefore, we ask you to answer as truthfully as possible. After PART 1, we will move onto PART 2 and PART 3.

In PART 2 and 3, you will be asked to do two different activities. In one activity, I will ask you to make a series of THREE choices, which will be explained to you very clearly. Your spouse will be asked to make the same choices. The choices made here will determine the type of gift your household may receive. So your choices have **REAL** consequences. In another activity, I will demonstrate to you a game that you will be asked to complete **TWO** times. Your performance here will determine how much money you might win to take home.

This next part is very important to understand

What you will win to take home will be RANDOMLY selected from PART 2 and PART 3. This means that for your payment **ONLY ONE** of the two activities will be implemented. At the end of the interview, our computer will randomly select ONE from the three choices you made, the three choices your spouse made, and the two tasks which you perform, to determine what you will take home. Therefore, **BOTH** activities are equally important because you do not know which one will be chosen for your final payment. Once you receive the final compensation, you can leave. Any money you earn will be paid out to you privately and confidentially **IMMEDIATELY** after the exercise is complete, with the use of Tigo Peza to a mobile number that you specify.

ROUND 1

[*Explain PART III using the following script.*]

I will now ask you to perform a task which I will explain in a minute. The task requires you to separate the match sticks in this box into the separate colours.

[Enumerator must now show them how to do this by taking 10 matches out of the box and putting them into two separate stacks; after this, the enumerator puts all matches on the table, and mixes them very well. Then continue with explanation below].

Those who finish the task in 30 seconds will earn Tsh 4,000. For every second after 30 seconds, we deduct Tsh 10 from this. So for example, if you finish the task in 45 seconds, you will earn Tsh 3,850. We will ask you to do this task twice. So there will be two rounds of this game. Only ONE round will be randomly selected for payment, so they are both equally important. Before we begin, I would like to ask you the following questions.

[Enumerator now asks self-confidence questions provided in Appendix E.]

[*You must now ask the respondent to perform the task as a first attempt, and then record their timing in minutes and seconds.*]

{Respondent performs Round 1 of task.}

ROUND 2

[*Please read out the following script.*]

Your spouse has been doing the same task in the next room. We would like to ask you some questions about what you think about their performance.

[Enumerator now asks self-confidence questions provided in Appendix E, but for the spouse.]

Control treatment (T1):

- *You must now ask the spouse to perform the task as a second attempt, and then record the completion time.*
- **{Respondent performs Round 2 of task.}**

Confidence treatment (T2):

- *You must now **FILL IN** your task card with **ONLY** the information about spousal **rankings**. This means that you must fill in **ONLY** the table. Do **NOT** fill in the timings. Then go find your enumerator partner. You need their help to fill in the second column with the spouse's information.*
- *You must now **TELL** your respondent what their spouse said about their performance, by showing them the paper. You must now ask the respondent to perform the task as a first attempt, and then record their timing in minutes and seconds.*
- *You must now ask the spouse to perform the task as a second attempt, and then record the completion time.*
- **{Respondent performs Round 2 of task.}**

Comparison treatment (T3):

- *You must now **FILL IN** your task card with **ONLY** the information about spousal **timings**. This means that you must fill in **ONLY** the time. Do **NOT** fill in the table. Then go find your enumerator partner. You need their help to fill in the spouse's time.*
- *You must now **TELL** your respondent what their spouse's Round 1 timing was, and remind them what their own Round 1 timing was, by showing them the paper.*
- *You must now ask the spouse to perform the task as a second attempt, and then record the completion time.*
- **{Respondent performs Round 2 of task.}**

Combined treatment (T4):

- *You must now **FILL IN** your task card with **BOTH** the information about spousal **rankings** AND spousal **timings**. Then go find your enumerator partner to fill in **BOTH** sections with the spouse's information. You need their help to fill in the second column with the spouse's information.*
- *You must now **TELL** your respondent what their spouse said about their performance, by showing them the paper.*
- *You must now **TELL** your respondent what their spouse's Round 1 timing was, and remind them what their own Round 1 timing was, by showing them the paper.*
- *You must now ask the spouse to perform the task as a second attempt, and then record the completion time.*
- **{Respondent performs Round 2 of task.}**

END

[*Please read out the following script.*]

Before we end, I would like to ask you the following questions

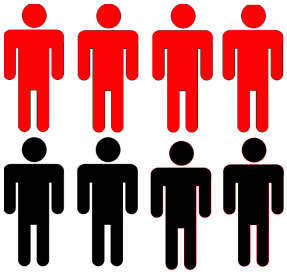
[Enumerator now repeats self-confidence questions provided in Appendix E.]

E. Self-confidence/spouse-confidence beliefs questions

For each set of questions eliciting self-confidence beliefs, we use the following set of questions. Note that the images use male or female stick figures, depending on whether the question is related to the husband or the wife, respectively. Also note that the same questions are used for the 'spouse-confidence' belief elicitation. The only difference is that instead of referring to 'your performance' and 'your timing' we ask about 'your spouse's performance' and 'your spouse's timing'.

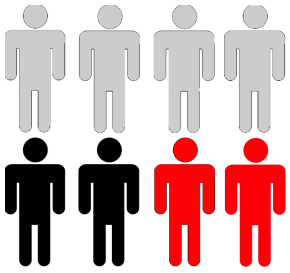
Q1 If we compare the timing of your performance in Round 1 with eight other men in today's session, do you think your timing would be in the top half or the bottom half? (Enumerator now shows the respondent the following image).

- I will be in the bottom half (if selected, go to Q2)
- I will be in the top half (if selected, go to Q4)



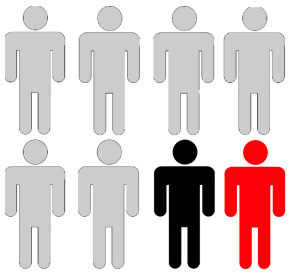
Q2 If we compare the timing of your performance in Round 1 with the bottom four men in today's session, do you think your timing would be better than the bottom two men, or would you be in the bottom two? (Enumerator now shows the respondent the following image).

- I will be in the bottom two (if selected, go to Q3)
- I will not be in the bottom two (if selected, stop ranking questions)



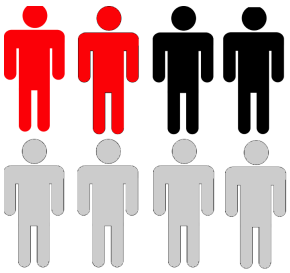
Q3 If we compare the timing of your performance in Round 1 with the bottom two men in today's session, do you think your timing would make you the bottom person or not? (Enumerator now shows the respondent the following image).

- I will be the bottom person (if selected, stop ranking questions)
- I will not be the bottom person (if selected, stop ranking questions)



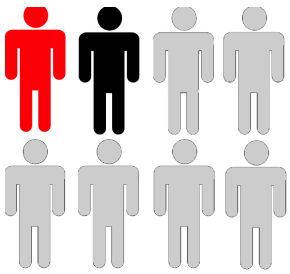
Q4 If we compare the timing of your performance in Round 1 with the top four other men in today's session, do you think your timing would be in the top two men or not? (Enumerator now shows the respondent the following image).

- I will not be in the top two (if selected, stop ranking questions)
- I will be in the top two (if selected, go to Q5)



Q5 If we compare the timing of your performance in Round 1 with the top two men in today's session, do you think your timing would make you the top person or not? (Enumerator now shows the respondent the following image).

- I will not be the top person (if selected, stop ranking questions)
- I will be the top person (if selected, stop ranking questions)



F. Average treatment effects

Table F.1: Average treatment effects (pooled sample)

	Time improvements		% Time improvements	
	(1)	(2)	(3)	(4)
T2: Beliefs	0.559	0.426	1.333	0.721
	(0.518)	(0.476)	(1.164)	(0.981)
T3: Times	0.590	0.369	1.573	1.051
	(0.589)	(0.538)	(1.364)	(1.274)
T4: Both	0.652	0.298	1.671	0.832
	(0.702)	(0.619)	(1.510)	(1.354)
Enumerator FEs & controls		✓		✓
Predicted mean under T1	5.17	5.34	12.16	12.65
F-test p-value for T2 vs T4	0.872	0.811	0.794	0.930
F-test p-value for T3 vs T4	0.869	0.828	0.923	0.810

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(2)), and % timing improvements between rounds (Models (3)-(4)). All regressions have 1269 observations. Enumerator fixed effects are included in Models (2) and (4), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F.2: Average treatment effects (gender-disaggregated sample with cutoff=120)

	Time improvements		% Time improvements		Time improvements		% Time improvements	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.044	0.925	-0.308	0.971	1.351	1.452	0.441	1.028
	(0.804)	(1.068)	(0.872)	(1.048)	(1.573)	(1.963)	(1.454)	(1.927)
T3: Times	0.344	0.112	0.108	-0.113	1.992	0.933	1.896	0.097
	(0.915)	(1.233)	(0.949)	(1.116)	(1.722)	(2.405)	(1.703)	(2.250)
T4: Both	-0.244	0.519	-0.498	0.095	0.370	2.523	-0.371	1.750
	(0.933)	(1.094)	(0.933)	(0.985)	(1.801)	(2.091)	(1.739)	(1.861)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.53	4.83	6.74	4.98	14.96	9.78	15.40	10.28
F-test p-value for T2 vs T4	0.708	0.702	0.796	0.342	0.575	0.567	0.655	0.656
F-test p-value for T3 vs T4	0.393	0.629	0.230	0.758	0.331	0.378	0.105	0.294

Notes: OLS regressions. Wife and husband regressions have 640 and 647 observations, respectively. Dependent variables are absolute timing improvements between rounds (calculated as timing in Round 1 minus timing in Round 2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F.3: Average treatment effects for couples

	Total time improvements (couples)
T2: Beliefs	0.948
	(1.051)
T3: Times	1.245
	(1.196)
T4: Both	1.178
	(1.414)
Mean in T1	10.36
p-value: T2=T4	0.847
p-value: T3=T4	0.930
N	628

Notes: OLS regression with 628 couples. The dependent variable is the total timing improvements of the couple (calculated as the sum of the timing improvements of husband and wife). Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G. Robustness checks

G.1. Regressions with ahead and behind in the same model

In the main text, 'Ahead' and 'Behind' were never included in the same regression model. We also did not combine 'Ahead & Poseval/Negeval' with 'Behind & Poseval/Negeval'. Table G.1 and Table G.2 present these combined models.

Table G.1: Ahead & Behind

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.773	0.901	0.710	0.429	3.612	1.931	3.456	0.458
	(1.024)	(1.155)	(1.012)	(1.103)	(2.450)	(2.782)	(2.451)	(2.689)
T3: Times	0.279	-0.049	0.049	-0.390	3.124	1.696	2.749	0.345
	(1.139)	(1.140)	(1.083)	(1.125)	(2.782)	(2.870)	(2.683)	(2.874)
T4: Both	-0.473	1.331	-0.343	1.522	-0.471	4.360*	0.082	4.874**
	(0.966)	(1.038)	(0.961)	(1.003)	(2.506)	(2.516)	(2.449)	(2.461)
Ahead	-0.046	0.069	-0.030	0.048	-0.111	0.078	-0.031	0.064
	(0.051)	(0.110)	(0.054)	(0.100)	(0.133)	(0.269)	(0.140)	(0.246)
T2: Beliefs × Ahead	-0.090	-0.053	-0.121	-0.009	-0.288	-0.127	-0.377**	-0.018
	(0.071)	(0.136)	(0.074)	(0.121)	(0.176)	(0.342)	(0.186)	(0.298)
T3: Times × Ahead	-0.059	-0.074	-0.046	-0.047	-0.244	-0.299	-0.225	-0.224
	(0.084)	(0.126)	(0.083)	(0.120)	(0.230)	(0.313)	(0.228)	(0.300)
T4: Both × Ahead	0.026	-0.126	-0.038	-0.167	0.012	-0.329	-0.181	-0.435
	(0.075)	(0.122)	(0.078)	(0.119)	(0.211)	(0.308)	(0.216)	(0.300)
Behind	0.296***	0.272***	0.279***	0.277***	0.334**	0.336***	0.379**	0.362***
	(0.078)	(0.053)	(0.078)	(0.053)	(0.165)	(0.114)	(0.169)	(0.132)
T2: Beliefs × Behind	0.030	0.012	0.011	0.024	-0.151	0.008	-0.214	0.054
	(0.113)	(0.094)	(0.108)	(0.101)	(0.217)	(0.175)	(0.215)	(0.193)
T3: Times × Behind	0.067	0.217**	0.082	0.197**	-0.064	0.282	-0.033	0.273
	(0.132)	(0.093)	(0.123)	(0.095)	(0.232)	(0.180)	(0.223)	(0.194)
T4: Both × Behind	0.127	0.051	0.110	-0.002	0.198	-0.010	0.154	-0.143
	(0.120)	(0.079)	(0.120)	(0.087)	(0.235)	(0.145)	(0.241)	(0.169)
Eumerator FEs & controls			✓	✓			✓	✓
N	631	638	631	638	631	638	631	638

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Behind is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.2: Ahead & Behind, Poseval & Negeval

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.181	0.831	0.078	0.666	0.942	1.657	0.533	0.915
	(0.892)	(0.840)	(0.862)	(0.835)	(1.930)	(1.962)	(1.880)	(1.922)
T3: Times	0.184	0.559	0.074	0.191	1.432	2.317	1.420	1.036
	(0.951)	(0.841)	(0.917)	(0.832)	(2.110)	(2.024)	(1.999)	(1.975)
T4: Both	-0.357	1.650**	-0.391	1.317*	-0.204	4.117**	-0.340	3.279*
	(0.844)	(0.797)	(0.845)	(0.780)	(1.931)	(1.803)	(1.948)	(1.768)
aheadnpos	0.031	0.174	-0.028	0.123	0.175***	0.396	0.086	0.294
	(0.024)	(0.195)	(0.047)	(0.184)	(0.061)	(0.461)	(0.103)	(0.439)
T2: Beliefs×aheadnpos	-0.305*	-0.083	-0.310	-0.090	-0.290	-0.211	-0.271	-0.221
	(0.157)	(0.228)	(0.199)	(0.220)	(0.451)	(0.566)	(0.556)	(0.541)
T3: Times×aheadnpos	-0.180*	-0.236	-0.200	-0.201	-0.227	-0.641	-0.304	-0.529
	(0.092)	(0.204)	(0.135)	(0.198)	(0.330)	(0.496)	(0.305)	(0.497)
T4: Both×aheadnpos	-0.467*	-0.263	-0.571**	-0.199	-1.604	-0.504	-1.818*	-0.390
	(0.268)	(0.201)	(0.284)	(0.206)	(0.989)	(0.476)	(1.045)	(0.477)
aheadnneg	-0.095	-0.007	-0.073	0.023	-0.217	-0.043	-0.115	0.113
	(0.067)	(0.129)	(0.073)	(0.161)	(0.168)	(0.307)	(0.185)	(0.364)
T2: Beliefs×aheadnneg	-0.058	-0.106	-0.119	-0.100	-0.143	-0.231	-0.284	-0.280
	(0.083)	(0.140)	(0.089)	(0.180)	(0.207)	(0.337)	(0.223)	(0.413)
T3: Times×aheadnneg	-0.093	-0.047	-0.106	-0.103	-0.322	-0.290	-0.367	-0.524
	(0.138)	(0.162)	(0.127)	(0.183)	(0.364)	(0.385)	(0.345)	(0.419)
T4: Both×aheadnneg	0.080	-0.285*	0.031	-0.381**	0.214	-0.676*	0.052	-0.870**
	(0.114)	(0.156)	(0.111)	(0.185)	(0.315)	(0.385)	(0.285)	(0.426)
behindnpos	-0.013	0.199**	-0.071	0.147	-0.167	0.241*	-0.057	0.109
	(0.091)	(0.086)	(0.115)	(0.095)	(0.193)	(0.145)	(0.258)	(0.170)
T2: Beliefs×behindnpos	0.301	0.175	0.317	0.206	0.475	0.347	0.251	0.405*
	(0.295)	(0.135)	(0.307)	(0.134)	(0.591)	(0.216)	(0.636)	(0.220)
T3: Times×behindnpos	0.237	0.309***	0.158	0.333***	0.479	0.369**	0.213	0.481**
	(0.297)	(0.114)	(0.318)	(0.123)	(0.628)	(0.182)	(0.711)	(0.209)
T4: Both×behindnpos	0.577***	0.142	0.560***	0.150	0.920***	0.149	0.714*	0.172
	(0.196)	(0.105)	(0.194)	(0.122)	(0.339)	(0.167)	(0.367)	(0.192)
behindnneg	0.219*	0.449***	0.226*	0.420***	0.202	0.886***	0.334	0.792***
	(0.126)	(0.103)	(0.120)	(0.121)	(0.231)	(0.241)	(0.214)	(0.301)
T2: Beliefs×behindnneg	0.314	-0.428*	0.178	-0.397	0.626	-0.901**	0.462	-0.737
	(0.263)	(0.224)	(0.294)	(0.277)	(0.456)	(0.427)	(0.529)	(0.510)
T3: Times×behindnneg	0.151	-1.261***	0.141	-1.061***	0.226	-2.918***	0.202	-2.498***
	(0.199)	(0.136)	(0.173)	(0.181)	(0.326)	(0.353)	(0.285)	(0.445)
T4: Both×behindnneg	0.024	-0.391*	-0.112	-0.405*	0.086	-0.743	-0.091	-0.723
	(0.341)	(0.229)	(0.353)	(0.213)	(0.676)	(0.598)	(0.664)	(0.546)
Enumerator FEs & controls			✓	✓			✓	✓
N	631	638	631	638	631	638	631	638

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind and Positive' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. 'Ahead and Positive' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.2. Regressions with spouse-confidence beliefs

Table G.3: Timings with stated spouse-confidence beliefs

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.754	1.002	-0.834	0.911	-0.858	2.012	-1.358	1.751
	(0.923)	(1.154)	(0.866)	(1.112)	(1.980)	(2.861)	(1.834)	(2.774)
T3: Times	0.426	1.139	0.379	1.104	1.283	2.370	1.558	2.079
	(0.852)	(1.394)	(0.755)	(1.284)	(1.743)	(3.387)	(1.575)	(3.237)
T4: Both	-0.025	1.332	-0.272	0.853	-0.322	3.978	-0.783	2.956
	(0.930)	(1.171)	(0.807)	(0.887)	(1.968)	(2.699)	(1.733)	(2.125)
Spouse confidence	0.248	1.595	-0.004	1.093	-1.622	2.918	-0.803	1.476
	(2.319)	(1.442)	(2.312)	(1.395)	(4.454)	(3.198)	(4.195)	(3.265)
T2 × Spouse confidence	8.678***	-0.400	8.122**	-0.087	16.416***	-1.145	15.042**	-1.508
	(3.007)	(1.778)	(2.958)	(1.909)	(5.625)	(4.201)	(5.481)	(4.480)
T3 × Spouse confidence	2.223	-1.526	2.114	-1.843	5.195	-2.656	3.956	-3.540
	(2.874)	(2.175)	(2.574)	(2.106)	(4.753)	(4.709)	(4.021)	(4.840)
T4 × Spouse confidence	0.162	-0.057	0.102	0.361	3.071	-2.156	1.542	-1.307
	(2.542)	(1.766)	(2.564)	(1.572)	(5.141)	(3.621)	(5.459)	(3.403)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.26	4.06	6.38	4.18	15.15	9.13	15.47	9.57
F-test p-value for T2 vs T4	0.343	0.695	0.465	0.934	0.768	0.333	0.755	0.475
F-test p-value for T3 vs T4	0.572	0.846	0.341	0.763	0.408	0.503	0.192	0.666

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Spouse-confidence' is a binary variable with value one if the respondent answers with 'Not top person' to the question "If we compare your spouse's timing in Round 1 with eight other men/women in the village, will s/he be the top 1 of 8?", and zero if the respondent answered with 'Top person'. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.3. Regressions with absolute spousal evaluation

Table G.4: Timings with absolute spouse evaluations

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.265	0.938	0.427	0.984	0.352	2.013	0.417	1.672
	(0.902)	(0.765)	(0.937)	(0.744)	(1.964)	(1.810)	(2.032)	(1.780)
T3: Times	0.598	1.249	0.516	1.125	1.824	3.101	2.062	2.548
	(1.319)	(0.912)	(1.220)	(0.821)	(2.754)	(2.195)	(2.544)	(2.011)
T4: Both	0.305	1.417	0.039	1.213	0.160	3.926*	-0.493	3.443
	(0.955)	(0.991)	(0.954)	(0.861)	(1.961)	(2.260)	(2.098)	(2.050)
Spousal evaluation	0.730	2.370	0.988	2.571	-0.166	6.819*	0.957	7.334*
	(1.520)	(1.534)	(1.252)	(1.651)	(3.137)	(3.458)	(2.640)	(3.552)
T2 × Spousal evaluation	0.327	-1.081	-0.696	-1.364	2.563	-3.930	0.285	-4.511
	(1.903)	(1.747)	(1.882)	(1.752)	(3.827)	(3.602)	(3.744)	(3.380)
T3 × Spousal evaluation	0.096	-5.725**	-0.012	-6.317***	0.124	-14.283**	-0.493	-16.205***
	(2.021)	(2.140)	(1.855)	(1.914)	(4.266)	(5.115)	(3.651)	(4.460)
T4 × Spousal evaluation	-0.733	-1.560	-0.859	-2.325	0.022	-6.364	-0.284	-8.093*
	(1.740)	(2.095)	(1.668)	(1.787)	(4.052)	(4.615)	(3.861)	(4.110)
Enumerator FEs & controls			✓	✓			✓	✓
Predicted mean under T1	6.25	4.11	6.40	4.24	15.16	9.25	15.53	9.74
F-test p-value for T2 vs T4	0.969	0.616	0.650	0.782	0.929	0.378	0.658	0.366
F-test p-value for T3 vs T4	0.784	0.818	0.566	0.899	0.455	0.645	0.162	0.599

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Spouse evaluation' is a binary variable with value one if the respondent's *spouse* answers with 'Not top person' to the question "If we compare your spouse's timing in Round 1 with eight other men/women in the village, will s/he be the top 1 of 8?", and zero if their spouse answered with 'Top person'. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.4. Regressions with cut-off at 120 seconds

In the main analysis, we restricted the sample to couples where spouses had a 60 second time difference in performances in Round 1 (along with some further restrictions). This was to capture treatment effects within a highly competitive environment. Here, we adjust this cut-off point to time differences of 120 seconds between spouses (imposing the same additional restrictions). Thus, the observations for wives rise from 631 to 640, and from 638 to 647 for husbands.

Table G.5: Timings with 'Ahead' (cut-off=120)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.276	1.326	-0.392	1.332	0.825	2.259	0.736	1.731
	(1.200)	(1.275)	(1.280)	(1.147)	(1.960)	(2.246)	(1.886)	(1.864)
T3: Times	0.055	1.098	-0.110	0.897	1.519	3.420	1.639	2.514
	(1.239)	(1.365)	(1.278)	(1.309)	(2.189)	(2.488)	(2.136)	(2.446)
T4: Both	-0.572	1.169	-0.533	0.918	-0.179	4.043**	0.162	3.655**
	(1.343)	(1.077)	(1.314)	(0.843)	(2.404)	(1.859)	(2.256)	(1.482)
Ahead	-0.159**	-0.044	-0.157**	-0.031	-0.285**	-0.023	-0.256**	0.013
	(0.060)	(0.059)	(0.057)	(0.057)	(0.112)	(0.128)	(0.103)	(0.122)
T2: Beliefs × Ahead	0.040	-0.101	0.022	-0.081	0.065	-0.191	-0.025	-0.148
	(0.084)	(0.128)	(0.077)	(0.112)	(0.144)	(0.289)	(0.122)	(0.239)
T3: Times × Ahead	0.029	-0.179*	0.038	-0.172	0.046	-0.456*	0.051	-0.428*
	(0.087)	(0.098)	(0.081)	(0.101)	(0.187)	(0.226)	(0.169)	(0.217)
T4: Both × Ahead	0.031	-0.179**	-0.016	-0.214***	0.049	-0.395*	-0.120	-0.477**
	(0.091)	(0.073)	(0.094)	(0.070)	(0.222)	(0.200)	(0.217)	(0.199)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regressions. Wife and husband regressions have 640 and 647 observations, respectively. Dependent variables are absolute timing improvements between rounds (calculated as timing in Round 1 minus timing in Round 2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.6: Timings with 'Behind' (cut-off=120)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.088	-0.127	0.730	-0.142	1.939	0.972	0.741	0.643
	(0.986)	(1.041)	(0.840)	(1.039)	(1.727)	(2.326)	(1.552)	(2.380)
T3: Times	0.827	0.949	0.657	0.957	1.770	1.224	1.460	0.713
	(0.932)	(0.913)	(0.824)	(0.967)	(1.912)	(2.537)	(1.815)	(2.464)
T4: Both	0.695	1.429	0.239	1.309*	0.055	2.761	-1.212	2.480
	(0.939)	(0.881)	(0.829)	(0.727)	(1.895)	(2.201)	(1.808)	(1.938)
Behind	0.524***	0.432***	0.509***	0.440***	0.396***	0.385***	0.364***	0.391***
	(0.122)	(0.059)	(0.123)	(0.057)	(0.082)	(0.063)	(0.082)	(0.065)
T2: Beliefs × Behind	-0.148	0.137	-0.145	0.118	-0.066	0.060	-0.012	0.024
	(0.164)	(0.113)	(0.151)	(0.115)	(0.168)	(0.107)	(0.149)	(0.134)
T3: Times × Behind	-0.118	-0.086	-0.108	-0.113	0.019	-0.007	0.082	-0.049
	(0.133)	(0.090)	(0.121)	(0.093)	(0.167)	(0.119)	(0.159)	(0.120)
T4: Both × Behind	-0.090	-0.094	-0.063	-0.124	0.186	0.004	0.296	-0.055
	(0.170)	(0.082)	(0.175)	(0.088)	(0.182)	(0.097)	(0.202)	(0.113)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Behind is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.7: Timings with 'Ahead' & 'Behind' (cut-off=120)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.175	-0.210	1.107	-0.285	1.633	1.707	1.400	1.299
	(1.060)	(1.366)	(1.018)	(1.249)	(1.877)	(2.809)	(1.902)	(2.621)
T3: Times	0.864	1.924*	0.656	1.896*	1.486	4.125	1.177	3.566
	(1.133)	(0.991)	(1.033)	(1.058)	(2.347)	(2.632)	(2.260)	(2.754)
T4: Both	0.480	2.330***	0.525	2.510***	-0.934	4.788**	-0.707	5.242***
	(1.102)	(0.778)	(0.918)	(0.600)	(2.334)	(1.863)	(2.117)	(1.540)
Ahead=1	-0.044	0.087	-0.036	0.097*	-0.210**	0.094	-0.176*	0.130
	(0.029)	(0.054)	(0.023)	(0.047)	(0.098)	(0.118)	(0.089)	(0.110)
T2: Beliefs × Ahead=1	-0.009	0.031	-0.038	0.041	0.038	-0.112	-0.056	-0.085
	(0.029)	(0.105)	(0.030)	(0.087)	(0.106)	(0.274)	(0.095)	(0.223)
T3: Times × Ahead=1	-0.001	-0.144	0.008	-0.142*	0.044	-0.423*	0.063	-0.408*
	(0.056)	(0.084)	(0.053)	(0.075)	(0.168)	(0.214)	(0.153)	(0.208)
T4: Both × Ahead=1	0.024	-0.144**	-0.035	-0.200**	0.110	-0.345*	-0.064	-0.472**
	(0.060)	(0.061)	(0.061)	(0.071)	(0.205)	(0.191)	(0.199)	(0.199)
Behind=1	0.511***	0.451***	0.500***	0.461***	0.334***	0.407***	0.319***	0.419***
	(0.123)	(0.057)	(0.125)	(0.055)	(0.079)	(0.063)	(0.078)	(0.072)
T2: Beliefs × Behind=1	-0.158	0.139	-0.167	0.122	-0.078	0.035	-0.061	0.002
	(0.169)	(0.117)	(0.162)	(0.118)	(0.170)	(0.110)	(0.153)	(0.144)
T3: Times × Behind=1	-0.128	-0.118	-0.114	-0.145	-0.002	-0.102	0.065	-0.146
	(0.142)	(0.089)	(0.133)	(0.096)	(0.164)	(0.102)	(0.158)	(0.121)
T4: Both × Behind=1	-0.087	-0.128	-0.090	-0.172*	0.198	-0.080	0.220	-0.173
	(0.177)	(0.084)	(0.180)	(0.095)	(0.191)	(0.094)	(0.200)	(0.116)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Behind is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.8: Timings with 'Spousal evaluations' (cut-off=120)

	Positive evaluations				Negative evaluations			
	Time improvements		% Time improvements		Time improvements		% Time improvements	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.058	0.549	0.720	0.616	-0.034	1.085	0.296	1.383
	(0.676)	(0.874)	(1.381)	(2.110)	(0.740)	(0.718)	(1.784)	(1.635)
T3: Times	0.047	0.433	1.825	1.577	0.250	0.602	1.875	1.655
	(0.665)	(1.086)	(1.444)	(2.449)	(1.085)	(0.830)	(2.410)	(1.847)
T4: Both	-0.386	0.223	-0.438	1.471	0.108	1.008	-0.171	3.179
	(0.766)	(0.821)	(1.606)	(1.819)	(0.839)	(0.824)	(1.897)	(1.889)
Spouse evaluation	-3.224	-0.602	-5.771	-0.585	0.863	3.831*	-0.176	9.966*
	(2.207)	(1.633)	(4.919)	(3.469)	(1.061)	(2.023)	(2.329)	(4.999)
T2: Beliefs × Spouse evaluation	0.529	0.793	-0.691	0.756	0.422	-5.201	1.723	-9.507
	(2.681)	(1.319)	(5.857)	(3.299)	(1.303)	(3.096)	(2.976)	(6.108)
T3: Timings × Spouse evaluation	1.002	-0.668	1.558	-3.434	-0.395	-7.248**	0.182	-21.395***
	(2.177)	(1.955)	(4.954)	(4.398)	(1.607)	(3.360)	(3.632)	(7.376)
T4: Both × Spouse evaluation	2.772	1.424	3.520	2.554	-0.637	-6.034**	-0.219	-15.635**
	(2.694)	(1.642)	(6.382)	(3.730)	(1.390)	(2.586)	(3.229)	(6.800)

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1), (2), (5) and (6)), and % timing improvements between rounds (Models (3), (4), (7) and (8)). 'Spousal evaluation' is a dummy variable for positive evaluations in Models (1)-(4), and a dummy for negative evaluations in Models (5)-(8). The positive evaluation dummy takes the value 1 if the self-confidence dummy is 1 (respondent believes they are not top person) whilst the spouse-confidence dummy is 0 (spouse believes the respondent is the top person), zero otherwise. The negative evaluation dummy takes the value 1 if the self-confidence dummy is 0 (respondent believes they are the top person) whilst the spouse-confidence dummy is 1 (spouse believes the respondent is not top person), zero otherwise. Wife and husband regressions have 640 and 647 observations, respectively. All models have enumerator fixed effects, along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Also included are control variables 'Ahead' and 'Behind': 'Ahead' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise, and 'Behind' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.9: Timings with 'Ahead & Positive/Negative' (cut-off=120)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.097	1.038	-0.462	1.136	1.094	1.708	0.281	1.423
	(0.952)	(1.104)	(1.086)	(1.062)	(1.697)	(2.007)	(1.669)	(1.896)
T3: Times	0.345	0.441	0.072	0.312	2.199	1.896	2.150	1.262
	(1.092)	(1.233)	(1.113)	(1.123)	(2.013)	(2.286)	(1.907)	(2.114)
T4: Both	-0.323	0.824	-0.660	0.403	0.447	3.210	-0.348	2.506
	(1.055)	(1.092)	(1.080)	(0.987)	(1.716)	(2.053)	(1.780)	(1.834)
Ahead and Positive	-0.164	0.037	-0.204*	0.049	-0.332	0.127	-0.362	0.158
	(0.120)	(0.066)	(0.109)	(0.055)	(0.308)	(0.151)	(0.270)	(0.132)
T2: Beliefs × Ahead and Positive	-0.153	-0.015	-0.205	-0.077	0.169	-0.014	0.089	-0.116
	(0.191)	(0.134)	(0.201)	(0.139)	(0.521)	(0.356)	(0.534)	(0.331)
T3: Times × Ahead and Positive	0.041	-0.160	0.015	-0.186*	0.111	-0.429	-0.050	-0.456
	(0.120)	(0.108)	(0.125)	(0.107)	(0.317)	(0.296)	(0.281)	(0.292)
T4: Both × Ahead and Positive	-0.341	-0.160**	-0.431*	-0.145	-1.187*	-0.277*	-1.473**	-0.266
	(0.210)	(0.066)	(0.221)	(0.086)	(0.579)	(0.158)	(0.607)	(0.182)
Ahead and Negative	-0.109**	-0.076	-0.115**	-0.072	-0.195	-0.122	-0.194*	0.075
	(0.052)	(0.121)	(0.047)	(0.152)	(0.124)	(0.292)	(0.100)	(0.326)
T2: Beliefs × Ahead and Negative	0.012	-0.083	0.015	-0.039	-0.015	-0.200	-0.039	-0.282
	(0.067)	(0.149)	(0.075)	(0.169)	(0.154)	(0.352)	(0.158)	(0.369)
T3: Times × Ahead and Negative	-0.148	-0.035	-0.121	-0.107	-0.430	-0.264	-0.383	-0.590
	(0.149)	(0.181)	(0.126)	(0.186)	(0.395)	(0.400)	(0.351)	(0.402)
T4: Both × Ahead and Negative	0.047	-0.256	0.041	-0.320*	0.131	-0.645	0.083	-0.896**
	(0.118)	(0.156)	(0.095)	(0.165)	(0.361)	(0.399)	(0.261)	(0.415)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead and Positive' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.10: Timings with 'Behind & Positive/Negative' (cut-off=120)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.539	1.486	0.273	1.463	1.005	1.719	0.121	1.171
	(0.850)	(1.194)	(0.784)	(1.174)	(1.795)	(2.139)	(1.648)	(2.094)
T3: Times	0.581	0.687	0.376	0.470	1.460	1.659	1.250	0.752
	(1.042)	(1.151)	(0.991)	(1.108)	(2.037)	(2.357)	(1.960)	(2.259)
T4: Both	0.158	1.002	-0.086	0.657	-0.052	2.966	-0.746	2.219
	(0.962)	(1.059)	(0.921)	(1.010)	(2.013)	(2.113)	(1.918)	(1.931)
Behind and Positive	0.043	0.407**	-0.011	0.361**	-0.096	0.382***	-0.051	0.303*
	(0.102)	(0.147)	(0.131)	(0.167)	(0.196)	(0.132)	(0.234)	(0.167)
T2: Beliefs × Behind and Positive	0.271	-0.087	0.243	-0.084	0.458	0.170	0.240	0.157
	(0.257)	(0.111)	(0.271)	(0.114)	(0.513)	(0.170)	(0.510)	(0.164)
T3: Times × Behind and Positive	0.229	-0.139	0.130	-0.145	0.520	-0.074	0.282	-0.055
	(0.302)	(0.216)	(0.332)	(0.236)	(0.632)	(0.239)	(0.695)	(0.278)
T4: Both × Behind and Positive	0.543**	-0.057	0.561**	-0.079	0.912**	0.028	0.827**	0.011
	(0.204)	(0.161)	(0.203)	(0.179)	(0.328)	(0.143)	(0.289)	(0.171)
Behind and Negative	0.569***	0.426***	0.567***	0.381***	0.337***	0.849***	0.327***	0.725***
	(0.100)	(0.106)	(0.093)	(0.107)	(0.055)	(0.237)	(0.051)	(0.251)
T2: Beliefs × Behind and Negative	0.005	-0.460*	-0.077	-0.377	0.579	-0.900*	0.603	-0.695
	(0.233)	(0.241)	(0.271)	(0.286)	(0.373)	(0.438)	(0.471)	(0.497)
T3: Times × Behind and Negative	-0.176	-1.280***	-0.164	-1.083***	0.146	-2.879***	0.255	-2.447***
	(0.149)	(0.140)	(0.126)	(0.148)	(0.221)	(0.339)	(0.214)	(0.369)
T4: Both × Behind and Negative	-0.298	-0.349	-0.364	-0.307	0.036	-0.668	0.015	-0.542
	(0.375)	(0.226)	(0.416)	(0.231)	(0.601)	(0.573)	(0.671)	(0.520)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind and Positive' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.11: Timings with 'Ahead/Behind & Positive/Negative' (cut-off=120)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.499	1.641	0.287	1.661	0.739	2.007	0.086	1.594
	(0.892)	(1.224)	(0.864)	(1.189)	(1.865)	(2.152)	(1.798)	(2.044)
T3: Times	0.660	1.049	0.473	0.907	1.720	2.735	1.662	2.003
	(1.227)	(1.096)	(1.134)	(1.076)	(2.405)	(2.147)	(2.254)	(2.052)
T4: Both	0.137	1.356	-0.116	1.001	0.014	3.741*	-0.587	3.047
	(0.994)	(1.073)	(0.951)	(1.038)	(1.801)	(2.090)	(1.806)	(1.924)
Ahead and Positive	-0.136	0.059	-0.169*	0.063	-0.317	0.151	-0.338	0.168
	(0.112)	(0.070)	(0.097)	(0.059)	(0.306)	(0.156)	(0.263)	(0.137)
T2: Beliefs × Ahead and Positive	-0.125	-0.013	-0.196	-0.059	0.236	0.004	0.128	-0.076
	(0.178)	(0.126)	(0.202)	(0.126)	(0.509)	(0.341)	(0.542)	(0.313)
T3: Times × Ahead and Positive	0.039	-0.156	0.006	-0.178	0.127	-0.431	-0.034	-0.445
	(0.115)	(0.105)	(0.107)	(0.108)	(0.322)	(0.293)	(0.275)	(0.296)
T4: Both × Ahead and Positive	-0.300	-0.147*	-0.385*	-0.128	-1.111*	-0.259	-1.384**	-0.241
	(0.196)	(0.072)	(0.197)	(0.091)	(0.564)	(0.162)	(0.587)	(0.187)
Ahead and Negative	-0.062	-0.021	-0.062	-0.027	-0.171	-0.061	-0.160	0.124
	(0.046)	(0.115)	(0.038)	(0.152)	(0.124)	(0.285)	(0.095)	(0.331)
T2: Beliefs × Ahead and Negative	-0.014	-0.122	-0.022	-0.081	-0.009	-0.233	-0.047	-0.324
	(0.059)	(0.142)	(0.061)	(0.163)	(0.147)	(0.341)	(0.145)	(0.370)
T3: Times × Ahead and Negative	-0.125	-0.066	-0.106	-0.107	-0.373	-0.305	-0.323	-0.598
	(0.158)	(0.166)	(0.132)	(0.178)	(0.404)	(0.388)	(0.361)	(0.408)
T4: Both × Ahead and Negative	0.047	-0.271	0.030	-0.319*	0.168	-0.659	0.103	-0.893*
	(0.118)	(0.159)	(0.087)	(0.169)	(0.359)	(0.411)	(0.254)	(0.434)
Behind and Positive	0.021	0.411**	-0.047	0.363**	-0.152	0.390***	-0.128	0.311*
	(0.098)	(0.148)	(0.126)	(0.168)	(0.183)	(0.132)	(0.216)	(0.166)
T2: Beliefs × Behind and Positive	0.278	-0.093	0.238	-0.092	0.484	0.160	0.221	0.142
	(0.259)	(0.112)	(0.275)	(0.113)	(0.517)	(0.169)	(0.512)	(0.155)
T3: Times × Behind and Positive	0.206	-0.151	0.078	-0.159	0.456	-0.108	0.151	-0.096
	(0.307)	(0.215)	(0.330)	(0.237)	(0.636)	(0.232)	(0.691)	(0.276)
T4: Both × Behind and Positive	0.544**	-0.070	0.549**	-0.094	0.905**	-0.001	0.772**	-0.023
	(0.211)	(0.165)	(0.211)	(0.184)	(0.328)	(0.150)	(0.292)	(0.179)
Behind and Negative	0.562***	0.432***	0.564***	0.384***	0.320***	0.864***	0.319***	0.736**
	(0.101)	(0.107)	(0.094)	(0.111)	(0.055)	(0.236)	(0.047)	(0.257)
T2: Beliefs × Behind and Negative	-0.013	-0.468*	-0.109	-0.390	0.545	-0.917*	0.535	-0.719
	(0.234)	(0.242)	(0.275)	(0.293)	(0.373)	(0.438)	(0.472)	(0.510)
T3: Times × Behind and Negative	-0.191	-1.307***	-0.192	-1.103***	0.104	-2.960***	0.174	-2.514***
	(0.156)	(0.136)	(0.130)	(0.148)	(0.226)	(0.321)	(0.215)	(0.347)
T4: Both × Behind and Negative	-0.319	-0.374	-0.399	-0.332	-0.032	-0.721	-0.103	-0.602
	(0.381)	(0.228)	(0.421)	(0.236)	(0.609)	(0.577)	(0.671)	(0.527)
Enumerator FE & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead and Positive' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. 'Behind and Positive' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.5. Regressions with binary ahead/behind

Table G.12: Timings with 'Ahead' (binary)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.345	1.191	-0.210	1.284	0.771	1.503	1.505	1.043
	(1.754)	(1.357)	(1.747)	(1.274)	(2.778)	(2.206)	(2.512)	(1.646)
T3: Times	-0.040	1.374	0.031	1.000	2.192	3.390	2.791	2.171
	(1.571)	(1.487)	(1.596)	(1.476)	(2.904)	(2.249)	(2.953)	(2.305)
T4: Both	-0.231	0.284	-0.082	-0.133	2.062	1.389	2.570	1.045
	(1.746)	(1.434)	(1.744)	(1.156)	(2.729)	(2.404)	(2.846)	(1.723)
Ahead=1	-4.949***	-3.769***	-4.171**	-3.412***	-6.426**	-6.150***	-4.749*	-5.865***
	(1.544)	(0.988)	(1.584)	(1.010)	(2.300)	(1.773)	(2.580)	(1.847)
T2: Beliefs × Ahead=1	0.898	-0.707	0.094	-0.842	1.295	-0.155	-1.520	-0.130
	(2.234)	(1.701)	(2.089)	(1.735)	(3.666)	(3.992)	(3.543)	(4.000)
T3: Times × Ahead=1	-0.101	-1.593	-0.410	-1.297	-1.634	-3.387	-2.462	-2.577
	(1.883)	(1.499)	(1.837)	(1.721)	(4.157)	(2.778)	(4.167)	(2.919)
T4: Both × aAhead=1	-0.435	0.998	-1.207	0.968	-3.853	3.335	-6.271*	2.340
	(1.893)	(1.082)	(1.915)	(0.919)	(3.362)	(2.370)	(3.526)	(1.858)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a binary variable which takes the value 1 when the respondent is ahead, zero otherwise. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.13: Timings with 'Behind' (binary)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.710	-0.245	0.253	-0.087	2.635	-0.598	1.008	-0.859
	(0.865)	(1.187)	(0.730)	(1.269)	(2.145)	(3.209)	(2.033)	(3.554)
T3: Times	0.161	-0.719	0.127	-0.570	1.393	-1.441	1.632	-1.669
	(0.885)	(1.096)	(0.857)	(1.112)	(2.344)	(2.908)	(2.278)	(2.906)
T4: Both	-0.371	0.701	-0.850	0.486	-0.758	3.064	-2.356	2.078
	(0.677)	(0.784)	(0.714)	(0.777)	(1.859)	(2.181)	(1.859)	(2.240)
Behind=1	6.082***	3.280***	5.710**	3.282***	8.694**	4.209**	7.875**	4.378**
	(2.092)	(1.097)	(2.054)	(1.077)	(3.459)	(1.963)	(3.421)	(1.845)
T2: Beliefs × Behind=1	-1.686	1.878	-1.289	1.682	-3.271	3.358	-1.246	3.057
	(2.798)	(1.951)	(2.520)	(1.900)	(4.863)	(3.955)	(4.238)	(3.905)
T3: Times × Behind=1	-1.094	2.373	-1.197	1.645	-0.902	5.832*	-1.021	4.670
	(2.357)	(1.556)	(2.246)	(1.866)	(4.868)	(2.904)	(4.721)	(3.243)
T4: Both × Behindi=1	-0.437	-0.069	0.123	-0.455	1.526	-0.692	3.565	-0.227
	(2.379)	(1.262)	(2.264)	(1.123)	(4.079)	(2.340)	(4.008)	(2.040)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Behind is a binary variable which takes the value 1 when the respondent is behind, and zero otherwise. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.14: Timings with 'Ahead' & 'Behind' (binary variables)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	2.267	-5.867**	3.706	-4.717**	8.322	-15.106**	10.920	-15.307**
	(3.475)	(2.289)	(3.082)	(2.166)	(9.923)	(6.626)	(8.866)	(6.974)
T3: Times	4.350	-3.950*	5.717	-1.939	12.057	-10.651*	16.371*	-10.625
	(3.241)	(1.999)	(3.393)	(2.697)	(8.470)	(5.395)	(9.252)	(7.610)
T4: Both	2.475	-3.200**	3.228*	-1.948	9.202	-8.037**	10.185*	-6.756
	(2.756)	(1.162)	(1.804)	(2.015)	(7.571)	(3.145)	(5.333)	(4.125)
Ahead=1	1.408	-3.684**	3.424	-2.223	4.951	-10.703**	9.329	-8.968**
	(2.263)	(1.577)	(2.205)	(1.376)	(6.517)	(3.796)	(6.213)	(3.182)
T2: Beliefs × Ahead=1	-1.714	6.351**	-3.817	5.175*	-6.256	16.454*	-10.936	16.231*
	(3.526)	(2.709)	(3.486)	(2.602)	(10.111)	(7.930)	(9.680)	(8.163)
T3: Times × Ahead=1	-4.491	3.731	-6.058*	1.624	-11.499	10.654*	-15.988	10.167
	(3.065)	(2.253)	(3.323)	(2.796)	(8.514)	(6.009)	(9.330)	(7.403)
T4: Both × Ahead=1	-3.141	4.481**	-4.526**	2.805	-10.993	12.762**	-13.908*	10.181*
	(2.894)	(1.675)	(2.135)	(2.521)	(8.556)	(4.778)	(6.829)	(5.014)
Behind=1	7.350*	0.093	8.799**	1.360	13.154	-5.048	16.297**	-3.385
	(3.582)	(1.821)	(3.226)	(1.620)	(8.306)	(3.828)	(7.307)	(2.869)
T2: Beliefs × Behind=1	-3.242	7.500**	-4.745	6.304*	-8.959	17.866**	-11.157	17.476**
	(4.864)	(3.464)	(4.387)	(3.308)	(12.159)	(7.313)	(10.734)	(7.676)
T3: Times × Behind=1	-5.284	5.603**	-6.808	3.010	-11.566	15.042**	-15.818	13.622
	(4.017)	(2.444)	(3.930)	(3.245)	(9.410)	(5.878)	(9.743)	(7.916)
T4: Both × Behind=1	-3.282	3.832	-3.963	1.986	-8.434	10.409**	-9.007	8.651
	(4.043)	(2.265)	(2.880)	(2.920)	(9.339)	(4.699)	(6.949)	(5.194)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a binary variable which takes the value of 1 when the respondent is ahead, and zero otherwise. Behind is a binary variable which takes the value 1 when the respondent is behind, and zero otherwise. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.15: Timings with 'Ahead & Positive/Negative' (binary variables)

	Time improvements			% Time improvements				
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.154	1.261	-0.500	1.447	1.311	2.236	0.558	2.179
	(1.069)	(1.081)	(1.173)	(0.960)	(1.799)	(1.950)	(1.610)	(1.628)
T3: Times	0.343	0.983	0.133	0.847	2.176	3.244	2.262	2.739
	(1.076)	(1.169)	(1.048)	(1.069)	(2.156)	(2.061)	(1.984)	(1.875)
T4: Both	-0.154	0.740	-0.549	0.304	1.226	2.960	0.279	2.347
	(1.060)	(1.108)	(1.099)	(0.951)	(1.603)	(2.054)	(1.729)	(1.648)
Ahead and Positive=1	-2.749	-0.430	-4.854	-0.875	-4.010	1.724	-7.407	1.348
	(3.491)	(2.450)	(3.392)	(2.387)	(7.635)	(5.868)	(6.900)	(5.824)
T2: Beliefs × Ahead and Positive=1	-2.608	-2.347	-1.484	-3.618	-4.665	-6.343	-3.158	-8.918
	(3.902)	(2.702)	(3.499)	(2.442)	(10.511)	(7.005)	(8.874)	(6.683)
T3: Times × Ahead and Positive=1	-0.485	-4.098	-1.478	-4.086	1.121	-12.196	-1.816	-12.763
	(3.406)	(3.120)	(3.044)	(3.025)	(8.095)	(8.439)	(7.270)	(8.116)
T4: Both × Ahead and Positive=1	-4.400	-0.312	-3.854	0.249	-16.699*	-0.876	-16.778*	-0.789
	(3.895)	(2.476)	(4.208)	(2.470)	(9.099)	(5.908)	(8.851)	(5.994)
Ahead and Negative=1	-2.589*	-0.892	-2.297**	-0.399	-4.204	-1.271	-3.276	3.201
	(1.242)	(4.129)	(1.039)	(4.474)	(2.940)	(10.052)	(2.296)	(10.463)
T2: Beliefs × Ahead and Negative=1	0.635	-4.261	-0.022	-3.924	-0.435	-8.408	-2.267	-11.160
	(2.058)	(4.800)	(1.770)	(5.350)	(3.987)	(11.924)	(3.451)	(12.397)
T3: Times × Ahead and Negative=1	-0.800	-4.316	-1.020	-5.488	-3.140	-12.380	-3.972	-18.484*
	(2.092)	(4.595)	(1.945)	(4.409)	(6.076)	(10.530)	(5.658)	(10.222)
T4: Both × Ahead and Negative=1	-0.043	-4.240	-0.322	-5.104	-1.677	-10.158	-2.355	-14.909
	(1.726)	(4.657)	(1.420)	(4.719)	(5.024)	(11.699)	(3.835)	(11.929)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead and Positive' is a binary variable which takes the value 1 when the respondent is ahead *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a binary variable which takes the value 1 when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.16: Timings with 'Behind & Positive/Negative' (binary variables)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.276	0.877	-0.036	0.695	1.158	0.909	0.260	0.176
	(0.804)	(1.210)	(0.759)	(1.137)	(1.710)	(2.182)	(1.591)	(2.207)
T3: Times	0.356	0.048	0.079	-0.306	1.351	0.756	1.205	-0.336
	(1.160)	(1.310)	(1.110)	(1.198)	(2.352)	(2.674)	(2.209)	(2.553)
T4: Both	-0.260	0.580	-0.539	0.023	-0.407	2.544	-1.275	1.490
	(0.971)	(1.102)	(1.004)	(1.021)	(2.107)	(2.151)	(2.135)	(1.959)
Behind and Positive=1	0.305	2.036	-0.809	0.291	-0.597	1.536	-1.438	-1.037
	(1.745)	(2.196)	(2.592)	(2.405)	(3.582)	(3.184)	(5.110)	(3.483)
T2: Beliefs × Behind and Positive=1	2.724	1.938	2.340	2.311	2.511	6.547*	0.612	6.355
	(2.728)	(1.732)	(3.425)	(2.056)	(6.102)	(3.620)	(5.473)	(4.307)
T3: Times × Behind and Positive=1	0.358	2.511	-0.196	2.761	1.248	5.625	-0.572	6.248
	(2.706)	(2.739)	(2.623)	(2.930)	(6.037)	(4.193)	(5.500)	(4.380)
T4: Both × Behind and Positive=1	5.315**	1.876	5.737*	1.968	11.058**	3.826	10.816**	4.225
	(2.248)	(2.457)	(2.743)	(2.315)	(4.473)	(4.161)	(4.804)	(4.034)
Behind and Negative=1	7.147	6.714***	6.899	5.090**	6.534	15.646***	6.156	11.664**
	(4.333)	(1.909)	(4.205)	(2.219)	(4.509)	(3.620)	(4.091)	(4.451)
T2: Beliefs × Behind and Negative=1	-2.318	-5.306	-2.782	-2.331	2.398	-11.203	2.646	-4.906
	(4.697)	(3.663)	(4.678)	(3.944)	(5.464)	(6.780)	(5.620)	(7.244)
T3: Times × Behind and Negative=1	-3.065	-12.048***	-2.474	-9.505**	0.421	-26.887**	1.539	-22.451**
	(4.476)	(3.810)	(4.146)	(3.397)	(6.264)	(9.980)	(5.794)	(8.817)
T4: Both × Behind and Negative=1	-5.171	-5.580	-4.507	-4.245	-3.476	-12.675	-1.311	-9.109
	(4.910)	(3.240)	(5.024)	(3.522)	(5.245)	(7.440)	(5.521)	(7.328)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind and Positive' is a binary variable which takes the value 1 when the respondent is behind *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a binary variable which takes the value 1 when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.17: Timings with 'Ahead/Behind & Positive/Negative' (binary variables)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.180	1.278	-0.085	1.215	1.107	1.786	0.483	1.433
	(0.904)	(1.208)	(0.918)	(1.010)	(1.961)	(2.049)	(1.749)	(1.779)
T3: Times	0.533	1.052	0.329	0.820	1.631	3.537	1.830	2.865
	(1.507)	(1.173)	(1.297)	(1.051)	(3.206)	(2.190)	(2.681)	(1.935)
T4: Both	0.012	0.822	-0.341	0.205	0.710	3.039	-0.246	2.081
	(1.029)	(1.147)	(1.004)	(1.028)	(1.894)	(2.141)	(1.927)	(1.771)
Ahead and Positive=1	-1.731	0.183	-3.865	-0.734	-3.224	2.597	-6.697	1.397
	(2.962)	(2.565)	(2.759)	(2.530)	(7.380)	(6.104)	(6.542)	(6.099)
T2: Beliefs × Ahead and Positive=1	-2.942	-2.365	-2.089	-3.271	-4.461	-5.893	-3.622	-7.934
	(3.457)	(2.725)	(2.935)	(2.538)	(10.427)	(6.970)	(8.745)	(6.729)
T3: Times × Ahead and Positive=1	-0.676	-4.167	-1.667	-4.005	1.666	-12.489	-1.525	-12.750
	(3.143)	(3.115)	(2.822)	(3.015)	(8.374)	(8.563)	(7.470)	(8.220)
T4: Both × Ahead and Positive=1	-4.566	-0.394	-3.986	0.378	-16.183*	-0.956	-16.159*	-0.425
	(3.453)	(2.690)	(3.537)	(2.698)	(8.875)	(6.261)	(8.330)	(6.367)
Ahead and Negative=1	-1.572	-0.278	-1.242	-0.235	-3.418	-0.398	-2.324	3.307
	(1.236)	(4.008)	(0.993)	(4.504)	(3.184)	(9.854)	(2.333)	(10.464)
T2: Beliefs × Ahead and Negative=1	0.301	-4.278	-0.446	-3.851	-0.231	-7.958	-2.257	-10.622
	(1.780)	(4.670)	(1.500)	(5.340)	(4.132)	(11.601)	(3.480)	(12.337)
T3: Times × Ahead and Negative=1	-0.991	-4.385	-1.208	-5.440	-2.595	-12.674	-3.550	-18.494*
	(2.570)	(4.487)	(2.195)	(4.541)	(6.705)	(10.460)	(5.824)	(10.433)
T4: Both × Ahead and Negative=1	-0.209	-4.322	-0.577	-4.940	-1.161	-10.237	-1.893	-14.482
	(1.824)	(4.633)	(1.357)	(4.787)	(5.341)	(11.741)	(3.914)	(12.132)
Behind and Positive=1	-0.160	2.043	-1.924	0.060	-1.581	1.775	-3.562	-1.076
	(1.761)	(2.237)	(2.645)	(2.522)	(3.548)	(3.250)	(5.198)	(3.646)
T2: Beliefs × Behind and Positive=1	2.820	1.537	2.481	1.804	2.561	5.670	0.511	5.107
	(2.780)	(1.637)	(3.431)	(1.963)	(6.339)	(3.321)	(5.606)	(4.011)
T3: Times × Behind and Positive=1	0.181	1.507	-0.436	1.626	0.968	2.843	-1.104	3.041
	(2.814)	(2.634)	(2.648)	(2.832)	(6.248)	(3.860)	(5.622)	(3.915)
T4: Both × Behind and Positive=1	5.044*	1.635	5.575*	1.816	9.941**	3.331	9.743*	3.732
	(2.426)	(2.623)	(2.882)	(2.553)	(4.683)	(4.441)	(5.047)	(4.482)
Behind and Negative=1	6.682	6.722***	6.378	4.920**	5.549	15.884***	5.248	11.657**
	(4.329)	(1.930)	(4.226)	(2.228)	(4.681)	(3.513)	(4.069)	(4.285)
T2: Beliefs × Behind and Negative=1	-2.222	-5.707	-2.745	-2.790	2.448	-12.080*	2.345	-5.985
	(4.543)	(3.660)	(4.614)	(4.027)	(5.539)	(6.626)	(5.657)	(7.172)
T3: Times × Behind and Negative=1	-3.242	-13.052***	-2.821	-10.643***	0.140	-29.669***	0.606	-25.678***
	(4.674)	(3.895)	(4.206)	(3.441)	(6.901)	(10.098)	(5.847)	(8.774)
T4: Both × Behind and Negative=1	-5.442	-5.822*	-4.827	-4.367	-4.593	-13.169	-2.707	-9.566
	(4.824)	(3.353)	(4.980)	(3.678)	(4.939)	(7.600)	(5.074)	(7.621)
Enumerator FE & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead and Positive' is a binary variable which takes the value 1 when the respondent is ahead *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a binary variable which takes the value 1 when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. 'Behind and Positive' is a binary variable which takes the value 1 when the respondent is behind *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a binary variable which takes the value 1 when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Wife and husband regressions have 640 and 647 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.6. Regressions with bootstrapped standard errors

In the main analysis, we account for within-village correlation by clustering standard errors at the village level. However, with only 19 village clusters, conventional cluster-robust standard errors may underestimate true sampling variability and yield overly optimistic p-values. To address this, we bootstrap standard errors clustered at the village level and report the same regression results below using these more reliable estimates as recommended by Cameron, Gelbach and Miller (2008).

Table G.18: Timings when ahead (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.020	0.965	1.020	0.957	2.638	1.959	2.431	1.401
	(0.908)	(1.042)	(0.927)	(0.775)	(1.986)	(2.461)	(1.974)	(1.842)
T3: Times	1.143	1.446	1.029	1.266	3.008	3.648	2.995	2.739
	(1.170)	(1.038)	(1.135)	(0.898)	(2.541)	(2.372)	(2.474)	(2.018)
T4: Both	-0.043	1.920**	-0.002	1.774***	0.324	4.458***	0.706	4.018***
	(1.139)	(0.754)	(1.095)	(0.536)	(2.427)	(1.570)	(2.352)	(1.166)
Ahead	-0.148**	-0.049	-0.135***	-0.049	-0.226*	-0.069	-0.178	-0.067
	(0.058)	(0.104)	(0.051)	(0.059)	(0.131)	(0.248)	(0.129)	(0.138)
T2: Beliefs × Ahead	-0.120*	-0.042	-0.138**	-0.025	-0.247	-0.111	-0.316**	-0.047
	(0.066)	(0.145)	(0.066)	(0.108)	(0.157)	(0.366)	(0.160)	(0.259)
T3: Times × Ahead	-0.112	-0.152	-0.096	-0.141	-0.244	-0.401	-0.206	-0.348*
	(0.121)	(0.136)	(0.105)	(0.091)	(0.302)	(0.335)	(0.269)	(0.198)
T4: Both × Ahead	0.019	-0.174*	-0.039	-0.212***	-0.010	-0.349	-0.200	-0.408*
	(0.088)	(0.102)	(0.083)	(0.070)	(0.237)	(0.284)	(0.236)	(0.218)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.19: Timings when behind (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.126	0.550	-0.146	0.416	1.521	1.091	0.606	0.449
	(0.716)	(1.026)	(0.624)	(0.937)	(1.687)	(2.435)	(1.646)	(2.293)
T3: Times	-0.142	-0.543	-0.251	-0.686	1.341	-0.545	1.223	-1.322
	(0.794)	(0.973)	(0.710)	(0.771)	(1.882)	(2.566)	(1.840)	(1.959)
T4: Both	-0.274	0.555	-0.747	0.417	-0.417	2.451	-1.595	2.142
	(0.739)	(0.892)	(0.715)	(0.522)	(1.863)	(2.259)	(1.871)	(1.522)
Behind	0.317***	0.248***	0.298***	0.249***	0.383**	0.310***	0.384**	0.312***
	(0.078)	(0.049)	(0.079)	(0.044)	(0.167)	(0.093)	(0.168)	(0.096)
T2: Beliefs × Behind	0.058	0.030	0.060	0.030	-0.056	0.052	-0.040	0.066
	(0.124)	(0.090)	(0.113)	(0.077)	(0.231)	(0.180)	(0.214)	(0.165)
T3: Times × Behind	0.089	0.242***	0.099	0.227***	0.029	0.403**	0.060	0.402***
	(0.106)	(0.076)	(0.102)	(0.066)	(0.213)	(0.158)	(0.209)	(0.135)
T4: Both × Behind	0.117	0.089	0.142	0.064	0.199	0.079	0.285	0.019
	(0.126)	(0.070)	(0.130)	(0.079)	(0.238)	(0.125)	(0.258)	(0.156)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Behind is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.20: Timings when ahead/behind (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.773	0.901	0.684	0.554	3.612	1.931	3.296	0.725
	(1.075)	(1.381)	(0.993)	(1.182)	(2.505)	(3.159)	(2.531)	(2.828)
T3: Times	0.279	-0.049	0.056	-0.359	3.124	1.696	2.703	0.382
	(1.283)	(1.178)	(1.149)	(0.964)	(3.018)	(2.921)	(2.832)	(2.519)
T4: Both	-0.473	1.331*	-0.401	1.410**	-0.471	4.360**	0.038	4.498***
	(1.060)	(0.744)	(0.990)	(0.585)	(2.636)	(1.817)	(2.554)	(1.558)
Ahead	-0.046	0.069	-0.034	0.062	-0.111	0.078	-0.050	0.074
	(0.049)	(0.102)	(0.047)	(0.064)	(0.124)	(0.260)	(0.128)	(0.163)
T2: Beliefs × Ahead	-0.090	-0.053	-0.109*	-0.020	-0.288*	-0.127	-0.352**	-0.033
	(0.067)	(0.135)	(0.066)	(0.098)	(0.170)	(0.364)	(0.179)	(0.259)
T3: Times × Ahead	-0.059	-0.074	-0.038	-0.051	-0.244	-0.299	-0.185	-0.220
	(0.105)	(0.129)	(0.089)	(0.077)	(0.297)	(0.340)	(0.257)	(0.203)
T4: Both × Ahead	0.026	-0.126	-0.038	-0.168**	0.012	-0.329	-0.189	-0.410*
	(0.076)	(0.105)	(0.077)	(0.085)	(0.231)	(0.295)	(0.236)	(0.240)
Behind	0.296***	0.272***	0.283***	0.270***	0.334**	0.336***	0.363**	0.336**
	(0.080)	(0.053)	(0.081)	(0.054)	(0.168)	(0.110)	(0.171)	(0.131)
T2: Beliefs × Behind	0.030	0.012	0.024	0.023	-0.151	0.008	-0.161	0.053
	(0.138)	(0.092)	(0.125)	(0.085)	(0.261)	(0.171)	(0.237)	(0.183)
T3: Times × Behind	0.067	0.217**	0.082	0.210***	-0.064	0.282*	-0.020	0.307*
	(0.111)	(0.086)	(0.107)	(0.073)	(0.204)	(0.166)	(0.200)	(0.163)
T4: Both × Behind	0.127	0.051	0.119	0.017	0.198	-0.010	0.180	-0.092
	(0.133)	(0.080)	(0.137)	(0.095)	(0.248)	(0.138)	(0.265)	(0.186)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Behind is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, attitude to risk, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.21: Timings with spousal evaluations (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.126	0.550	-0.146	0.416	1.521	1.091	0.606	0.449
	(0.716)	(1.026)	(0.624)	(0.937)	(1.687)	(2.435)	(1.646)	(2.293)
T3: Times	-0.142	-0.543	-0.251	-0.686	1.341	-0.545	1.223	-1.322
	(0.794)	(0.973)	(0.710)	(0.771)	(1.882)	(2.566)	(1.840)	(1.959)
T4: Both	-0.274	0.555	-0.747	0.417	-0.417	2.451	-1.595	2.142
	(0.739)	(0.892)	(0.715)	(0.522)	(1.863)	(2.259)	(1.871)	(1.522)
Behind	0.317***	0.248***	0.298***	0.249***	0.383**	0.310***	0.384**	0.312***
	(0.078)	(0.049)	(0.079)	(0.044)	(0.167)	(0.093)	(0.168)	(0.096)
T2: Beliefs × Behind	0.058	0.030	0.060	0.030	-0.056	0.052	-0.040	0.066
	(0.124)	(0.090)	(0.113)	(0.077)	(0.231)	(0.180)	(0.214)	(0.165)
T3: Times × Behind	0.089	0.242***	0.099	0.227***	0.029	0.403**	0.060	0.402***
	(0.106)	(0.076)	(0.102)	(0.066)	(0.213)	(0.158)	(0.209)	(0.135)
T4: Both × Behind	0.117	0.089	0.142	0.064	0.199	0.079	0.285	0.019
	(0.126)	(0.070)	(0.130)	(0.079)	(0.238)	(0.125)	(0.258)	(0.156)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are absolute timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(2) and (5)-(6)), and % timing improvements between rounds (Models (3)-(4) and (7)-(8)). Relative confidence is a dummy variable for positive evaluations in Models (1)-(4), and a dummy for negative evaluations in Models (5)-(8). The positive evaluation dummy takes the value 1 if the self-confidence dummy is 1 (respondent believes they are not top person) whilst the spousal evaluation dummy is 0 (spouse believes the respondent is the top person), zero otherwise. The negative evaluation dummy takes the value 1 if the self-confidence dummy is 0 (respondent believes they are the top person) whilst the spousal evaluation dummy is 1 (spouse believes the respondent is not top person), zero otherwise. Enumerator fixed effects are included in all models, along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. We use the 'Ahead' and 'Behind' continuous variables as further controls. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.22: Timings with 'Ahead & Positive/Negative' (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.549	0.961	0.458	1.039	1.704	1.900	1.216	1.617
	(0.699)	(0.840)	(0.728)	(0.662)	(1.580)	(2.025)	(1.595)	(1.578)
T3: Times	0.859	0.974	0.745	0.878	2.383	2.518	2.487	1.921
	(0.850)	(0.859)	(0.810)	(0.704)	(1.819)	(2.046)	(1.749)	(1.631)
T4: Both	0.143	1.697**	-0.272	1.387**	0.650	4.019**	-0.137	3.292**
	(0.824)	(0.821)	(0.863)	(0.612)	(1.625)	(1.800)	(1.715)	(1.359)
Ahead and Positive	0.021	0.148	-0.050	0.109	0.168	0.358	0.060	0.271
	(0.154)	(0.213)	(0.221)	(0.087)	(0.349)	(0.491)	(0.536)	(0.209)
T2: Beliefs × Ahead and Positive	-0.352	-0.087	-0.394	-0.113	-0.363	-0.220	-0.369	-0.222
	(0.621)	(0.246)	(0.656)	(0.163)	(1.815)	(0.651)	(1.871)	(0.431)
T3: Times × Ahead and Positive	-0.255	-0.256	-0.388	-0.242**	-0.320	-0.651	-0.609	-0.572*
	(0.195)	(0.249)	(0.396)	(0.117)	(0.469)	(0.626)	(0.718)	(0.302)
T4: Both × Ahead and Positive	-0.526	-0.271	-0.598	-0.228	-1.687*	-0.508	-1.899*	-0.391
	(0.361)	(0.221)	(0.392)	(0.171)	(0.978)	(0.544)	(1.087)	(0.431)
Ahead and Negative	-0.113	-0.035	-0.112	0.027	-0.230	-0.084	-0.190	0.141
	(0.081)	(0.369)	(0.081)	(0.441)	(0.222)	(0.869)	(0.197)	(1.042)
T2: Beliefs × Ahead and Negative	-0.073	-0.098	-0.120	-0.108	-0.176	-0.221	-0.290	-0.331
	(0.088)	(0.383)	(0.095)	(0.441)	(0.247)	(0.910)	(0.268)	(1.039)
T3: Times × Ahead and Negative	-0.147	-0.062	-0.140	-0.163	-0.394	-0.294	-0.394	-0.634
	(0.169)	(0.989)	(0.156)	(1.342)	(0.455)	(2.683)	(0.429)	(4.358)
T4: Both × Ahead and Negative	0.051	-0.296	0.046	-0.422	0.166	-0.683	0.081	-0.968
	(0.139)	(0.395)	(0.117)	(0.459)	(0.427)	(0.948)	(0.332)	(1.096)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Ahead and Positive' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.23: Timings with 'Behind & Positive/Negative' (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	-0.023	0.645	-0.151	0.578	0.478	1.204	-0.089	0.658
	(0.718)	(0.789)	(0.714)	(0.701)	(1.581)	(1.984)	(1.569)	(1.715)
T3: Times	0.028	0.170	-0.057	-0.108	0.940	1.092	0.974	0.095
	(0.939)	(0.857)	(0.914)	(0.700)	(1.919)	(2.181)	(1.876)	(1.745)
T4: Both	-0.424	1.179	-0.630	0.902*	-0.633	3.121	-1.195	2.384*
	(0.870)	(0.866)	(0.892)	(0.523)	(1.899)	(1.985)	(1.910)	(1.260)
Behind and Positive	0.003	0.188*	-0.048	0.143	-0.135	0.218	-0.074	0.103
	(0.181)	(0.099)	(0.289)	(0.109)	(0.370)	(0.165)	(0.606)	(0.181)
T2: Beliefs × Behind and Positive	0.309	0.184	0.271	0.200*	0.494	0.368	0.242	0.394**
	(0.451)	(0.153)	(0.560)	(0.117)	(1.059)	(0.248)	(1.122)	(0.191)
T3: Times × Behind and Positive	0.265	0.327**	0.155	0.349***	0.553	0.424**	0.312	0.515***
	(0.354)	(0.128)	(0.418)	(0.112)	(0.767)	(0.195)	(0.905)	(0.187)
T4: Both × Behind and Positive	0.582**	0.162	0.590*	0.175*	0.951**	0.192	0.850	0.229
	(0.254)	(0.101)	(0.335)	(0.103)	(0.462)	(0.190)	(0.635)	(0.172)
Behind and Negative	0.231**	0.436	0.257**	0.416***	0.226	0.858	0.341	0.752***
	(0.110)	(0.504)	(0.108)	(0.112)	(0.207)	(1.787)	(0.213)	(0.244)
T2: Beliefs × Behind and Negative	0.341	-0.417	0.225	-0.400**	0.685	-0.875	0.566	-0.708**
	(0.240)	(0.576)	(0.289)	(0.164)	(0.433)	(1.886)	(0.520)	(0.300)
T3: Times × Behind and Negative	0.160	-1.234	0.141	-1.059	0.254	-2.830	0.232	-2.362
	(0.131)	(1.196)	(0.138)	(1.267)	(0.241)	(3.474)	(0.250)	(3.827)
T4: Both × Behind and Negative	0.040	-0.359	-0.054	-0.374	0.147	-0.677	0.019	-0.602
	(0.330)	(0.553)	(0.407)	(0.262)	(0.667)	(1.876)	(0.866)	(0.613)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind and Positive' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.24: Timings with 'Ahead/Behind & Positive/Negative' (bootstrapped SEs)

	Time improvements				% Time improvements			
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.181	0.831	0.208	0.780	0.942	1.657	0.773	1.130
	(0.775)	(0.871)	(0.821)	(0.707)	(1.684)	(2.096)	(1.726)	(1.657)
T3: Times	0.184	0.559	0.149	0.314	1.432	2.317	1.678	1.373
	(1.117)	(0.763)	(1.051)	(0.638)	(2.288)	(1.930)	(2.136)	(1.575)
T4: Both	-0.357	1.650**	-0.571	1.354***	-0.204	4.117**	-0.559	3.321***
	(0.928)	(0.767)	(0.936)	(0.515)	(1.753)	(1.755)	(1.817)	(1.227)
Ahead and Positive	0.031	0.174	-0.037	0.129	0.175	0.396	0.075	0.295
	(0.157)	(0.209)	(0.218)	(0.086)	(0.350)	(0.489)	(0.531)	(0.212)
T2: Beliefs × Ahead and Positive	-0.305	-0.083	-0.355	-0.081	-0.290	-0.211	-0.317	-0.177
	(0.579)	(0.232)	(0.612)	(0.150)	(1.748)	(0.628)	(1.810)	(0.412)
T3: Times × Ahead and Positive	-0.180	-0.236	-0.296	-0.202*	-0.227	-0.641	-0.496	-0.529*
	(0.186)	(0.240)	(0.355)	(0.115)	(0.519)	(0.615)	(0.712)	(0.300)
T4: Both × Ahead and Positive	-0.467	-0.263	-0.515	-0.206	-1.604*	-0.504	-1.796	-0.369
	(0.355)	(0.228)	(0.386)	(0.181)	(0.971)	(0.558)	(1.095)	(0.432)
Ahead and Negative	-0.095	-0.007	-0.085	0.054	-0.217	-0.043	-0.152	0.176
	(0.083)	(0.348)	(0.074)	(0.360)	(0.225)	(0.807)	(0.188)	(0.846)
T2: Beliefs × Ahead and Negative	-0.058	-0.106	-0.117	-0.129	-0.143	-0.231	-0.280	-0.357
	(0.088)	(0.368)	(0.096)	(0.387)	(0.240)	(0.867)	(0.263)	(0.906)
T3: Times × Ahead and Negative	-0.093	-0.047	-0.093	-0.136	-0.322	-0.290	-0.330	-0.610
	(0.184)	(0.433)	(0.163)	(0.412)	(0.470)	(0.997)	(0.432)	(0.962)
T4: Both × Ahead and Negative	0.080	-0.285	0.055	-0.398	0.214	-0.676	0.094	-0.945
	(0.144)	(0.370)	(0.117)	(0.379)	(0.427)	(0.880)	(0.328)	(0.905)
Behind and Positive	-0.013	0.199**	-0.084	0.151	-0.167	0.241	-0.125	0.123
	(0.164)	(0.100)	(0.282)	(0.107)	(0.334)	(0.174)	(0.593)	(0.181)
T2: Beliefs × Behind and Positive	0.301	0.175	0.251	0.191	0.475	0.347	0.177	0.371*
	(0.453)	(0.156)	(0.577)	(0.118)	(1.055)	(0.255)	(1.142)	(0.190)
T3: Times × Behind and Positive	0.237	0.309**	0.090	0.330***	0.479	0.369*	0.171	0.457**
	(0.351)	(0.128)	(0.416)	(0.108)	(0.756)	(0.197)	(0.907)	(0.179)
T4: Both × Behind and Positive	0.577**	0.142	0.576*	0.154	0.920**	0.149	0.773	0.184
	(0.249)	(0.102)	(0.336)	(0.107)	(0.441)	(0.188)	(0.641)	(0.173)
Behind and Negative	0.219**	0.449	0.241**	0.424***	0.202	0.886	0.321	0.768***
	(0.112)	(0.529)	(0.105)	(0.112)	(0.207)	(1.847)	(0.200)	(0.246)
T2: Beliefs × Behind and Negative	0.314	-0.428	0.182	-0.414**	0.626	-0.901	0.467	-0.735**
	(0.240)	(0.601)	(0.290)	(0.165)	(0.429)	(1.947)	(0.514)	(0.301)
T3: Times × Behind and Negative	0.151	-1.261	0.124	-1.091	0.226	-2.918	0.171	-2.466
	(0.144)	(1.221)	(0.144)	(1.188)	(0.255)	(3.557)	(0.249)	(3.689)
T4: Both × Behind and Negative	0.024	-0.391	-0.084	-0.406	0.086	-0.743	-0.103	-0.666
	(0.332)	(0.576)	(0.405)	(0.265)	(0.653)	(1.933)	(0.842)	(0.619)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regressions. Wife and husband regressions have 631 and 638 observations, respectively. Dependent variables are timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). 'Behind and Positive' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Behind and Negative' is a continuous variable which takes the value of the time difference between respondent and spouse, when the respondent is behind *and* the respondent also has a negative evaluation from their spouse, zero otherwise. 'Ahead and Positive' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values) *and* when the respondent also has a positive evaluation from their spouse, zero otherwise. 'Ahead and Negative' is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead *and* the respondent also has a negative evaluation from their spouse, zero otherwise. For definitions of positive and negative evaluations, please refer to section 2.4. Enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Bootstrapped standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.7. Regressions with SUREGs

Table G.25: Treatment effects on timing improvements when behind, using SUREG

	Time improvements				% Time improvements			
	(1) est1		(2) est2		(3) est3		(4) est4	
	Wives	Husbands	Wives	Husbands	Wives	Husbands	Wives	Husbands
T2: Beliefs	0.089	0.375	-0.240	-0.076	1.476	0.718	0.386	-0.671
	(0.676)	(1.011)	(0.564)	(1.088)	(1.645)	(2.353)	(1.383)	(2.543)
T3: Timings	-0.197	-0.592	-0.237	-0.899	1.207	-0.676	1.221	-1.598
	(0.837)	(0.956)	(0.780)	(0.923)	(1.942)	(2.528)	(1.867)	(2.432)
T4: Both	-0.375	0.592	-0.987	0.146	-0.600	2.459	-2.152	1.556
	(0.755)	(0.886)	(0.702)	(0.668)	(1.893)	(2.236)	(1.813)	(1.830)
Behind	0.303***	0.256***	0.296***	0.232***	0.375**	0.319***	0.393**	0.245**
	(0.084)	(0.043)	(0.085)	(0.047)	(0.173)	(0.086)	(0.178)	(0.104)
T2: Beliefs × Behind	0.066	0.046	0.067	0.102	-0.046	0.086	-0.043	0.221
	(0.126)	(0.095)	(0.115)	(0.094)	(0.230)	(0.187)	(0.209)	(0.185)
T3: Timings × Behind	0.101	0.257***	0.094	0.261***	0.054	0.430***	0.046	0.447***
	(0.113)	(0.065)	(0.116)	(0.058)	(0.223)	(0.149)	(0.237)	(0.133)
T4: Both × Behind	0.120	0.085	0.124	0.084	0.219	0.080	0.250	0.073
	(0.116)	(0.064)	(0.112)	(0.066)	(0.236)	(0.113)	(0.241)	(0.136)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Village and enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.26: Treatment effects on timing improvements when ahead, using SUREG

	Time improvements				% Time improvements			
	(1)		(2)		(3)		(4)	
	Wives	Husbands	Wives	Husbands	Wives	Husbands	Wives	Husbands
T2: Beliefs	1.025	0.771	1.213	1.085	2.648	1.530	2.457	1.550
	(0.938)	(1.076)	(0.986)	(0.960)	(2.019)	(2.519)	(1.872)	(2.243)
T3: Timings	1.326	1.564	1.511	1.573	3.131	3.604	3.431	3.054
	(1.147)	(1.011)	(1.110)	(1.036)	(2.605)	(2.346)	(2.528)	(2.542)
T4: Both	-0.084	1.976***	-0.208	1.723***	0.171	4.526***	0.039	3.980***
	(1.118)	(0.748)	(0.988)	(0.628)	(2.370)	(1.555)	(2.110)	(1.361)
Ahead	-0.198***	-0.106	-0.172***	-0.085	-0.258**	-0.102	-0.188	-0.026
	(0.054)	(0.110)	(0.048)	(0.100)	(0.128)	(0.254)	(0.126)	(0.232)
T2: Beliefs × Ahead	-0.122*	-0.040	-0.156**	-0.055	-0.250	-0.083	-0.326**	-0.109
	(0.068)	(0.149)	(0.064)	(0.120)	(0.155)	(0.366)	(0.146)	(0.286)
T3: Timings × Ahead	-0.153	-0.157	-0.158*	-0.165	-0.272	-0.384	-0.259	-0.381
	(0.114)	(0.138)	(0.092)	(0.128)	(0.306)	(0.336)	(0.267)	(0.310)
T4: Both × Ahead	0.013	-0.196*	-0.048	-0.222**	-0.003	-0.369	-0.197	-0.435
	(0.076)	(0.105)	(0.073)	(0.109)	(0.216)	(0.280)	(0.217)	(0.281)
Enumerator FEs & controls			✓	✓			✓	✓

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Village and enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.27: Treatment effects on timing improvements with spousal evaluations, using SUREG

	Positive evaluations				Negative evaluations			
	(1) Time improvements		(2) % Time improvements		(3)Time improvements		(4) % Time improvements	
	Wives	Husbands	Wives	Husbands	Wives	Husbands	Wives	Husbands
T2: Beliefs	0.152	0.206	1.388	0.664	-0.484	0.991	-0.819	1.477
	(0.666)	(0.957)	(1.603)	(2.336)	(0.711)	(0.712)	(1.633)	(1.798)
T3: Timings	0.285	0.565	1.312	2.512	0.013	1.057	1.159	2.155
	(0.615)	(1.049)	(1.507)	(2.516)	(1.035)	(0.773)	(2.302)	(1.986)
T4: Both	-0.405	0.239	0.299	2.291	-0.552	1.136	-1.228	3.093*
	(0.667)	(0.768)	(1.738)	(2.324)	(0.734)	(0.787)	(1.843)	(1.858)
Spousal evaluation	-1.507	-0.847	-1.757	0.432	-0.211	4.269**	-0.663	10.806**
	(1.778)	(1.490)	(5.195)	(3.671)	(1.166)	(2.086)	(2.691)	(4.623)
T2: Beliefs × Spousal evaluation	-0.996	1.400	-2.402	2.441	2.153	-6.274**	4.141	-11.340*
	(2.502)	(1.396)	(6.628)	(3.620)	(1.375)	(3.030)	(2.896)	(6.061)
T3: Timings × Spousal evaluation	-0.591	0.175	1.322	-2.314	0.605	-8.099***	0.958	-19.301***
	(2.043)	(1.643)	(5.719)	(4.892)	(1.680)	(2.911)	(3.775)	(6.729)
T4: Both × Spousal evaluation	0.419	1.643	1.262	2.582	0.482	-6.660***	0.506	-15.820***
	(2.367)	(1.610)	(6.634)	(4.015)	(1.217)	(2.488)	(3.191)	(5.912)
Enumerator FEs & controls	✓	✓	✓	✓	✓	✓	✓	✓
N	628		628		628		628	

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1)-(4)), and % timing improvements between rounds (Models (5)-(8)). Ahead is a continuous variable which takes the value of the time difference between spouse and respondent, when the respondent is ahead (thus they are all positive values), and zero otherwise. Wife and husband regressions have 631 and 638 observations, respectively. Village and enumerator fixed effects are included in Models (3), (4), (7), and (8), along with control variables capturing the respondent's age education, occupation, monthly income in a good month, monthly income in a bad month, number of sons and daughters, and number of married years with spouse. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.8. Testing for order effects and spillover effects

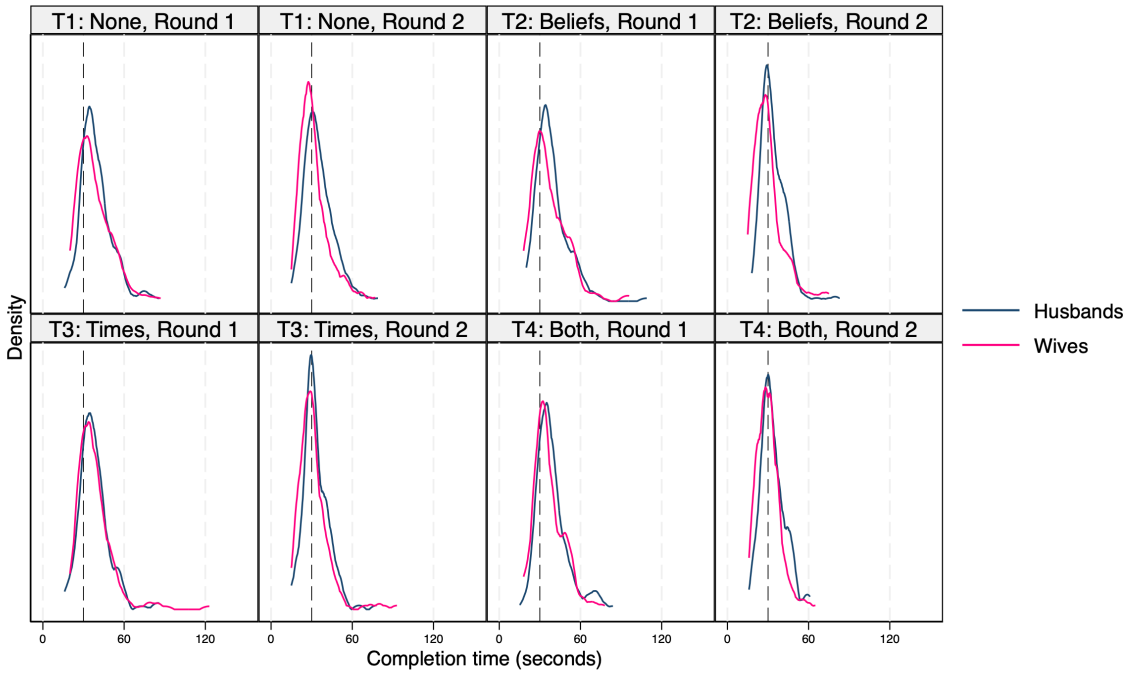
Table G.28: Order effects and spillover effects of two experiments from chapters 2 & 3

	Order effects				Spillover effects			
	Time improvements		% Time improvements		Time improvements		% Time improvements	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
CH3 T2: Beliefs	1.461	0.506	4.193	0.620				
	(1.320)	(1.408)	(2.803)	(3.825)				
CH3 T3: Times	0.910	1.353	3.590	2.678				
	(1.092)	(1.392)	(2.314)	(3.538)				
CH3 T4: Both	-0.900	0.922	-0.939	1.863				
	(1.334)	(1.625)	(3.023)	(3.781)				
CH3 second	-0.450	-0.432	0.107	-1.087				
	(1.166)	(1.650)	(3.037)	(3.855)				
CH3 T2: Beliefs × CH3 second	-2.202	0.568	-5.848	1.731				
	(1.775)	(2.269)	(4.228)	(5.539)				
CH3 T3: Timings × CH3 second	-0.478	-1.555	-3.312	-2.563				
	(1.030)	(1.768)	(3.080)	(4.537)				
CH3 T4: Both × CH3 second	1.850	0.695	2.239	2.530				
	(1.537)	(2.171)	(4.232)	(4.998)				
CH2 T2: Soap					0.337	0.790	1.215	1.486
					(0.729)	(0.790)	(1.572)	(1.968)
CH2 T3: Books					0.654	0.557	1.870	1.362
					(0.823)	(0.944)	(1.649)	(2.295)
CH2 T4: Rice					0.022	1.270	0.172	3.132
					(0.844)	(0.927)	(1.808)	(2.054)
N	631	638	631	638	631	638	631	638

Notes: OLS regression with dependent variables: timing improvements between rounds (calculated as r1 minus r2) (Models (1), (2), (5), (6)), and % timing improvements between rounds (Models (3), (4), (7), (8)). *CH3 second* is a binary variable that takes the value one if the experiment from Chapter 3 was conducted after the experiment from Chapter 2, and zero otherwise. Models (1)-(4) interact *CH3 second* with the treatment indicators from Chapter 3, whereas models (5)-(8) use treatment indicators for treatments from Chapter 2. Wife and husband regressions have 631 and 638 observations, respectively. No controls or fixed effects are used. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.9. Testing for bunching at 30 seconds

Figure G.1: Distribution of task completion times by round, treatments, and gender



Notes: Kernel density estimates of completion times for husbands (navy) and wives (pink). The vertical dashed line represents the 30-second maximum payoff threshold.

Table G.29: "Under 30 seconds" binary dependent variable

	Var = Ahead		Var = Behind		Var = Poseval		Var = Negeval	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	0.023	0.000	-0.008	0.005	-0.001	-0.005	-0.005	0.016
	(0.040)	(0.064)	(0.031)	(0.061)	(0.026)	(0.048)	(0.044)	(0.058)
T3: Times	0.029	0.034	0.026	-0.036	0.007	0.002	0.011	0.019
	(0.052)	(0.048)	(0.052)	(0.052)	(0.037)	(0.044)	(0.045)	(0.047)
T4: Both	0.009	-0.011	0.022	-0.001	-0.004	-0.016	0.018	0.013
	(0.036)	(0.061)	(0.041)	(0.053)	(0.036)	(0.051)	(0.039)	(0.055)
Var	0.002	0.001	-0.013**	-0.013***	-0.109	-0.107	-0.038	0.021
	(0.003)	(0.005)	(0.004)	(0.002)	(0.113)	(0.074)	(0.077)	(0.137)
T2 × Var	-0.004	0.002	0.002	0.002	-0.013	0.050	0.007	-0.107
	(0.005)	(0.007)	(0.005)	(0.003)	(0.178)	(0.078)	(0.085)	(0.182)
T3 × Var	-0.002	-0.005	-0.000	0.007**	0.124	0.040	0.015	-0.202
	(0.008)	(0.006)	(0.006)	(0.003)	(0.116)	(0.109)	(0.094)	(0.214)
T4 × Var	-0.001	0.002	-0.006	0.001	0.118	0.047	-0.065	-0.177
	(0.005)	(0.007)	(0.004)	(0.003)	(0.161)	(0.093)	(0.105)	(0.219)
Under 30 seconds (R1)	0.511***	0.629***	0.434***	0.565***	0.512***	0.623***	0.506***	0.628***
	(0.035)	(0.037)	(0.041)	(0.037)	(0.038)	(0.035)	(0.038)	(0.038)
N	631	638	631	638	631	638	631	638

Notes: OLS regression with binary dependent variable *Under 30 seconds*, which takes the value one if the respondent's timing in Round 2 was equal to or less than 30 seconds. Wife and husband regressions have 631 and 638 observations, respectively. No controls or fixed effects are used. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table G.30: "Over 30 seconds" restricted sample

	Var = Ahead		Var = Behind		Var = Poseval		Var = Negeval	
	(1) Wife	(2) Husband	(3) Wife	(4) Husband	(5) Wife	(6) Husband	(7) Wife	(8) Husband
T2: Beliefs	1.089	1.790	-0.068	1.397	0.598	0.923	-0.287	1.902**
	(1.355)	(1.071)	(1.182)	(1.183)	(1.247)	(1.032)	(1.352)	(0.773)
T3: Times	0.703	2.134*	-0.990	-0.106	0.307	0.907	-0.446	1.801**
	(1.413)	(1.052)	(1.241)	(1.186)	(1.281)	(1.205)	(2.009)	(0.798)
T4: Both	-0.463	2.163**	-0.928	0.377	-0.747	1.117	-0.121	2.060
	(1.360)	(0.988)	(1.239)	(1.170)	(1.373)	(1.314)	(1.212)	(1.246)
Var	-0.118	-0.014	0.204***	0.219***	-0.484	0.180	-1.478	3.000
	(0.078)	(0.126)	(0.069)	(0.055)	(1.551)	(1.757)	(1.846)	(2.672)
T2 × Var	-0.120	-0.056	0.093	0.008	0.402	2.582	2.449	-4.302
	(0.106)	(0.157)	(0.127)	(0.097)	(3.331)	(1.735)	(2.583)	(3.163)
T3 × Var	-0.112	-0.167	0.144	0.218**	-0.593	1.170	1.741	-8.134*
	(0.146)	(0.161)	(0.116)	(0.093)	(3.662)	(2.174)	(3.126)	(4.568)
T4 × Var	0.058	-0.235*	0.146	0.105	2.890	1.447	-0.773	-6.417*
	(0.113)	(0.117)	(0.125)	(0.075)	(2.635)	(2.166)	(1.706)	(3.485)
Mean in T1	8.46	4.72	8.46	4.79	8.43	4.72	8.50	4.74
p-value: T2=T4	0.105	0.765	0.250	0.411	0.124	0.879	0.878	0.902
p-value: T3=T4	0.163	0.981	0.953	0.575	0.243	0.835	0.818	0.788
N	400	498	400	498	400	498	400	498

Notes: OLS regression with binary dependent variable *Under 30 seconds*, which takes the value one if the respondent's timing in Round 2 was equal to or less than 30 seconds. Wife and husband regressions have 631 and 638 observations, respectively. No controls or fixed effects are used. Standard errors clustered at the village level between parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Conclusion

"Choices are the hinges of destiny"

– Edwin Markham

In my thesis, I contribute to the field of behavioural development economics a more nuanced understanding of the decision processes and behavioural motivations that drive the intrahousehold choices of children, parents, and spouses. This is of central interest to economists as intrahousehold choices have the power to perpetuate or eliminate gender inequalities, both within and across generations. Building on the significant contributions made by related scholarship in the field, I shed light on three specific dimensions of intrahousehold decision-making that have received little to no attention: the role of children's agency in parental decision-making, the dynamic process of spousal involvement in decision-making, and the influence of spousal comparisons and evaluations on spousal effort.

To provide specific insights into these three aspects of household behaviour, I use primary data from three field experiments and accompanying surveys, administered to 1,272 children and their parents, and 1,328 spouses, in coastal Zanzibar, Tanzania. Where rapid environmental change intersects with longstanding Islamic and patriarchal gender norms, Zanzibar presents an ideal setting for studying intrahousehold decision-making, one in which household choices carry substantial weight for the economic trajectories of the next generation.

In **Chapter 1**, I investigate the role of children's agency in parental decision-making, challenging the standard economic assumption that children are passive recipients of parental decisions made on their behalf. In many real-world settings, children are the first point of contact for opportunities related to their own human capital development (e.g., joining an extracurricular activity), and may hold explicit preferences about which parent to involve in a decision to pursue this. I posit that child preferences over parental involvement are not incidental, but can meaningfully shape parental decision outcomes, especially in the instance where children expect differences in outcomes of approaching one parent over the other.

To examine this, I designed and implemented two complementary experiments. The first is a Randomised Controlled Trial (RCT) with 1,272 children, where each child was instructed to approach either their mother or father for a signature on a consent form permitting them to enrol in swimming lessons. By exogenously assigning the parent that needed to be approached, the RCT captures the decision outcome under a scenario where the child's agency in choosing the parent is intentionally restricted. Here I find that girls are 38% less likely to secure paternal than maternal consent, but that there is no significant difference for boys. This is consistent with evidence provided by Dizon-Ross and Jayachandran (2023) that fathers have a lower willingness-to-pay (WTP) for their daughters' human capital than mothers.

In parallel, a choice experiment (CE) on a subsample of 848 children elicited child preferences over parental involvement. This was done by asking them to choose between the mother-form and the father-form, thereby restoring agency and allowing me to observe how children would choose if they were the decision initiators. Further, to investigate the determinants of children's choices, the CE experimentally varied the framing of swimming lessons as neutral, high-risk, inducing potential parental disagreement (high-conflict), or both high-risk and high-conflict. These variations reveal context-dependent preferences: for instance, girls are more likely to prefer involving mothers when parental conflict is made salient, whereas boys tend to shift toward involving fathers when risk or conflict is emphasised.

Linking the CE with the RCT allows me to examine child agency more closely. I find that RCT treatment effects are driven by girls who did *not* receive the form they expressed a preference for in the CE. This is the first indication that child agency, at the very least, matters for their own development outcomes. Moreover, making parental conflict salient in the CE reduces girls' success in obtaining paternal consent but increases it

when maternal consent is required, which also indicates that children's own perceptions and preferences can determine parental decision outcomes. It is worth noting that although I assume children do not have agency once they are given a form (regardless of whether the allocated form aligns with their preference), children may still have some agency in swaying the outcome, especially if the parent involved is not the parent the child wants, i.e., children can influence parents' decisions once the parent is chosen. This is an extension worth exploring in further research.

Taken together, my results show that children have systematic preferences over parental involvement, that this is responsive to anticipated parental disagreement and relative risk attitudes of parents, and that child agency in parental involvement shapes the outcome of parental decisions. Moreover, by incorporating children as actors who can influence the outcome, I extend the conceptual framework of studies linking spouses' involvement in decision processes to key child development outcomes (Ashraf, 2009; D'Exelle and Ringdal, 2022, 2025; D'Exelle et al., 2023; McCarthy, 2019).

Having looked at parental involvement from the child's perspective, in **Chapter 2**, I switch the lens to the spouse's perspective of spousal involvement in decision-making. Specifically, I aim to answer three research questions: (1) how much autonomy do spouses retain in household decisions? (2) what influences whether individuals retain or relinquish autonomy? and (3) does autonomy affect the final uptake of the services offered? To answer these questions, I draw on a field experiment with 664 spouses in rural Zanzibar that used the following experimental approach to *endogenise* spousal involvement. In private, I offered each spouse in a married or co-habiting couple a choice between two goods, where one option was always swimming lessons and the other a consumption good (Decision Block 1), then a choice to make the decision themselves or defer decision-making to their spouse (Decision Block 2), and finally a choice to consult the decision with their spouse where, if consultation was chosen, the chance to switch to the spouse's choice was offered (Decision Block 3). I also used a 2×2 factorial treatment design which varied (i) whether swimming lessons were for the wife or the child and hence whether the decision falls in the decision-making domain of the wife or a joint domain that explicitly required the involvement of the husband, and (ii) whether the paired consumption good benefitted an individual or the household.

Endogenising spousal involvement in this way serves a triple purpose. For one, it yields a better approximation of real-life decision-making. Many of the experiments that compare individual with joint decision-making try to exclude post-experimental influences by exogenously varying spousal involvement (Ashraf, 2009; D'Exelle and Ringdal, 2022, 2025; D'Exelle et al., 2023; McCarthy, 2019). In reality, couples play a repeated game and their anticipation of their spouses' actions can influence their real-life decisions. For another, the decision blocks allow for a close examination of autonomy as a process rather than an outcome, which highlights where in the decision process autonomy might be retained, relinquished or reconsidered. Moreover, it yields a more realistic estimate of whether which spouse moves first i.e., which spouse is approached with an investment opportunity or a cash transfer, really *matters* for uptake of the service or allocative efficiency of the intervention. By mapping the level of autonomy in each decision block onto the uptake of swimming lessons, I am able to shed light on whether autonomy matters for uptake and whether spouses really do retain autonomy from the first reflection through to the final decision.

Across the three decision blocks, the study yields three sets of findings. First, in eliciting individual preferences (Decision Block 1), clear gender differences emerge: husbands are consistently more likely than wives to choose swimming lessons—whether for the wife or the child—and both spouses shift their preferences upward when the beneficiary is the child rather than the wife. These patterns appear to be moderated by expectations about the other spouse's preferences, as individuals who believe their spouse prefers swimming lessons are significantly more likely to choose it themselves, and correlations between own choices and beliefs about what one's spouse thinks they would choose suggest that spouses take into account anticipated second-order beliefs. Second, in examining spousal involvement (Decision Blocks 2 and 3), I find that wives are more

likely to defer to, or consult with, their husbands when decisions concern the child rather than themselves, whereas husbands show the opposite pattern, deferring and consulting more when benefits accrue to the wife individually. This behaviour cannot be explained by a desire to avoid conflict and reconcile expected spousal preference differences. Instead, domain-based norms appear to be the mechanism through which treatment variation shapes spousal involvement. A related result is that spouses systematically underestimate the extent of their actual preference differences, implying that decisions are often made under incorrect assumptions about alignment. Third, regarding uptake, I find that who moves first appears to matter when looking only at initial choices in Decision Block 1, with higher uptake among husbands than wives across treatments. However, once the spousal involvement process is endogenised through deferral, consultation, and switching, the gender gap in final outcomes is substantially narrowed. In short, while autonomy shapes early decisions, the intrahousehold decision process itself ultimately narrows these differences, bringing spouses' final choices closer together. In effect, what *appears* as an autonomous decision may be less autonomous in practice through an endogenous spousal involvement process. However, I note an important nuance to consider: as Hillesland et al. (2025) argue, there can also be autonomy in choosing *not* to decide. Thus, although my study finds that decision outcomes are better for children under husbands when spouses make decisions alone, the high deferral rates of wives that bring these differences in outcomes closer together demonstrate that wives choosing not to decide leads to better outcomes for their children. Moreover, wives' beliefs around husbands' choices indicate that this outcome may already be anticipated by women from the onset of decision-making.

In **Chapter 3**, I examine the role of confidence and competition in influencing spousal effort. The motivation for this chapter lies in the research gap in the intrahousehold literature around what drives effort in the setting of a spousal relationship, the most gendered of partnerships, and seminal work in the economics literature which finds significant gender differences in confidence and competition as effort motivators (Buser et al., 2020; Gneezy, Niederle and Rustichini, 2003; Möbius et al., 2022; Niederle and Vesterlund, 2007, 2010, 2011; Sarsons and Xu, 2021).

I posit that, given the close proximity which characterises the spousal relationship, there is a regular tendency for (i) spousal comparisons which may facilitate competition (since actions are observable in close proximity) and (ii) spousal evaluations which may influence confidence (since spouses often exchange opinions about one another). To test this, I use a field experiment with 1,328 spouses in Tanzania, where spouses must complete two rounds of an incentivised real-effort task. Before the first round, I elicit self-confidence beliefs, which capture how a respondent expects to perform. Before the second round of the task, I elicit what I term 'spouse-confidence' beliefs, which capture how a respondent expects their spouse to perform. I then use a 2 $\times$ 2 treatment design to experimentally vary whether, prior to round two, the individual receives information about their spouse's actual performance in round 1 (i.e., spousal comparison), their spouse's response to the spouse-confidence belief question (i.e., spousal evaluation), neither, or both types of feedback combined.

With this study, I add to the literature on effort and its motivators. Seminal work has linked gender inequalities in what motivates effort to a range of unequal gender outcomes in competitive environments such as education (Buser, Niederle and Oosterbeek, 2014; Buser, Peter and Wolter, 2017; Niederle and Vesterlund, 2010), labour markets (Buser, Niederle and Oosterbeek, 2014, 2024; Flory, Leibbrandt and List, 2015), tournaments (Backus et al., 2023), and even amongst top economists (Sarsons and Xu, 2021). However, to my knowledge, this is the first study to examine effort and its motivators in a spousal setting. This is surprising, not least because gender inequalities in outcomes are heavily documented in the household across several indicators such as wages, labour market choices, or even farm productivity, all of which depend on effort. If an individual's economic circumstances are shaped by how much effort they are willing to put in, and if spouses can reinforce or undermine that effort through their comparisons and evaluations which affect competitiveness and confidence, then understanding how spouses influence each other's effort is essential for explaining persistent gender inequalities within the household.

My results reveal interesting insights into spousal effort. From my summary statistics, I find that wives have larger improvements than men across *all* treatments, even when controlling for ability. I also find that husbands have less 'spouse-confidence' in their spouse than wives do (i.e., they are less likely than wives to believe that their spouse will be a top performer), and they also have lower spouse-*over*confidence (i.e., they believe the spouse will not be the top person when they actually are). From the experiment results, I observe gender differences in spousal responses to the two different types of feedback. When spousal comparisons are prompted, I find husbands to be more behind-averse than wives. When spousal evaluations are shared, I find that wives are more responsive than husbands to positive spousal evaluations, reducing (increasing) effort if they are ahead (behind). On the contrary, husbands are more responsive than wives to negative spousal evaluations, reducing effort irrespective of whether they are ahead or behind.

This study has two main policy implications. One implication relates to a future where dual career households become more widespread, and assortative mating leads to couples with similar interests and skills. A natural corollary to this could be that spousal comparisons become more common: where spouses perform similar tasks, work in comparable occupations, and pursue increasingly convergent career paths, each other's effort and performances would become a more salient reference point for comparison. In this scenario, my results show that both spouses are susceptible to comparisons, and appear to moderate effort accordingly. Where larger inequalities exist, this dynamic may offer an unexpected channel for narrowing gaps in wages or labour outcomes between spouses: if competitive individuals increase their effort in response to being comparatively behind, spousal comparisons could serve as an internal household mechanism that nudges performance upward. However, this same mechanism could entrench disparities if the more advantaged spouse responds by increasing effort even further. Understanding these dynamics is therefore key for policies that rely on household-level production, entrepreneurship, or shared labour decisions.

A second implication relates to sectors in which spouses' livelihoods begin to converge—a trend that may accelerate as labour markets become less gender-segregated due to environmental change or shifting economic demands. In such settings, confidence-building (or confidence-undermining) evaluations within the household can meaningfully shape effort and labour supply. For example, in Zanzibar, where environmental pressures and economic transitions are pushing households toward tourism, fisheries, and seaweed, spouses with sector-specific expertise (often the husbands) may be well positioned to encourage skill acquisition among their wives. Positive spousal evaluations could therefore support women's entry into emerging sectors, helping to narrow gender gaps in earnings and economic opportunities. This suggests that policies promoting joint training, shared performance feedback, or household-level labour interventions may generate behavioural spillovers that amplify their impact on gender equality.

In conclusion, my three chapters approach the nebulous field of intrahousehold decision-making by delineating three targeted lines of inquiry around child agency, spousal involvement, and spousal effort. They each contribute novel insights with theoretical and policy relevance for economists and policymakers working at the intersection of economic welfare, intrahousehold gender equity, and behavioural development economics. My hope is that by wading into the hidden choice architecture of household decisions—with equal power to perpetuate or diminish gender inequalities—I convincingly demonstrate to readers the merits of more systematically considering the behavioural motivators of household choices, the hinges of destiny for this generation and the next.

References

Ashraf, Nava. 2009. "Spousal control and intra-household decision making: An experimental study in the Philippines." *American Economic Review*, 99(4): 1245–1277.

Backus, Peter, Maria Cubel, Matej Guid, Santiago Sánchez-Pagés, and Enrique López Mañas. 2023. "Gender, competition, and performance: Evidence from chess players." *Quantitative Economics*, 14(1): 349–380.

Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. 2020. "Overconfidence and gender gaps in redistributive preferences: Cross-Country experimental evidence." *Journal of Economic Behavior & Organization*, 178: 267–286.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, competitiveness, and career choices." *The Quarterly Journal of Economics*, 129(3): 1409–1447.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2024. "Can Competitiveness Predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures." *The Review of Economics and Statistics*, 1–45.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *The American Economic Review*, 107(5): 125–130.

Dizon-Ross, Rebecca, and Seema Jayachandran. 2023. "Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation." *American Economic Review: Insights*, 5(4): 445–459.

D'Exelle, Ben, and Charlotte Ringdal. 2022. "Women's use of family planning services: An experiment on the husband's involvement." *Journal of Development Economics*, 158: 102915.

D'Exelle, Ben, and Charlotte Ringdal. 2025. "Who decides in the household when risk preferences conflict?" *Review of Economics of the Household*.

D'Exelle, Ben, Aurélia Lépine, Richard Bakyono, and Ludovic D. G. Tapsoba. 2023. "Fertility and polygyny: Experimental evidence from Burkina Faso." *Journal of Development Economics*, 164: 103134.

Flory, Jeffrey A, Andreas Leibbrandt, and John A List. 2015. "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions." *The Review of Economic Studies*, 82(1): 122–155.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics*, 118(3): 1049–1074.

Hillesland, Marya, Cheryl R Doss, Serena Masino, Martina Querejeta, Aletheia Amalia Donald, Greg Seymour, and Clare Clingain. 2025. "Deciding Not to Decide: When Is There Power in Not Deciding?" The World Bank. Policy Research Working Paper. Report Number WPS11065. Washington, D.C. : World Bank Group.

McCarthy, Aine Seitz. 2019. "Intimate partner violence and family planning decisions: Experimental evidence from rural Tanzania." *World Development*, 114: 156–174.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2022. "Managing self-confidence: Theory and experimental evidence." *Management Science*, 68(11): 7793–7817.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122(3): 1067–1101.

Niederle, Muriel, and Lise Vesterlund. 2010. "Explaining the gender gap in math test scores: The role of competition." *Journal of Economic Perspectives*, 24(2): 129–144.

Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and competition." *Annual Review of Economics*, 3(1): 601–630.

Sarsons, Heather, and Guo Xu. 2021. "Confidence Men? Evidence on Confidence and Gender among Top Economists." *AEA Papers and Proceedings*, 111: 65–68.